KOME – An International Journal of Pure Communication Inquiry
DOI: 10.17646/KOME.of.22

ARTICLE

Generating a Deepfake Frame

A Text Mining Study Based on Reddit

Genlong Zhou,* Fei Qiao**

* School of Journalism and Communication, Guangdong University of Foreign Studies, China
** 1. School of Journalism and Communication, Guangdong University of Foreign Studies, China;
2. Key Laboratory of Cross-Cultural Information Analysis and Intelligent Decision-Making;
3. Center for Public Opinion Governance and International Image Communication

This study investigates the understanding of deepfake, a highly realistic AI mimicry technique that is rapidly evolving to produce increasingly realistic videos and explores the construction of a deepfake framework through the lens of audience communication using framing theory. It identifies three key findings. First, the public discourse on deepfakes forms a concept hierarchy emphasising technology and its entertainment applications, with core concepts including AI, voice, actor and job, while peripheral concerns such as consent and company receive less focus. Second, employing the BERTopic algorithm, latent themes in public discussions were categorised into two dimensions: social dynamics and cultural phenomena. Third, sentiment analysis reveals predominantly neutral or negative attitudes, indicating concerns over the risks and societal impacts of deepfake technology. The deepfakes framework developed here provides a structured approach to understanding these impacts, highlighting the need for ethical considerations in technological development, regulatory measures and public education.

Keywords: deepfakes, frame theory, BERTopic model, word2vec, sentiment analysis

Address for correspondence: Fei Qiao, e-mail: jennifer.qf@gmail.com
Article received on 19 April 2024. Article accepted on 22 November 2024.
Conflict of interest: The authors declare no conflicts of interest.
Funding: This research was supported by the Guangdong Basic and Applied Basic Research Foundation (No. 2023A1515010670).

Introduction

Deepfakes are defined as "hyper-realistic videos digitally manipulated to depict people saying and doing things that never actually happened" as, for example, in the AI face swap used to "reanimate" the actor Paul Walker in *Fast & Furious 7* (Westerlund, 2019, p. 51). The popularity of generative AI applications has led to the widespread and pervasive presence of deepfakes. They are increasingly being spread across the web via social media, where deepfake content is convincingly and widely distributed. Powerful new AI software has made it easy to create and disseminate videos of people saying and doing things they never actually said or did.

Whittaker et al. (2023), in a systematic literature review on deepfakes identified several knowledge gaps and opportunities for future research. These gaps and opportunities fall under five research streams of interest: generation, information dissemination, adoption and rejection, impact and regulation and ethics. Current research tends to focus more on the development and detection of deepfake technology. However, from the public's perspective, the mere existence of deepfakes has already eroded confidence in digital content because seeing no longer equates to believing. The ability to create deepfakes is becoming increasingly accessible, and online platforms can rapidly distribute false content, threatening both belief systems and truth itself (Heidari et al., 2024). Although deepfakes present challenges, the technology also demonstrates artistic potential in virtual communication, entertainment and visual effects. Future research must continue to focus on balancing the beneficial potential of deepfakes while minimising their negative impact (Naitali et al., 2023).

Recent surveys indicate that the U.S. public perceives both potential benefits and risks associated with AI (Aoun, 2018; West, 2018; Zhang & Dafoe, 2019). These surveys also reveal that opinions about AI differ across political and demographic lines. However, existing research has paid less attention to the potential for communication to shape public attitudes toward deepfakes, an application of AI, or their potential to interact with one another in doing so. While spreading false information is easy, correcting the record and combating deepfakes are more challenging (De keersmaecker & Roets, 2017). Current research on deepfakes has concentrated on algorithmic approaches for detecting deepfake content (Bappy et al., 2019), the dangers and hazards of deepfakes (Godulla et al., 2021) and users' feelings about sexual deepfakes (Wang & Kim, 2022). With regard to deepfakes of the deceased, the resurrection of figures through deepfakes has met with complex responses: while some audiences found the deepfakes to be a powerful prosocial message, others reported feeling uncomfortable seeing the deceased being manipulated through deepfakes (Kneese, 2020).

This brings us to a crucial question: Given their complex attitudes, how do people talk about deepfakes? This is a matter we need to clarify. *Framing theory* offers a potential avenue through which to explore this issue. Framing theory addresses several key questions about the framing process (Reese, 2001; Chong & Druckman, 2007; Scheufele, 1999; Walsh, 1995), including the frames that the public uses to understand issues *(cognitive frames).* Such cognitive or *mental frames* refer to the mental templates individuals use to process information, interpret entities and environments, and determine appropriate

actions (Walsh, 1995). This concept provides a foundation for understanding deepfakes from the public's perspective.

There is growing research on artificial intelligence (AI) and its implications within the framework theory paradigm, aiming to construct different methodologies to explain the effects of AI and its reproductions on individuals, organisations or society (Makarius et al., 2020; Ashok et al., 2022; Huang & Rust, 2021). However, current discussions on AI frameworks rarely delve into AI simulations. Due to the unique nature of deepfakes – marked by their aggressiveness and clearer intentions – the AI framework does not fully accommodate the complexities of deepfake scenarios. Therefore, this research seeks to address the question: *What are the audience frames used by the public when engaging with deepfakes*? As a metatheory, framing itself encompasses various aspects of modern human life, and deepfakes, as a highly controversial use of technology, offer a valuable contribution to the application of framing theory.

This study is set against the backdrop of the increasing maturity and development of AI applications such as ChatGPT, Sora and ZAO. Through the lens of framing theory (Entman, 1993; Gamson & Modigliani, 1989; Reese, 2001), it aims to clarify how the public perceives deepfake technology and explore the underlying themes in public discussions of deepfakes. The goal is to analyse public discourse on deepfakes, identify key themes in public cognition on deepfakes and ultimately develop a framework for understanding how deepfakes are framed by the public.

Literature review and research questions

AI frameworks and public perceptions of AI

The study begins with a theoretical exploration of how past experience might influence users' understanding and attitudes towards deepfake technology. Framing, as defined by Gamson & Modigliani (1987, p. 143), refers to "a central organizing idea or storyline that provides meaning to unfolding events and connects them. Frames suggest what the controversy is about and the nature of the problem". Frames consist of metaphors, buzzwords, images and other symbolic devices that help construct meanings for topics, such as emerging technologies (Gamson & Modigliani, 1989). According to Entman (1993), framing involves selecting certain aspects of perceived reality and highlighting them in communicative texts to facilitate specific definitions of problems, causal explanations, moral evaluations and/or treatment recommendations (p. 52). Framing occurs at multiple levels: in communicative texts like news stories, Hollywood movies, television shows, and interpersonal conversations; in the minds of the audience; and as part of the broader culture (Entman, 1993).

A review of existing literature reveals that AI frameworks are generally categorised into four domains: physical, cognitive, informational and governance. Among these, the governance domain has been emphasised by many scholars because it addresses the controversies surrounding AI and how these challenges should be overcome. The ethical implications of AI span multiple scientific disciplines, which provides for ethical AI

analyses in various contexts (Ashok et al., 2022; Ulnicane et al., 2021; Akter et al., 2023). Interestingly, much research on AI frameworks also relies on the underlying logic of AI itself, using methods such as computational modelling, emotion recognition, sentiment analysis and human attention and performance monitoring (Górriz et al., 2023).

In studies focusing on AI frameworks, many adopt computational methods (Gourlet et al., 2024; Natale & Henrickson, 2024; Nguyen, 2023; Zeng et al., 2023), employing deep learning techniques to address real world applications and research problems related to AI frameworks. It is noteworthy that review articles tend to engage more deeply with AI frameworks than empirical studies, a distinction due not only to meta-analysis but also to the different levels of AI literacy represented by the research samples. AI literacy, a subset of digital literacy, refers to the ability to understand and apply AI technologies in the AI era, as well as the capacity to comprehend AI's societal impact. This literacy is crucial in adapting to rapidly changing technological environments (Domínguez Figaredo & Stoyanovich, 2023).

AI literacy involves not only interacting with AI applications but also recognising ethical issues (Steinbauer et al., 2021). In this context, ethics are defined as awareness of the responsibilities and risks associated with AI usage, requiring users to ensure AI technologies are used correctly and appropriately (Wang et al., 2023). However, current research remains largely concentrated in data science and computer science fields, leaving room for complementary studies that explore the ethical implications of AI from the perspective of media or audiences.

Audience discourse and audience framing of technology on social media

Social media platforms are increasingly recognised as valuable sources of information and public discourse. As social media has gained popularity within public communities, there has been a significant shift in how public opinion is analysed, moving beyond traditional surveys and interviews. Public perceptions of AI are still in the process of being shaped and developed, which makes understanding and analysing public opinion surrounding AI critically important (Zhou et al., 2024). By examining the public's perspectives, attitudes, and concerns, we can gain valuable insights into the current state of public perception, bridge knowledge gaps, and ensure that the development and deployment of AI technologies align with societal expectations and values (Qi et al., 2024).

Reddit, a popular social networking platform, stands out as a valuable source of data for research due to its large user base, diverse topic communities, and the anonymity of its users. These characteristics make it particularly useful for exploring topics such as politics and mental health, as its data offers unique perspectives and rich material for academic analysis (Li et al., 2023). We have chosen Reddit as the focus of this study because of its distinct social networking features. Unlike Facebook and Twitter, Reddit allows users to post longer comments and threads, encouraging more thoughtful discourse. This, in turn, can generate higher quality electronic word-of-mouth (eWoM). Such reflective communication offers a new lens through which we can explore how users disseminate and

interpret information, making Reddit a particularly valuable platform for in-depth social media research (Bonifazi et al., 2023).

Our research focuses on public discussions about technology on social media, which means we have taken an audience-centred approach. The influence of public discourse on people's perceptions of deepfakes has been well documented in public opinion research (Zaller, 1992). Audience frames refer to an individual's perceptions regarding an issue. They are defined as "interpretive patterns that enable individuals to perceive, organize, and understand incoming information" (Valkenburg et al., 1999). The public can use media messages or interact with peers to understand and evaluate AI (Claessens & Van den Bulck, 2016). For instance, research conducted through focus groups has shown that individuals acquire knowledge about topics such as nuclear energy (Gamson, 1992) and genetic technology (Bates, 2005) by engaging in conversations with their peers. During these discussions, the public not only draws from media discourse but also from their own values, experiences and reasoning abilities (Gamson, 1992, p. 117). Moreover, research has shown that interpersonal communication can influence people's attitudes towards various issues (De Vreese & Boomgaarden, 2006; Price et al., 2005). Consequently, some studies suggest that discussing science and technology can reconstruct attitudes towards emerging technologies by providing information and linking it to existing knowledge (Ho et al., 2013; Liu & Priest, 2009). The interaction between audience frames can reshape the original frames.

