Human-Intelligent Systems Integration
https://doi.org/10.1007/s42454-025-00060-4

RESEARCH ARTICLE

Unmasking deepfakes: a multidisciplinary examination of social impacts and regulatory responses

Sami Alanazi[1] · Seemal Asif[1] · Antoinette Caird-daley[2] · Irene Moulitsas[3]

Received: 30 September 2024 / Accepted: 29 January 2025
© The Author(s) 2025

Abstract
This paper presents a comprehensive analysis of deepfake technology and its multifaceted impacts on society, privacy, trust, and information integrity. Deepfakes, synthetic media generated using AI-powered algorithms, pose significant challenges to individual privacy, societal trust, and the integrity of information. To explore these issues, we employed a mixed-methods approach that included in-depth expert interviews with professionals from diverse fields such as law, ethics, artificial intelligence, cybersecurity, and social sciences, along with a dichotomous question survey, which provided comprehensive insights from multiple perspectives. This methodological approach facilitated a multidimensional perspective on the potential risks and benefits of deepfakes. Our findings reveal unanimous concern among experts regarding the profound societal implications of deepfakes, particularly their capacity to amplify disinformation, erode public trust, and inflict psychological harm on individuals. Key themes identified include the urgent need for robust regulatory frameworks, the critical role of media literacy in enhancing public resilience, and the varying impacts of deepfakes across different demographic groups. The consensus among experts points out the necessity for an ethically guided approach to the development and deployment of deepfake technology, emphasizing the importance of interdisciplinary collaboration in crafting effective policy responses. This study advances the ongoing discourse on deepfake technology by providing stakeholders and policymakers with evidence-based recommendations aimed at mitigating the associated risks and harnessing potential benefits. These recommendations promote a balanced and informed approach to navigating the complexities of this emerging technological challenge.

Keywords Deepfake threads · Social impact · Misinformation · Ethical implication · AI regulations · Media literacy · Policy recommendations

1 Introduction

Deepfake technology, driven by advances in generative adversarial networks (GANs) and other AI-powered algorithms, has emerged as a transformative force in digital media, significantly affecting societal trust, individual privacy, and the integrity of information. Unlike traditional image or video manipulation, deepfakes can produce highly realistic yet entirely fabricated visual and auditory content, challenging our conventional understanding of authenticity in digital communication. Although deepfakes are frequently associated with misuse, they also present promising positive applications across various fields, including entertainment, education, accessibility, and healthcare (Kietzmann et al. 2020; Navarro Martínez et al. 2024; Patterson 2023). For instance, deepfake technology can enable realistic voiceovers in film production (Chesney and Citron, 2019) and improve language translation through advanced lip-syncing

✉ Sami Alanazi
sami.alanazi@cranfield.ac.uk

Seemal Asif
s.asif@cranfield.ac.uk

Antoinette Caird-daley
a.caird-daley@cranfield.ac.uk

Irene Moulitsas
i.moulitsas@cranfield.ac.uk

[1] Centre of Robotics and Assembly, Cranfield University, Bedford, UK

[2] Centre for Electronic Warfare, Information and Cyber, Cranfield University, Bedford, UK

[3] Centre for Computational Engineering Sciences, Cranfield University, Bedford, UK

Published online: 20 February 2025

techniques (Suwajanakorn et al. 2017). However, these benefits are often overshadowed by the potential for widespread misuse, underscoring the lack of comprehensive understanding of both their societal risks and their innovative potential (Mirsky & Lee 2020).

The growing prevalence of deepfakes highlights a critical gap in targeted regulatory responses, public awareness, and protective frameworks to address their multifaceted risks effectively. Prior studies have documented their disruptive potential in amplifying uncertainty on social media platforms (Vaccari & Chadwick 2020), with significant ramifications for civic culture and democratic processes. Recent incidents, such as fabricated content involving political figures during the Russian-Ukrainian conflict (Wakefield 2022), point out the need for more comprehensive regulatory strategies. Moreover, deepfake technology's alarming psychological and social toll remains underexplored, particularly its ability to exploit individuals on