Previous research has identified two primary modes of public discussion regarding technology. One view presents new technologies as tools for solving problems and improving lives, while the other sees them as potential factors that could lead to uncontrolled or catastrophic consequences (Nisbet, 2009). According to Gamson and Modigliani (1989), framing is not simply about taking a positive or negative stance on an issue. A frame can encompass a range of positions, even though media messages may be dominated by a particular viewpoint. Moreover, there can be multiple "pro" and "con" positions on any given issue (Nisbet, 2009). This suggests that, depending on the perspective, audiences may hold different stances on a specific subject.

Individual attitudes toward deepfakes

Deepfake technology, which utilises deep learning and generative adversarial networks to create or manipulate videos, audio or images with high realism has both positive and negative implications. While it has promising applications in creative and entertainment fields, such as enhancing user experiences in Metaverse applications, improving the realism of virtual customer service agents, and serving as educational tools (Tricomi et al., 2023; Mustak et al., 2023), it also poses significant risks. These include the potential for misuse in generating false information, spreading rumours and committing identity fraud.

Unlike other AI technologies, deepfakes create highly realistic, computer-generated human representations, typically in video format. This realism implies that the content produced appears to be of *actual people* interacting in a human-like manner. Current research on deepfakes often focuses on their risks and negative impacts, such as information manipulation and fake news (Gamage et al., 2022). However, it is also important to

acknowledge the benefits of deepfake technology (Mustak et al., 2023), as mentioned above, its potential to enhance user experiences in virtual environments and serve educational purposes.

Discussions surrounding deepfake technology reveal diverse perspectives. While many users find deepfakes fascinating and impressive, there are significant concerns about potential misuse (Cleveland, 2022). On the one hand, deepfakes are criticised for their potential in creating fake news, invading privacy and manipulating public opinion (Saif et al., 2024). For instance, when potential voters are aware of the existence of deepfakes, they may even question the authenticity of genuine videos, potentially undermining their trust in political institutions and reinforcing beliefs about conspiracy (Ternovski et al., 2022). These concerns highlight the potential for deepfakes to disrupt social trust and legal systems. On the other hand, some argue that deepfake technology itself is not inherently harmful; rather, it is the application and regulation of the technology that need to be managed to ensure its positive development, such as its innovative uses in filmmaking and virtual reality (Ahmed, 2021).

Public attitudes toward deepfakes are central to the ongoing debate surrounding this technology. These attitudes not only influence the level of trust the public places in deepfake content but also determine the acceptance of deepfake technology across various applications, including virtual reality environments, virtual assistants and educational programs (Seymour et al., 2021). If the public is generally sceptical about deepfakes, the adoption and positive utilisation of these technologies may be hindered. In consequence, understanding and shaping public perceptions of deepfakes is crucially important in ensuring the responsible and healthy development of the technology. This requires balancing technological advancements with ethical considerations and enhancing public education and awareness, so that people can enjoy the benefits of the technology while recognising and mitigating potential risks.

Our research aims to delineate the core dimensions of the deepfake framework by exploring the central themes in public discussions about deepfakes. Guided by the motivating question, given their complex attitudes, how do people talk about deepfakes? We seek to understand the audience-based framework of deepfakes. Based on this guiding question, we have focused on three research questions (RQs):

RQ1: What is the conceptual hierarchy in public discussions about deepfakes? Which concepts are central and which are marginal?
RQ2: What are the potential themes in public discussions about deepfakes, and how are they categorised into dimensions?
RQ3: What are the emotional attitudes of the public within each theme?
RQ1 examines the priority structure of topics in public discourse. RQ2 explores the dimensional categorisation of the deepfake framework. RQ3 investigates public attitudes within specific dimensions. To address these RQs, we employed text mining computational methods on the collected data. Our findings reveal the potential of the deepfake framework, particularly in highlighting themes of public concern. This provides pathways for further empirical work in theorising and better understanding human–deepfake interactions.

Research design

Data

This study employed the Python Reddit API Wrapper (PRAW, https://praw.readthedocs.io/en/stable/) to collect comments related to deepfakes. The comments on the 20 most popular videos were collected using the keyword "deepfake" from 1 January, 2022 to 18 December, 2023. PRAW was used to retrieve and print comments from specific posts on Reddit. To begin, a Reddit object is created using PRAW and the necessary credentials – including client ID, client secret password, user agent, and username – are provided. Next, the URL of the target Reddit post from which comments are to be fetched is specified. Using the extracted post ID, the specific Reddit post object is obtained via the "reddit.submission (id=submission_id)" method. All the top-level comments are iterated and printed. In this process, "MoreComments" objects, which represent collapsed comments, are excluded as they do not require processing. After this, "submission.comments.replace_more(limit=None)" is called to retrieve all collapsed comments, which are replaced with actual comment objects. Finally, "submission.comments.list()" is used to obtain a list of all comments, which are then printed individually.

PRAW has been used by many previous studies to collect social media data (Deas et al., 2023). A total of 19,910 comment texts, containing 749,375 words, were collected from the 20 most popular videos on Reddit. The raw data was then processed using SPSS 29.0 software's data cleaning tools to eliminate incomplete samples and advertisements, and comments unrelated to deepfakes were manually removed. This process yielded 15,132 comment samples, totalling 569,045 words. The URLs of the twenty most popular videos of deepfakes on Reddit are listed below (see Table 1).

Table 1:
The most popular videos of deepfakes on Reddit

URL	Comments	Video name
www.reddit.com/r/technology/comments/14t4hd7/louisiana_outlaws_sexual_deepfakes_of_children/	509	Louisiana Outlaws Sexual Deepfakes of Children
www.reddit.com/r/Livestream-Fail/comments/10q7pot/destiny_reasons_out_why_deepfakes_fundamentally/	548	Destiny reasons out why deepfakes fundamentally feel violating
www.reddit.com/r/Livestream-Fail/comments/10pkdpn/xqc_take_on_people_saying_that_it_comes_with_the/	540	xQc take on people saying that "It comes with the territory" about deepfakes
www.reddit.com/r/Showerthoughts/comments/10t5pkg/deepfakes_are_ironically_taking_us_back_to_the/	560	Deepfakes are ironically taking us back to the pre-photography era of information where the only things we can be totally certain actually happened are events that we personally witnessed.

URL	Comments	Video name
www.reddit.com/r/TwoXChromosomes/comments/10pcawi/the_reaction_to_this_streamer_watching_deepfake/	557	The reaction to this streamer watching deepfake porn of women he knows is so scary to me.
www.reddit.com/r/conspiracy/comments/12nop3q/this_is_what_it_takes_for_normies_to_realize_the/	572	This is what it takes for normies to realize the danger of AI. Not deepfakes not their ability to clone your voice. I'm so tired.
www.reddit.com/r/LivestreamFail/comments/10sa1hj/dr_k_on_deepfake_pornography/	578	Dr. K on Deepfake Pornography
www.reddit.com/r/Futurology/comments/133m3vg/aigenerated_deepfakes_are_moving_fast/	630	AI-generated deepfakes are moving fast. Policymakers can't keep up.
www.reddit.com/r/Futurology/comments/1131q2r/keanu_reeves_says_deepfakes_are_scary_confirms/	635	Keanu Reeves Says Deepfakes Are Scary, Confirms His Film Contracts Ban Digital Edits to His Acting.
www.reddit.com/r/LivestreamFail/comments/11l3nh9/twitch_makes_some_changes_regarding_deepfakes/	676	Twitch makes some changes regarding "Deepfakes".
www.reddit.com/r/technology/comments/170iddp/deepfake_celebrities_begin_shilling_products_on/	796	Deepfake celebrities begin shilling products on social media, causing alarm.
www.reddit.com/r/skyrimmods/comments/12zel2z/it_happened_somebody_took_a_skyrim_voice_actors/	884	It happened. Somebody took a Skyrim voice actor's performance, fed through Eleven Labs to create AI-generated voices for a porn mod, and uploaded it to Nexus Mods. This is not acceptable.
www.reddit.com/r/ChatGPT/comments/156hcz7/chatgpt_wrote_all_the_words_coming_out_of_this/	938	ChatGPT wrote ALL the words coming out of this hyper-realistic deepfake – INSANE.
www.reddit.com/r/Damnthatsinteresting/comments/13l19qd/deepfakes_are_getting_too_good/	997	Deepfakes are getting too good.
www.reddit.com/r/technology/comments/16z1hyk/tiktok_ran_a_deepfake_ad_of_an_ai_mrbeast_hawking/	1,100	TikTok ran a deepfake ad of an AI MrBeast hawking iPhones for $2 – and it's the 'tip of the iceberg'
www.reddit.com/r/justneckbeardthings/comments/10rdpt9/how_dare_you_be_sad_about_people_making_deepfake/	1,200	How dare you be sad about people making deepfake porn of yourself? Like, grow up!
www.reddit.com/r/movies/comments/1130ocr/keanu_reeves_says_deepfakes_are_scary_confirms/	1,600	Keanu Reeves Says Deepfakes Are Scary, Confirms His Film Contracts Ban Digital Edits to His Acting.

URL	Comments	Video name
www.reddit.com/r/technology/comments/13einfa/deepfake_porn_election_disinformation_move_closer/	2,200	Deepfake porn, election disinformation move closer to being crimes in Minnesota.
www.reddit.com/r/gaming/comments/14tdayz/pc_gamer_anger_from_voice_actors_as_nsfw_mods_use/	3,600	[PC Gamer] Anger from voice actors as NSFW mods use AI deepfakes to replicate their voices: 'This is NOT okay.'
www.reddit.com/r/collapse/comments/12e0zv6/society_is_absolutely_asleep_at_the_wheel_in/	790	Society is absolutely asleep at the wheel in regards to the impact LLM's & AGI are going to have on the working class.

Source: compiled by the authors

It is important to note that Reddit does not allow us to extract information about users' geographic locations. Additionally, due to the lack of descriptive information about users, we are unable to collect additional data on factors such as age, gender or education level, which could potentially influence attitudes. According to a survey by the Pew Research Center, Reddit has unique mechanisms, and its user demographics differ from those of other social media platforms (Auxier & Anderson, 2021). These differences may provide supplementary insights into understanding public perceptions of deepfakes.

Text mining

The Gensim library was used for data processing. Gensim is a Python library designed for topic modelling, document indexing and similarity retrieval with large corpora, primarily servicing the natural language processing (NLP) and information retrieval (IR) users (Řehůřek & Sojka, 2010). The study utilised three text mining methods: The study began with semantic network analysis, using the Word2vec model to generate a semantic network graph in Gephi. Sentiment analysis was then conducted to determine the text's sentiment orientation. Finally, the BERTopic model was used for topic analysis. The program as a whole was completed using Python 3.10 software.

Gensim was used for text preprocessing, which involved tokenisation, lowercasing, stop word removal, and retaining only alphanumeric tokens. Next, a Word2Vec model was trained on the preprocessed text data using specific parameters such as vector size, window size, minimum count and number of workers (Church, 2017). Pairwise semantic similarity scores were then computed for each word in the vocabulary of the Word2Vec model. An edge was added between two words in the semantic network graph if their similarity score exceeded a threshold of 0.5. The resulting graph was then visualised using Gephi software (Bastian et al., 2009).

BERTopic, as an efficient text clustering tool, excels in extracting contextual meanings and semantic relationships from text by leveraging pretrained BERT models to capture deep semantic features. It automatically identifies and interprets interpretable topics, providing highly interpretable output results. This allows users to further simplify

topics based on domain knowledge. Empirical validations have demonstrated BERTopic's superiority over traditional topic modelling techniques such as LDA in handling large volumes of unstructured text data (Tang et al., 2024). The latest BERTopic algorithm has gained prominence in the field of topic modelling, with researchers from various domains applying it and validating its superiority and adaptability compared to other algorithms (Egger & Yu, 2022; Chen et al., 2023).

Mendonça & Figueira (2024) reviewed several studies using the BERTopic method and emphasised that UMAP better retains local and global features of high-dimensional data compared to alternatives such as PCA or t-SNE. HDBSCAN allows noise to be treated as outliers and does not assume centroid-based clusters, which provides advantages over other topic modelling techniques. Additionally, the classic TF-IDF variant used during the c-TF-IDF process generates a word bag at the cluster level, connecting all documents within the same cluster. TF-IDF is then applied to the word bag of each cluster, providing a measure for each cluster rather than for the entire corpus.

Past experiences support the use of our method, and BERTopic is recognised as a reliable topic classification tool. We have, therefore, utilised BERTopic for topic classification of our samples. In this study, we employed BERTopic topic modelling techniques to analyse text data, identify, as well as interpreting the underlying topic structures. Initially, we converted the text into vector form using the multilingual model from the Sentence-Transformer library for further analysis. Next, we reduced the dimensionality of the vector data using the UMAP algorithm and identified topics in the data using the HDBSCAN algorithm based on the set minimum cluster size parameters. During the training process, we applied a ClassTfidfTransformer to enhance the textual representation of topics. After training, we saved the BERTopic model to a file and created a DataFrame containing each document and its corresponding topic, which was then exported to an Excel file. Additionally, we generated various visualisations, including bar charts, distribution plots and hierarchical charts of topics and saved these visualisations as HTML files for intuitive presentation of the topic modelling results.

Although Gensim does not provide direct sentiment analysis capabilities, it can be a valuable tool for text processing to facilitate sentiment analysis tasks. Sentiment analysis is a crucial tool for mining social media opinions and can be categorised into two approaches: dictionary based and machine learning based (Bordoloi & Biswas, 2023). Dictionary based methods classify emotions by mapping words to emotional directions using predefined dictionaries. However, in practice, user-generated social media content often contains misspellings and internet slang, making dictionary based methods less effective (Chatterjee et al., 2019). Consequently, this study will rely on artificial intelligence and deep learning based sentiment analysis toolkits to improve robustness and flexibility.

We used Gensim along with the machine learning library Scikit-learn (Pedregosa et al., 2011) for sentiment analysis. We first converted each text sample into vector representations by averaging the word vectors derived from the Word2Vec model. This step transforms raw text data into a format suitable for machine learning algorithms. Next, we trained the vectorised text data using Scikit-learn's logistic regression classifier and generate sentiment analysis results. By leveraging the semantic information captured by Word2Vec embeddings, this method allows the sentiment analysis model to learn and

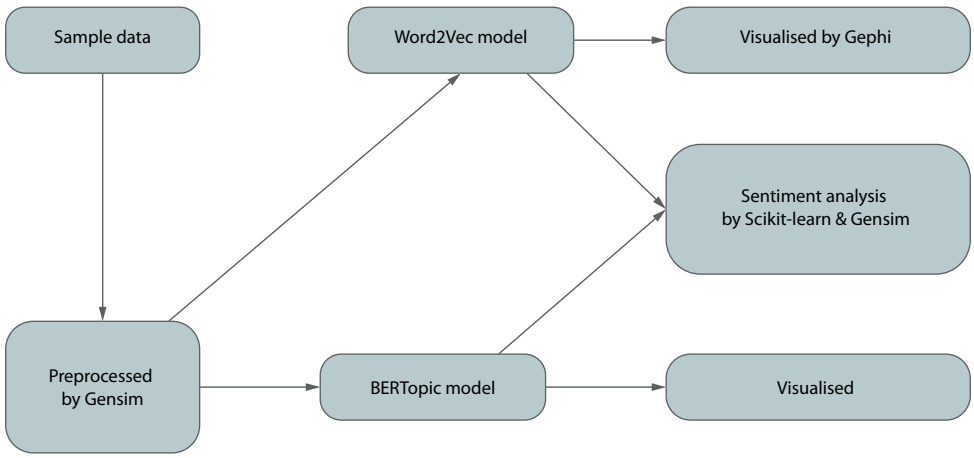

Figure 1:
Visualisation of the research process
Source: compiled by the authors

recognise emotional patterns in the text data. The combination of Gensim and Scikit-learn provides a powerful framework for conducting efficient and effective sentiment analysis tasks.

Specifically, each $post_c$ is labeled with positive sentiment QUOTE p_c^+ and negative sentiment p_c^-, with their sum equal to 1. Based on previous research (Stieglitz & Dang-Xuan, 2013), this study aggregates these two probability values into their difference to obtain a single sentiment polarity score $p_c = p_c^+ - p_c^-$. A positive polarity score indicates that $post_c$ is more likely to be positive (i.e. favourable) rather than negative, and vice versa. This study expects to perform sentiment polarity scoring for all posts. The resulting scores will operationalise the variables discussed in the framework. Figure 1 illustrates the operational steps mentioned in the research design.

Results

Construct semantic network of deepfakes

The data processed by the Word2vec program in Gensim was exported to Microsoft Excel and imported into the network visualisation program of the Gephi software (version 0.1.0). We created a semantic network graph for Reddit comments of deepfakes, where words are nodes and relationships between them are edges. The semantic networks are weighted undirected networks. Weighted degree and eigenvector centrality are used to identify key topics. Nodes with a higher degree of weighting are more strongly linked to other nodes, indicating their importance in the domain represented by the semantic network. Figure 2 shows the results of the comments in question are from the 20 Reddit videos that were analysed.

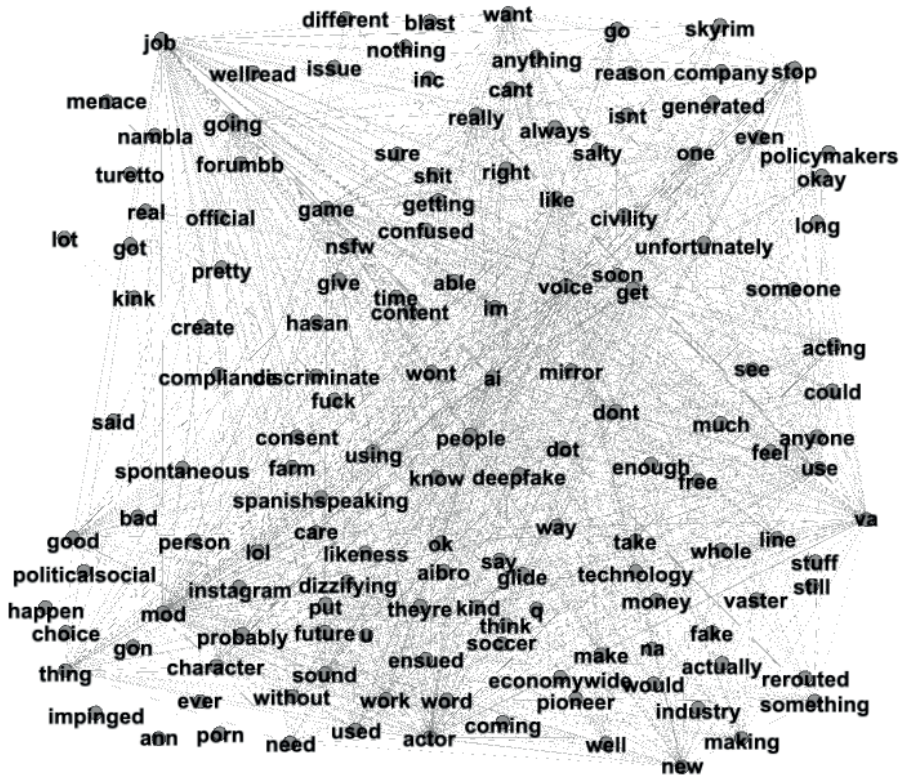

Figure 2:
Reddit – Deepfakes semantic network graph
Source: compiled by the authors

Table 2:
Top 10 calculations of weightiness and eigenvector centrality

Node	Weighted Degree Centrality	Eigenvector Centrality
people	52.074165	0.17240552768185900
like	40.911532	0.1499078425130630
ai	61.753613	0.1901658607102540
dont	43.468393	0.15751680700516600
would	13.922207	0.0661679654480382
think	28.429315	0.11528943155156200
voice	63.991547	0.1944615061402010
make	25.782831	0.10529209455008400
get	47.715491	0.16655182472667600
thing	31.827821	0.12745003892592100

Source: compiled by the authors

This semantic network graph comprises 147 nodes and 1,814 edges. In the public discussion of deepfakes, the conceptual hierarchy reveals that concepts related to *technology* and *applications* are positioned at the core, while discussions on *ethics* and *social impacts* are more peripheral. Core concepts include *AI, Voice, Actor, Job* and *Use.* The high centrality and feature vector centrality of these concepts indicate that public discourse is primarily focused on the development of deepfake technology and its applications within the entertainment industry, particularly concerning its impact on the careers of actors.

At the same time, the concept of *People* is also centrally positioned, reflecting public concern about the effects of deepfakes on privacy, reputation and social trust. In contrast, concepts such as *Consent*, *Company*, *Stop*, *Sound*, *Different* and *Bad* are positioned at the margins. Although these concepts address privacy issues, company policies and negative ethical viewpoints, they are emphasised relatively less in the overall discussion. This hierarchical structure indicates that the public is more focused on the technological development and application impacts of deepfakes rather than in-depth discussions of ethical or policy issues.

However, the high importance of the "people" node highlights the public's concern about the social impacts and moral consequences of deepfakes. It is crucial to consider the human aspect and how individuals should interact with deepfakes. This insight enriches our understanding of the deepfakes framework and underscores the importance of addressing the broader societal and ethical implications in future discussions and research.

BERTopic topic modelling and sentiment polarity analysis of deepfakes

This study applied the BERTopic algorithm, utilising individual modules SBERT, UMAP, HDBSCAN, and c-TF-IDF for modelling. The initial modelling was completed using default settings, incorporating the reduce_outliers algorithm to minimise noise interference. Without predefining the number of clusters, the model automatically generated 154 topics. As shown in Figure 3, the overall topic distribution exhibited characteristics of small scale aggregation and large scale dispersion, indicating potential for further aggregation between smaller topics. The figure illustrates the distribution of algorithm-generated topics in a two-dimensional scaling space. Each circle in the chart represents a distinct subtopic, and the size of the circle typically reflects the number of documents associated with that topic in the dataset. The position of the circles indicates the relative distance and similarity between topics: circles closer to each other suggest similar thematic content, while those further apart indicate greater content differences.

During the process of further determining the number of topics, we manually reviewed the original topic distribution in the figure above and the semantic network diagram to optimise and consolidate the topics. By continuously adjusting parameters related to BERTopic, such as "min_topic_size", we ultimately determined that 16 topics produced an optimal result. Figure 4 shows the distance distribution between topics, where each topic is relatively dispersed with minimal local overlap, indicating a relatively ideal clustering effect. Table 3 presents representative text content for each topic.

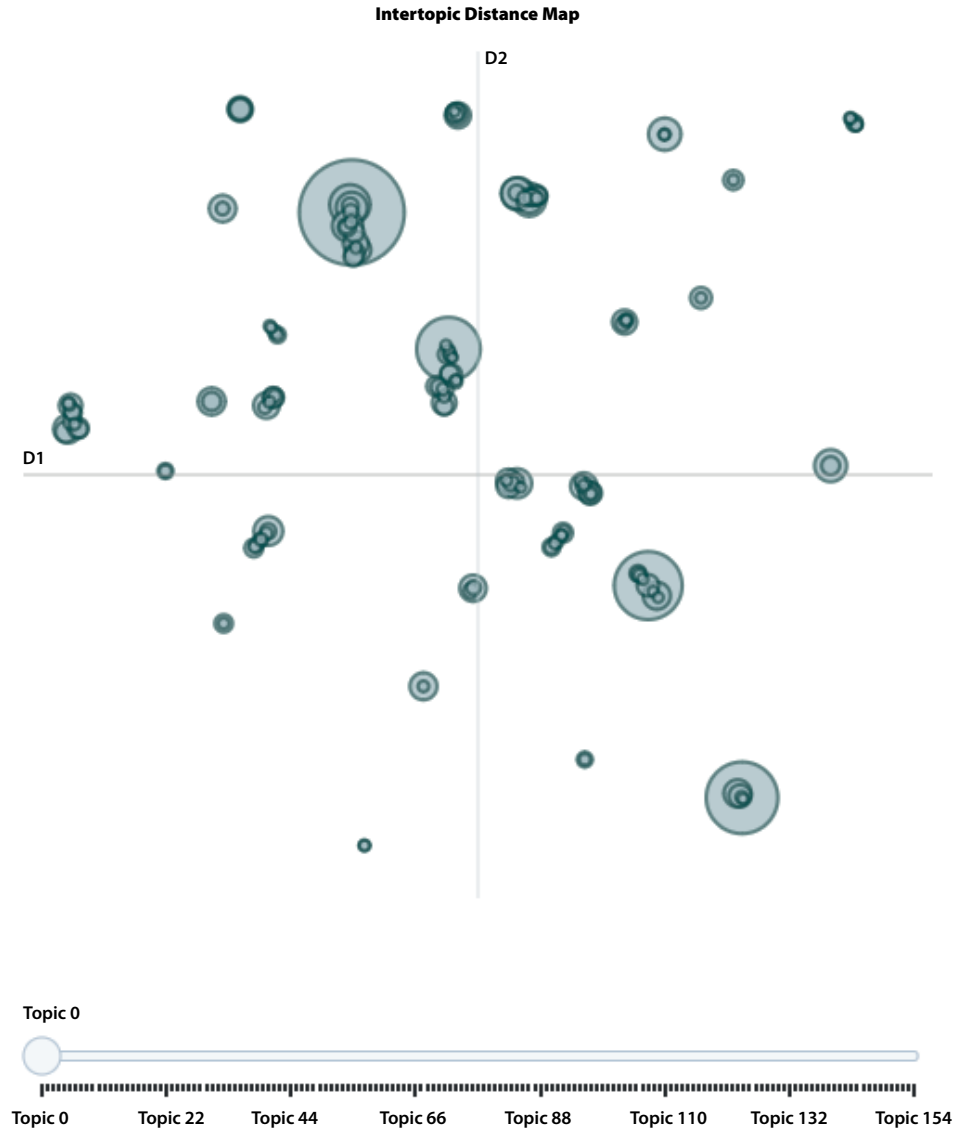

Figure 3:
Fully automated topic distribution
Source: compiled by the authors

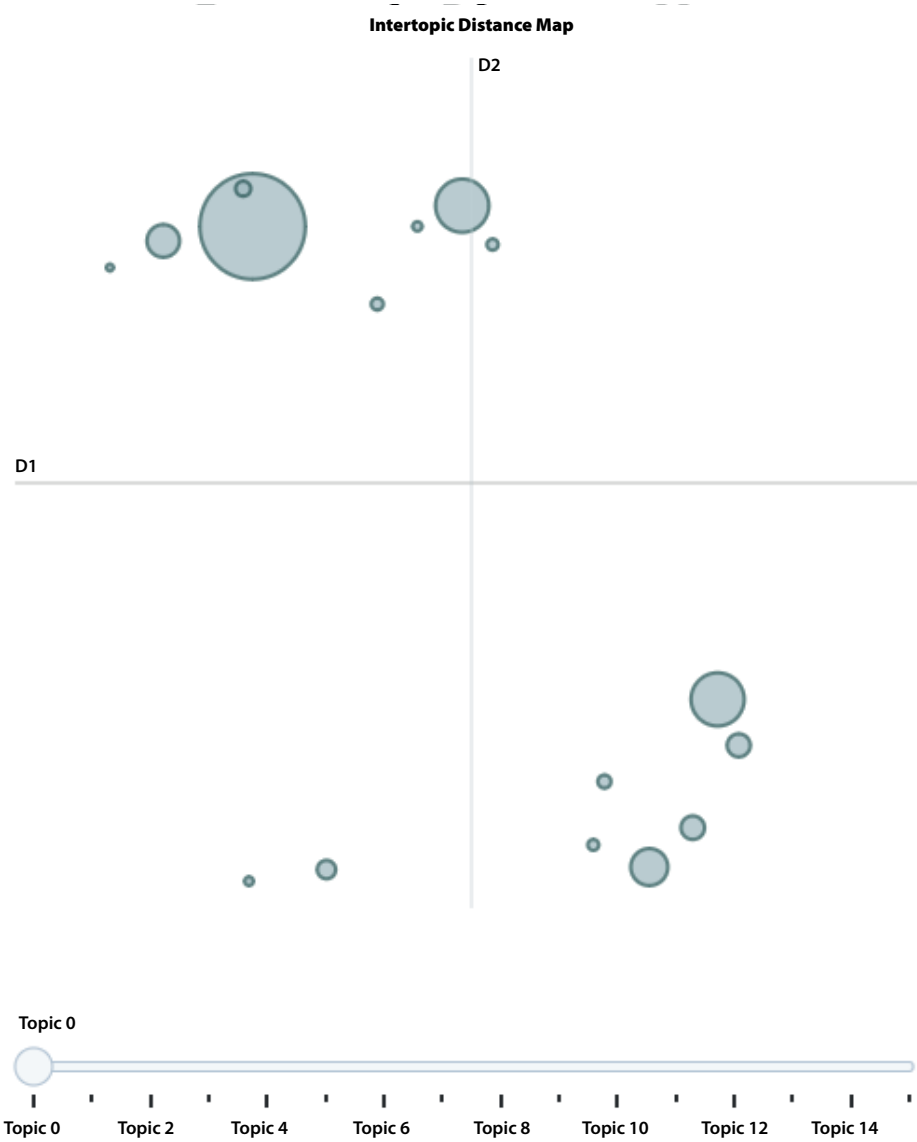

Figure 4:
The visualisation of the BERTopic model

Note: We uploaded the entire model to Google Drive and made it accessible via a dynamic web page after download (The visualisation of the BERTopic model: https://drive.google.com/file/d/1XoIL4IOgt0Ru3x-SJAmsXkzA3-1fM8bj/view?usp=share_link). Readers can explore the topic model for their own interests using an interactive, intuitive interface.

Source: compiled by the authors

Table 3:
Representative text content for each topic

Topic	Representation	Representative
0	['porn', 'voice', 'ai', 'like', 'people', 'actors', 'don't', 'women', 'think', 'deepfake']	['ai isn't limited voice cloning fact much better making new voices don't sound like particular person that's what's going replace voice actors', 'everyone porn made everyone make porn', 'honestly don't really see issue horny modders use computer skills put artificial voices mods long don't claim voices real vas don't see problem don't copyright voice don't think besides voice mods vas voice imitation essentially thing mod voiced someone really good impression original vat shouldn't person like able']
1	['school', 'people', 'election', 'don't', 'speech', 'law', 'laws', 'like', 'illegal', 'politicians']	['case let's say exact thing court case external event school participating likely involved bit planning school students community participating purely school activity', 'school trip kid outside school property school event happening also agree doesn't political expression value makes political expression vulnerable deeply political justices rather less you're implying honestly I'm confused stance okay limit freedom expression students don't matter ai generated political expression cant infringed draw lines supreme court', 'don't free speech school rules']
2	['jobs', 'ai', 'work', 'job', 'people', 'internet', 'technology', 'money', 'new', 'dont']	['want jobs', 'think people don't care say 34 years ai taking work forces jobs', 'omg I'm project manager advertising terrifying I've seriously considering else try work seems like blue collar jobs like plumbing carpentry construction etc. safe woman interest sure could actually type work I'm strong handy one thing saving grace white collar jobs ai advancement think people still want work people ai running projects clients want talk person companies super lean current job 3 full time employees I'm pm I've using ChatGPT able things quickly take projects simultaneously I'm worried shake']
3	['comment', 'edit', 'thanks', 'thank', 'fuck', 'read', 'that's', 'point', 'yeah', 'reading']	['yeah, sure comment', 'read comment lol', 'even read comment']
4	['reddit', 'social', 'media', 'mods', 'twitch', 'mod', 'facebook', 'streamers', 'nexus', 'content']	['much social media us talking social', 'reddit isn't technically social media least original definition social media follow people you're exposed content share interact reddit follow topics content see shared people social connection goes people interact comments social connections aspect defines social media less completely missing reddit', 'still social media']
5	['empathy', 'people', 'crying', 'trauma', 'feel', 'scary', 'victims', 'pain', 'emotional', 'like']	['don't need basic human empathy would enough', 'people like empathy would hate happened', 'I'm sure know empathy']
6	['world', 'time', 'future', 'years', 'humanity', 'live', 'old', 'ago', 'ai', 'end']	['could live real world', 'future ai', 'last watched 20 years ago still think time time']

7	['mouth', 'eyes', 'hands', 'facial', 'lips', 'movements', 'eye', 'face', 'expressions', 'movement']	['voice excellent mouth movements facial expression made clear fake still really impressive, good enough trick people sure', 'face eyes hair skin movements speaking mannerisms', 'facial expressions lips don't match intonation I'm sure it'll improve time really close']
8	['tom', 'cruise', 'scientology', 'hanks', 'fudge', 'guy', 'looks', 'hes', 'look', 'packer']	['tom cruise real tom cruise isn't real', 'except really tom cruise', 'tom cruise']
9	['deepfake', 'deepfakes', 'know', 'isnt', 'wait', 'willis', 'bruce', 'realistic', 'unrealkeanu', 'doesnt']	['someone makes deepfake', 'deepfake already ai', 'deepfake']
10	['jesus', 'christ', 'faith', 'religious', 'religion', 'pope', 'church', 'people', 'beliefs', 'bible']	['jesus christ clearly idea saying', 'jesus christ get ass', 'jesus christ lol']
11	['smoking', 'cigarettes', 'nicotine', 'vaping', 'health', 'smoke', 'cigarette', 'cancer', 'cardiovascular', 'smokers']	['mean said pretty straightforward I'm sure struggle smokers switch vaping cigarettes huge favor I'm saying people pick vaping don't smoke anything there's much harm reduction going cigarettes vaping suggest two pure ignorance', 'compared smoking cigarettes kills 6 million people around world every year nicotine less harmful inhaled cigarette smoke inhale nicotine wouldn't see kinds health risks health harms burdens see says nancy rigotti director tobacco research treatment center massachusetts general hospital cigarette smoke grab bag chemicals including 250 well-known bad us according national cancer institute includes heavy metals carbon monoxide hydrogen cyanide comes causing chronic lung disease cancer don't think nicotine big player rigotti says', 'you're talking smoking cigarettes vs vaping yes unlikely vaping cause damage long term smoking cigarettes could vaping possibly dangerous smoking cigarettes long term know ingredients']
12	['water', 'liquid', 'radioactive', 'peanut', 'sauce', 'diet', 'butter', 'balls', 'eat', 'boil']	['slow flowing liquid would like molasses tar pitch high viscosity liquid shown behave like liquid famous tar pitch drop experiment', 'edit rest story amorphous solid call amorphous liquid non flowing liquid whatever whole old glass windows flowing downward common misconception like always comes part discussion', 'know boil water']
13	['china', 'chinese', 'russia', 'russian', 'kinu', 'ukraine', 'offices', 'care', 'make', 'propaganda']	['would illegal china', 'yup even make laws west places like china russia ignore shit', 'also sue chinese companies think you'd china']
14	['mirror', 'black', 'episode', 'joan', 'season', 'awful', 'episodes', '6', 'new', 'steps']	['feels like black mirror episode', 'seen new black mirror episode', 'black mirror shit']
15	['friends', 'friend', 'elusive', 'friendship', 'asking', 'map', 'doors', 'open', 'maya', 'associated']	['big elusive friends ask', 'let take minute explain elusive friends' real life game mode called elusive friends', 'asking friend']

Source: compiled by the authors

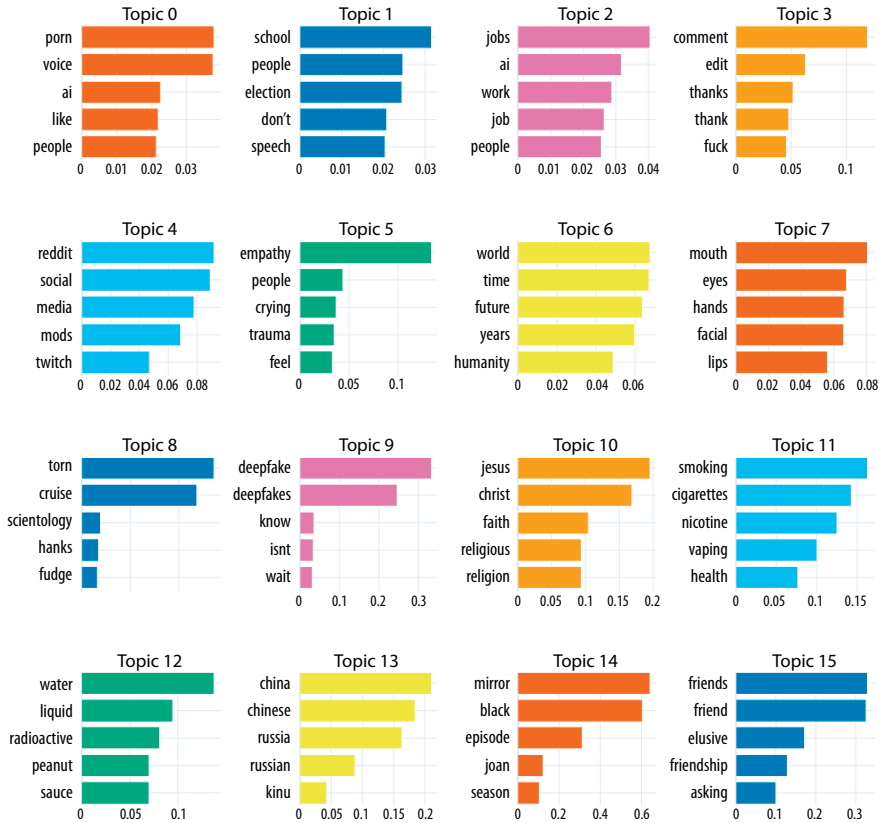

Figure 5:
The word score results by topic
Source: compiled by the authors

Figure 5 shows the keywords with the highest c-TF-IDF scores in each topic, illustrating the composition of different topics and identifying the most influential keywords defining these topics. Each subplot represents a topic, with the horizontal axis indicating the weight or contribution of the keywords within the respective topic, and the vertical axis listing the top-weighted keywords for each topic. The length of the bars in the bar chart represents the weight, with longer bars indicating a greater contribution of the keyword to the topic.

As shown in Figure 6, the hierarchical topic map generates a dendrogram to visualise the hierarchical clustering of the 16 topics, where topics of the same colour share greater similarity. The diagram illustrates the relative relationships and hierarchy between the extracted subtopics. Each topic is labelled with a number and descriptor on the left side of the chart, while the different coloured lines represent various topic categories and clustering branches, helping to distinguish broader group relationships between topics. On the horizontal axis, smaller values indicate greater content similarity between connected topics, while larger values indicate greater differences between topics, providing guidance for the hierarchical categorisation of the deepfakes framework.

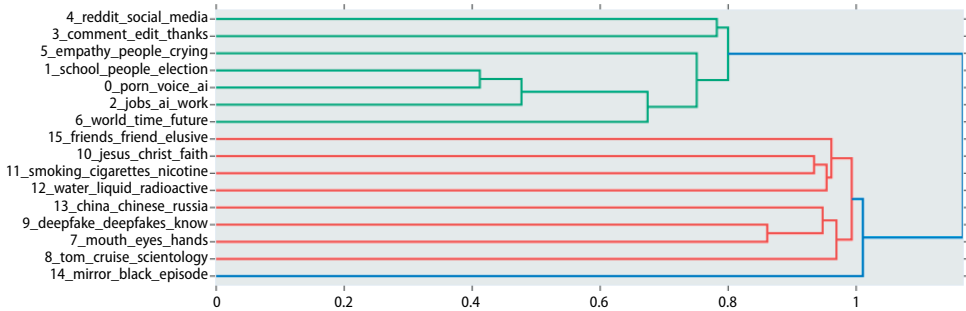

Figure 6:
The hierarchical map of topics
Source: compiled by the authors

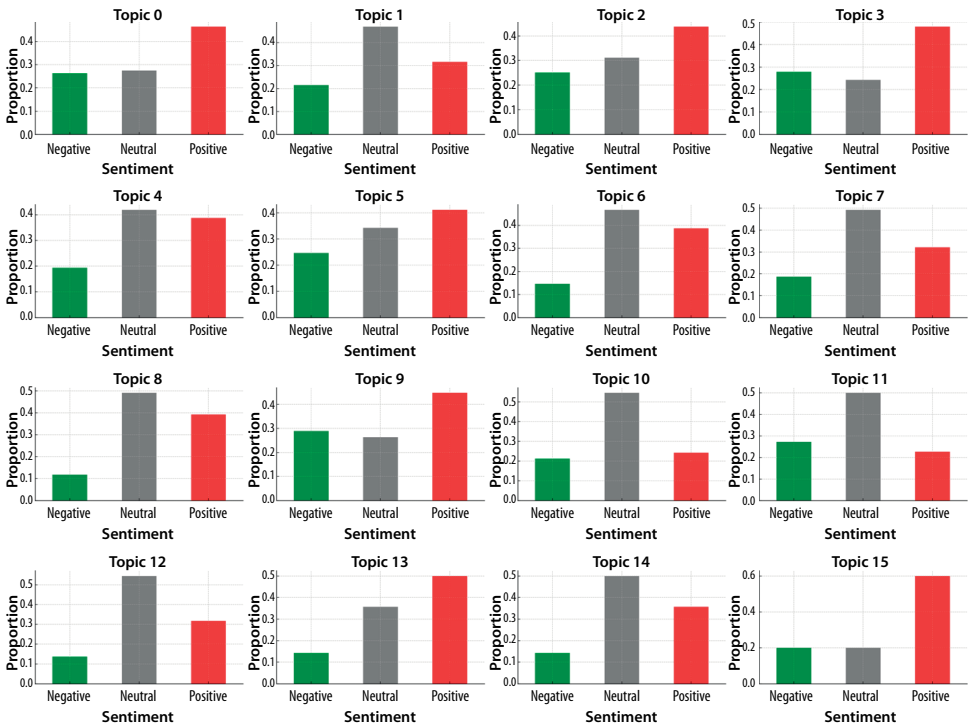

Figure 7:
The sentiment polarity analysis result of topics
Source: compiled by the authors

We generated a corresponding sentiment polarity analysis for each topic. The figure aims to show the sentiment tendencies within each topic. It is apparent that neutral or negative sentiments are more dominant in each topic, which also provides some insights for structuring our framework (see Figure 7).

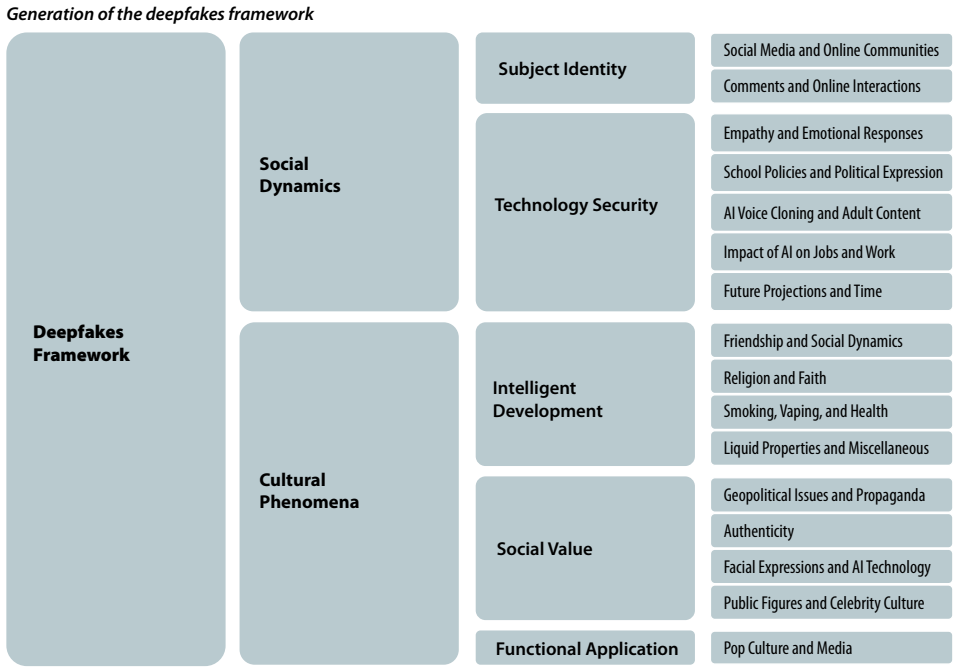

Figure 8:
The deepfakes framework of users
Source: compiled by the authors

In alignment with the findings, this study constructs a deepfakes framework that is exclusive to the audience, integrating the results from BERTopic's thematic exploration, thematic hierarchy and sentiment polarity analysis (see Figure 8). This framework illustrates the various levels involved when the audience discusses deepfakes. We have identified two significant themes: *Social Dynamics* and *Cultural Phenomena.* These dimensions not only reveal the mechanisms by which deepfakes operate within different social structures but also reflect their role in shaping contemporary cultural phenomena.

Social Dynamics refer to the patterns and changes in behaviours among individuals and collectives within social interactions. Within the context of deepfakes, this theme involves the way in which technology impacts the construction of personal identity, the nature of social interactions and the responses of legal and educational systems. Regarding the construction of *Subject Identity,* deepfake technology enables the facile imitation and reproduction of identity expressions and personal images. As demonstrated in representative documents from "Social Media and Online Communities", the public has a complex attitude towards AI-generated content, reflecting concerns about personal privacy and the authenticity of information. In the educational domain, the application of deepfake technology has sparked discussions about academic integrity and the veracity of knowledge. For instance, cases mentioned under the subtheme "School Policies and Political Expression" reveal the risks of deepfakes being used to disseminate misinformation or manipulate

public opinion. Furthermore, the subtheme "Empathy and Emotional Responses" within Technology Security underscores the influence of deepfakes in eliciting emotional resonance and moral considerations, particularly when dealing with content related to trauma.

Cultural Phenomena focus on how deepfakes shape and reflect societal values, belief systems and health perceptions. Under the framework of "Intelligent Development", deepfake technology has not only altered our understanding of friendship and social interaction but also triggered philosophical and ethical discussions on the authenticity of interpersonal relationships. The subtheme "Religion and Faith" explores the application of deepfakes in religious content creation and its potential impact on belief systems and spiritual practices. Additionally, the subtheme "Smoking, Vaping, and Health" addresses the role of deepfakes in health communication and their potential and challenges in educating the public about the risks of smoking and vaping products in particular. Against the backdrop of "Geopolitical Issues and Propaganda", deepfake technology is considered as a potential new propaganda tool, influencing international relations and political dynamics, as evidenced in discussions regarding countries such as "China" and "Russia". The subtheme "Pop Culture and Media" reflects a potential application of deepfakes in entertainment and media, namely the creation or imitation of celebrity images, which also raises discussions about celebrity imagery and personal privacy.

In summary, the role of deepfake technology in *Social Dynamics and Cultural Phenomena* is multifaceted and complex. It has not only changed the ways in which we communicate and express ourselves but also posed new challenges to fields such as law, education, health and international relations. As technology continues to evolve, we must adopt an interdisciplinary approach to understand and address these challenges, ensuring that the application of deepfake technology adheres to ethical standards and promotes the overall well-being of society.

Discussion and conclusion

This study was intended to examine the construction of the deepfake framework from the perspective of audience communication using framing theory. The research yielded three major findings. First, when discussing deepfakes, the public formed a concept hierarchy centred around technology and its applications. Core concepts such as *AI, Voice, Actor*, *Job* and *Use* reflect the public's primary focus on the development of deepfake technology and its applications in the entertainment industry. In contrast, peripheral concepts such as *Consent* and *Company* indicate a relatively lower emphasis on ethical and legal concerns. This hierarchy reveals both curiosity and concern about the potential of deepfake technology, highlighting the need for deeper discussions on the impact of its applications. Second, through the BERTopic algorithm, the study identified latent themes in public discussions, which were categorised into two main dimensions: social dynamics and cultural phenomena. In the social dynamics dimension, the themes addressed issues such as personal identity, social interactions and the responses of legal and educational systems. Meanwhile, in the cultural phenomena dimension, the themes focused on how deepfakes shape societal values, belief systems and perceptions of health. These dimensions encompass both the

social and cultural impacts of deepfakes, providing a multifaceted view of how the public perceives and understands this technology. Lastly, sentiment analysis showed that in each theme, public attitudes tend to lean toward neutral or negative emotions. This reflects public awareness of the potential risks of deepfake technology and uncertainty about the societal changes its applications may bring. This emotional tendency underscores the need for greater consideration of ethical and social impacts in the development and application of the technology, as well as the importance of cultivating a more balanced and comprehensive public understanding of deepfakes.

The deepfakes framework serves as an analytical tool for delving into the societal and cultural impacts of deepfake technology. By defining two core dimensions – Social Dynamics and Cultural Phenomena – it uncovers how deepfake technology shapes individual behaviours, social structures and cultural values.

Within the dimension of Social Dynamics, the framework thoroughly examines the effects of deepfake technology on the construction of personal identity, the nature of social interactions, the formulation of legal policies and the educational environment. For instance, deepfake technology's ability to easily impersonate and reproduce personal identities poses threats to individual privacy and reputation, challenging the authenticity of legal systems and educational content. The framework emphasises the close connection between technological development and social responsibility, indicating that the developers and users of technology must consider the societal impact and responsibilities alongside innovation.

The Cultural Phenomena dimension analyses how deepfake technology reflects and shapes religious beliefs, health perceptions, international relations and popular culture. This dimension illustrates the interplay between technology and culture, such as the application of deepfake technology in religious content creation, which may influence belief systems, or in health communication, potentially altering public awareness of health risks. Furthermore, the use of deepfake technology in international relations and political propaganda could impact national images and foreign policies, necessitating a collective effort from the global community to establish regulations and strategies for response.

The practical significance of this framework is that it provides a shared language and perspective for various stakeholders to understand and discuss the implications of deepfake technology. Raising awareness of the potential risks of deepfake technology not only helps the public but also guides developers in considering ethical, legal and social impacts during the design and implementation processes. Policymakers can utilise this framework to assess and formulate relevant policies to ensure that the application of technology does not harm societal interests.

In the academic realm, the deepfakes framework fosters the development of interdisciplinary research, integrating the findings from various fields such as communication studies, sociology, law and computer science. It offers researchers a comprehensive theoretical framework for analysing the societal impacts of deepfake technology and lays a solid theoretical foundation for future empirical research. Through this framework, researchers can explore the challenges and opportunities of deepfake technology applications in the real world more deeply, thereby providing scientific evidence for and practical guidance in constructing a more responsible and sustainable technological development environment.

It is important to note that most of the research on deepfakes has concentrated on the risks and negative perceptions associated with the technology, such as manipulation, misinformation and fake news (Gamage et al., 2022; Hancock & Bailenson, 2021; Lyu, 2020). The rapid advancement of deepfake technology, characterised by increasing quality, suggests a potential rise in its prevalence in the future. The application of deepfake technology in real-world scenarios is more likely to exacerbate a crisis of trust – a phenomenon reflected in the sentiment analysis across all subtopics, where neutral or negative emotions are more predominant. The ability of this technology to mimic the voices and images of real individuals with unprecedented realism allows for the creation of highly convincing false content on social media, political propaganda and even in everyday life. The direct consequence of such technological progress is a widespread public skepticism regarding the authenticity of media content, coupled with profound concerns about personal privacy and data security.

Understanding the potential risks associated with deepfakes, and accurately identifying them to ensure that artificial intelligence technologies such as deepfakes align with human values is of paramount importance. The regulation of AI, particularly AI-generated content, has become a new global risk and challenge. However, all discussions regarding regulation face a key dilemma: whether it is humans or AI that should be regulated? In addressing this question, let us return to the essence of artificial intelligence. Asimov (1942) proposed the "Three Laws of Robotics": First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm; Second Law: A robot must obey the orders given it by human beings, except where such orders would conflict with the First Law; Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

To some extent, the Three Laws of Robotics outline a general framework for *machine ethics,* implying that AI, under any circumstances, should be considered more as a tool than anything else. Whether in deepfakes or any of the other dilemmas brought about by AI technology, AI essentially exists as a tool rather than an independently thinking *person*. Thus, while many topics and scholars discuss how to constrain *AI,* fundamentally, we are discussing *the way in which people use tools*. To ensure that deepfake technology is used responsibly and to establish necessary trust in society, a multifaceted approach is required (Drabiak et al., 2023; Hagendorff, 2020). First, technology developers should incorporate ethical considerations from the design phase on to ensure that the use of technology does not harm society. Second, policymakers need to enact relevant laws and regulations to oversee the use of deepfake technology and prevent its misuse for improper purposes. Lastly, public education is crucial; we need to raise public awareness of deepfake technology and enhance their ability to discern truth from falsehood. Today, as deepfake technology matures, we face the challenge of rebuilding societal trust. By viewing AI technology as a tool and working together on ethical, legal and educational fronts, we hope to guide this technology towards beneficial societal development. Building trust takes time, but through responsible technology development and use, we can enjoy the conveniences it brings while maintaining a healthy, trustworthy social environment (Torresen, 2018).

Naturally, this study has areas that need further development. First, in terms of research methodology, while BERTopic is competitive with other models, it has certain

limitations. It assumes that each document contains only one topic, which may not always reflect reality. Additionally, a bag-of-words approach is used to generate topic representations, meaning it does not consider the relationships between words. As a result, the words within a topic may be redundant in explaining the theme, so it is essential to return to the original texts to validate the effectiveness of the results. Second, in terms of sample selection, different platforms foster different atmospheres, which can bias the research outcomes. Future studies could explore how people evaluate deepfakes across different platforms and even compare the discussions across various platforms to identify differences. Third, to date, there is no effective unified standard to evaluate or regulate AI use; much of the discussion revolves around ethical considerations. Therefore, we call for a policy framework – given the opportunities and risks that deepfake technology presents, what is needed now is a set of standards to align the utility of tools with humanity's rational values, ensuring that they develop together.

References

Ahmed, S. (2021). Fooled by the Fakes: Cognitive Differences in Perceived Claim Accuracy and Sharing Intention of Non-Political Deepfakes. *Personality and Individual Differences, 182*. Online: https://doi.org/10.1016/j.paid.2021.111074

Akter, S., Hossain, M. A., Sajib, S., Sultana, S., Rahman, M., Vrontis, D. & McCarthy, G. (2023). A Framework for AI-Powered Service Innovation Capability: Review and Agenda for Future Research. *Technovation, 125*. Online: https://doi.org/10.1016/j.technovation.2023.102768

Aoun, J. E. (2018). *Optimism and Anxiety: Views on the Impact of Artificial Intelligence and Higher Education's Response.* Gallup Inc.

Ashok, M., Madan, R., Joha, A. & Sivarajah, U. (2022). Ethical Framework for Artificial Intelligence and Digital Technologies. *International Journal of Information Management, 62*. Online: https://doi.org/10.1016/j.ijinfomgt.2021.102433

Asimov, I. (1942). *Runaround.* Street & Smith Publications, Inc.

Auxier, B. & Anderson, M. (2021). *Social Media Use in 2021.* Pew Research Center.

Bappy, J. H., Simons, C., Nataraj, L., Manjunath, B. S. & Roy-Chowdhury, A. K. (2019). Hybrid LSTM and Encoder–Decoder Architecture for Detection of Image Forgeries. *IEEE Transactions on Image Processing, 28*(7), 3286–3300. Online: https://doi.org/10.1109/TIP.2019.2895466

Bastian, M., Heymann, S. & Jacomy, M. (2009, March). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media, 3*(1), 361–362. Online: https://doi.org/10.1609/icwsm.v3i1.13937

Bates, B. R. (2005). Public Culture and Public Understanding of Genetics: A Focus Group Study. *Public Understanding of Science, 14*(1), 47–65. Online: https://doi.org/10.1177/0963662505048409

Bonifazi, G., Corradini, E., Ursino, D. & Virgili, L. (2023). Modeling, Evaluating, and Applying the eWoM Power of Reddit Posts. *Big Data and Cognitive Computing, 7*(1). Online: https://doi.org/10.3390/bdcc7010047

Bordoloi, M. & Biswas, S. K. (2023). Sentiment Analysis: A Survey on Design Framework, Applications and Future Scopes. *Artificial Intelligence Review, 56*(11), 12505–12560. Online: https://doi.org/10.1007/s10462-023-10442-2

Chatterjee, A., Gupta, U., Chinnakotla, M. K., Srikanth, R., Galley, M. & Agrawal, P. (2019). Understanding Emotions in Text Using Deep Learning and Big Data. *Computers in Human Behavior, 93*, 309–317. Online: https://doi.org/10.1016/j.chb.2018.12.029

Chen, W., Rabhi, F., Liao, W. & Al-Qudah, I. (2023). Leveraging State-of-the-Art Topic Modeling for News Impact Analysis on Financial Markets: A Comparative Study. *Electronics, 12*(12). Online: https://doi.org/10.3390/electronics12122605

Chong, D. & Druckman, J. N. (2007). Framing Theory. *Annual Review of Political Science, 10*(1), 103–126. Online: https://doi.org/10.1146/annurev.polisci.10.072805.103054

Church, K. W. (2017). Word2Vec. *Natural Language Engineering, 23*(1), 155–162. Online: https://doi.org/10.1017/S1351324916000334

Claessens, N. & Van den Bulck, H. (2016). A Severe Case of Disliking Bimbo Heidi, Scumbag Jesse and Bastard Tiger: Analysing Celebrities' Online Anti-Fans. In L. Duits, K. Zwaan & S. Reijnders (Eds.), *The Ashgate Research Companion to Fan Cultures* (pp. 63–75). Routledge.

Cleveland, K. (2022). *Creepy or Cool? An Exploration of Non-Malicious Deepfakes Through Analysis of Two Case Studies.* University of Maryland.

De keersmaecker, J. & Roets, A. (2017). 'Fake News': Incorrect, But Hard to Correct. The Role of Cognitive Ability on the Impact of False Information on Social Impressions. *Intelligence, 65*, 107–110. Online: https://doi.org/10.1016/j.intell.2017.10.005

De Vreese, C. H. & Boomgaarden, H. G. (2006). Media Message Flows and Interpersonal Communication: The Conditional Nature of Effects on Public Opinion. *Communication Research, 33*(1), 19–37. Online: https://doi.org/10.1177/0093650205283100

Deas, N., Kowalski, R., Finnell, S., Radovic, E., Carroll, H., Robbins, C., Cook, A., Hurley, K.,Cote, N., Evans, K., Lorenzo, I., Kiser, K., Mochizuki, G., Mock, M. & Brewer, L. (2023). I Just Want to Matter: Examining the Role of Anti-Mattering in Online Suicide Support Communities Using Natural Language Processing. *Computers in Human Behavior, 139*. Online: https://doi.org/10.1016/j.chb.2022.107499

Domínguez Figaredo, D. & Stoyanovich, J. (2023). Responsible AI Literacy: A Stakeholder-First Approach. *Big Data & Society, 10*(2). Online: https://doi.org/10.1177/20539517231219958

Drabiak, K., Kyzer, S., Nemov, V. & El Naqa, I. (2023). AI and Machine Learning Ethics, Law, Diversity, and Global Impact. *The British Journal of Radiology, 96*(1150). Online: https://doi.org/10.1259/bjr.20220934

Egger, R. & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology, 7*. Online: https://doi.org/10.3389/fsoc.2022.886498

Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication, 43*(4), 51–58. Online: https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Gamage, D., Ghasiya, P., Bonagiri, V., Whiting, M. E. & Sasahara, K. (2022). *Are Deepfakes Concerning? Analyzing Conversations of Deepfakes on Reddit and Exploring Societal Implications.* Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, USA (pp. 1–19). Online: https://doi.org/10.1145/3491102.3517446

Gamson, W. A. (1992). *Talking Politics.* Cambridge University Press.

Gamson, W. A. & Modigliani, A. (1987). The Changing Culture of Affirmative Action. *Research in Political Sociology, 3*(1), 137–177.

Gamson, W. A. & Modigliani, A. (1989). Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach. *American Journal of Sociology, 95*(1), 1–37. Online: https://psycnet.apa.org/doi/10.1086/229213

Godulla, A., Hoffmann, C. P. & Seibert, D. (2021). Dealing with Deepfakes – An Interdisciplinary Examination of the State of Research and Implications for Communication Studies. *Studies in Communication and Media, 10*(1), 72–96. Online: https://doi.org/10.5771/2192-4007-2021-1-72

Górriz, J. M., Álvarez-Illán, I., Álvarez-Marquina, A., Arco, J. E., Atzmueller, M., Ballarini, F., Barakova, E., Bologna, G., Bonomini, P., Castellanos-Dominguez, G., Castillo-Barnes, D., Cho, S. B., Contreras, R., Cuadra, J. M., Domínguez-Mateos, F., Duro, R. J., Elizondo, D., Fernández-Caballero, A., Fernandez-Jover, E. & Ferrández-Vicente, J. M. (2023). Computational Approaches to Explainable Artificial Intelligence: Advances in Theory, Applications and Trends. *Information Fusion, 100*. Online: https://doi.org/10.1016/j.inffus.2023.101945

Gourlet, P., Ricci, D. & Crépel, M. (2024). Reclaiming Artificial Intelligence Accounts: A Plea for a Participatory Turn in Artificial Intelligence Inquiries. *Big Data & Society, 11*(2). Online: https://doi.org/10.1177/20539517241248093

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines, 30*(1), 99–120. Online: https://doi.org/10.1007/s11023-020-09517-8

Hancock, J. T. & Bailenson, J. N. (2021). The Social Impact of Deepfakes. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 149–152. Online: https://doi.org/10.1089/cyber.2021.29208.jth

Heidari, A., Jafari Navimipour, N., Dag, H. & Unal, M. (2024). Deepfake Detection Using Deep Learning Methods: A Systematic and Comprehensive Review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14*(2). Online: https://doi.org/10.1002/widm.1520

Ho, S. S., Scheufele, D. A. & Corley, E. A. (2013). Factors Influencing Public Risk–Benefit Considerations of Nanotechnology: Assessing the Effects of Mass Media, Interpersonal Communication, and Elaborative Processing. *Public Understanding of Science, 22*(5), 606–623. Online: https://doi.org/10.1177/0963662511417936

Huang, M. H. & Rust, R. T. (2021). A Strategic Framework for Artificial Intelligence in Marketing. *Journal of the Academy of Marketing Science, 49*, 30–50. Online: https://doi.org/10.1007/s11747-020-00749-9

Kneese, T. (2020, November 2). How Data Can Create Full-On Apparitions of the Dead. *Slate.* Online: https://slate.com/technology/2020/11/robert-kardashian-joaquin-oliver-deepfakes-death.html

Li, S., Xie, Z., Chiu, D. K. & Ho, K. K. (2023). Sentiment Analysis and Topic Modeling Regarding Online Classes on the Reddit Platform: Educators Versus Learners. *Applied Sciences, 13*(4). Online: https://doi.org/10.3390/app13042250

Li, Y., Su, Z., Yang, J. & Gao, C. (2020). Exploiting Similarities of User Friendship Networks across Social Networks for User Identification. *Information Sciences, 506*, 78–98. Online: https://doi.org/10.1016/j.ins.2019.08.022

Liu, H. & Priest, S. (2009). Understanding Public Support for Stem Cell Research: Media Communication, Interpersonal Communication and Trust in Key Actors. *Public Understanding of Science, 18*(6), 704–718. Online: https://doi.org/10.1177/0963662508097625

Lyu, S. (2020). *Deepfake Detection: Current Challenges and Next Steps.* 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) IEEE. Online: https://doi.org/10.1109/ICMEW46912.2020.9105991

Makarius, E. E., Mukherjee, D., Fox, J. D. & Fox, A. K. (2020). Rising with the Machines: A Sociotechnical Framework for Bringing Artificial Intelligence into the Organization. *Journal of Business Research, 120*, 262–273. Online: https://doi.org/10.1016/j.jbusres.2020.07.045

Mendonça, M. & Figueira, Á. (2024). Topic Extraction: BERTopic's Insight into the 117th Congress's Twitterverse. *Informatics, 11*(1). Online: https://doi.org/10.3390/informatics11010008

Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A. & Dwivedi, Y. K. (2023). Deepfakes: Deceptions, Mitigations, and Opportunities. *Journal of Business Research, 154*. Online: https://doi.org/10.1016/j.jbusres.2022.113368

Naitali, A., Ridouani, M., Salahdine, F. & Kaabouch, N. (2023). Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions. *Computers, 12*(10). Online: https://doi.org/10.3390/computers12100216

Natale, S. & Henrickson, L. (2024). The Lovelace Effect: Perceptions of Creativity in Machines. *New Media & Society, 26*(4), 1909–1926. Online: https://doi.org/10.1177/14614448221077278

Nguyen, H. N. (2023). *Evaluation of Transition State Locating Algorithm Based on Artificial Intelligence.* Thesis, B.Sc. (Hons.) in Chemistry, University of Prince Edward Island.

Nisbet, M. C. (2009). Framing Science: A New Paradigm in Public Engagement. In L. Kahlor & P. Stout (Eds.) *Communicating Science: New Agendas in Communication.* Routledge. Online: https://doi.org/10.4324/9780203867631-10

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau D., Brucher, M., Perrot, M. & Duchesnay, É. (2011). Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. Online: https://doi.org/10.48550/arXiv.1201.0490

Price, V., Nir, L. & Cappella, J. N. (2005). Framing Public Discussion of Gay Civil Unions. *Public Opinion Quarterly, 69*(2), 179–212. Online: https://doi.org/10.1093/poq/nfi014

Qi, W., Pan, J., Lyu, H. & Luo, J. (2024). Excitements and Concerns in the Post-ChatGPT Era: Deciphering Public Perception of AI Through Social Media Analysis. *Telematics and Informatics, 92*. Online: https://doi.org/10.1016/j.tele.2024.102158

Reese, S. D. (2001). Framing Public Life: A Bridging Model for Media Research. In S. Reese, O. Gandy & A. Grant (Eds.), *Framing Public Life: Perspectives on Media and Our Understanding of the Social World* (pp. 7–31). Routledge.

Řehůřek, R. & Sojka, P. (2010). *Software Framework for Topic Modelling with Large Corpora.* Conference: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, Malta. Online: https://doi.org/10.13140/2.1.2393.1847

Saif, S., Tehseen, S. & Ali, S. S. (2024). Fake News or Real? Detecting Deepfake Videos Using Geometric Facial Structure and Graph Neural Network. *Technological Forecasting and Social Change, 205*. Online: https://doi.org/10.1016/j.techfore.2024.123471

Scheufele, D. A. (1999). Framing As a Theory of Media Effects. *Journal of Communication, 49*(1), 103–122. Online: https://doi.org/10.1111/j.1460-2466.1999.tb02784.x

Seymour, M., Riemer, K., Yuan, L. & Dennis, A. (2021). Beyond Deep Fakes: Conceptual Framework, Applications, and Research Agenda for Neural Rendering of Realistic Digital Faces. *Communications of the ACM, 66*(10), 56–67. Online: https://doi.org/10.1145/3584973

Shen, Q. & Rose, C. (2019). The Discourse of Online Content Moderation: Investigating Polarized User Responses to Changes in Reddit's Quarantine Policy. In S. T. Roberts, J. Tetreault, V. Prabhakaran, Z. Waseem (Eds.), *Proceedings of the Third Workshop on Abusive Language Online* (pp. 58–69). Association for Computational Linguistics. Online: https://doi.org/10.18653/v1/W19-3507

Steinbauer, G., Kandlhofer, M., Chklovski, T., Heintz, F. & Koenig, S. (2021). A Differentiated Discussion About AI Education K-12. *KI-Künstliche Intelligenz, 35*(2), 131–137. Online: https://doi.org/10.1007/s13218-021-00724-8

Stieglitz, S. & Dang-Xuan, L. (2013). Emotions and Information Diffusion in Social Media – Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems, 29*(4), 217–248. Online: https://doi.org/10.2753/MIS0742-1222290408

Tan, C. & Lee, L. (2015). *All Who Wander: On the Prevalence and Characteristics of Multi-Community Engagement*. Proceedings of the 24th International Conference on World Wide Web, Florence, Italy (pp. 1056–1066). Online: https://doi.org/10.1145/2736277.2741661

Tang, W., Bu, H., Zuo, Y. & Wu, J. (2024). Unlocking the Power of the Topic Content in News Headlines: BERTopic for Predicting Chinese Corporate Bond Defaults. *Finance Research Letters, 62*. Online: https://doi.org/10.1016/j.frl.2024.105062

Ternovski, J., Kalla, J. & Aronow, P. (2022). The Negative Consequences of Informing Voters about Deepfakes: Evidence from Two Survey Experiments. *Journal of Online Trust and Safety, 1*(2). Online: https://doi.org/10.54501/jots.v1i2.28

Torresen, J. (2018). A Review of Future and Ethical Perspectives of Robotics and AI. *Frontiers in Robotics and AI, 4*. Online: https://doi.org/10.3389/frobt.2017.00075

Tricomi, P. P., Nenna, F., Pajola, L., Conti, M. & Gamberini, L. (2023). You Can't Hide Behind Your Headset: User Profiling in Augmented and Virtual Reality. *IEEE Access, 11*, 9859–9875. Online: https://doi.org/10.1109/ACCESS.2023.3240071

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C. & Wanjiku, W. G. (2021). Framing Governance for a Contested Emerging Technology: Insights from AI Policy. *Policy and Society, 40*(2), 158–177. Online: https://doi.org/10.1080/14494035.2020.1855800

Valkenburg, P. M., Semetko, H. A. & De Vreese, C. H. (1999). The Effects of News Frames on Readers' Thoughts and Recall. *Communication Research, 26*(5), 550–569. Online: https://psycnet.apa.org/doi/10.1177/009365099026005002

Walsh, J. P. (1995). Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane. *Organization Science, 6*(3), 280–321. Online: https://psycnet.apa.org/doi/10.1287/orsc.6.3.280

Wang, B., Rau, P. L. P. & Yuan, T. (2023). Measuring User Competence in Using Artificial Intelligence: Validity and Reliability of Artificial Intelligence Literacy Scale. *Behaviour & Information Technology, 42*(9), 1324–1337. Online: https://doi.org/10.1080/0144929X.2022.2072768

Wang, S. & Kim, S. (2022). Users' Emotional and Behavioral Responses to Deepfake Videos of K-Pop Idols. *Computers in Human Behavior, 134*. Online: https://doi.org/10.1016/j.chb.2022.107305

West, D. M. (2018, May 21). Brookings Survey Finds Worries Over AI Impact on Jobs and Personal Privacy, Concern the U.S. Will Fall Behind China. *Brookings Institution.* Online: www.brookings.edu/articles/brookings-survey-finds-worries-over-ai-impact-on-jobs-and-personal-privacy-concern-u-s-will-fall-behind-china/

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review, 9*(11), 40–53. Online: https://doi.org/10.22215/timreview/1282

Whittaker, L., Mulcahy, R., Letheren, K., Kietzmann, J. & Russell-Bennett, R. (2023). Mapping the Deepfake Landscape for Innovation: A Multidisciplinary Systematic Review and Future Research Agenda. *Technovation,* 125(1). Online: http://dx.doi.org/10.1016/j.technovation.2023.102784

Zaller, J. (1992). *The Nature and Origins of Mass Opinion.* Cambridge University Press.

Zeng, C., Jian, Y., Vosoughi, S., Zeng, C. & Zhao, Y. (2023). Evaluating Native-Like Structures of RNA-Protein Complexes Through the Deep Learning Method. *Nature Communications, 14*(1), 1060. Online: https://doi.org/10.1038/s41467-023-36720-9

Zhang, B. & Dafoe, A. (2019). *Artificial Intelligence: American Attitudes and Trends.* Online: https://dx.doi.org/10.2139/ssrn.3312874

Zhou, Z., Zhou, X., Chen, Y. & Qi, H. (2024). Evolution of Online Public Opinions on Major Accidents: Implications for Post-Accident Response Based on Social Media Network. *Expert Systems with Applications, 235*. Online: https://doi.org/10.1016/j.eswa.2023.121307

This page intentionally left blank.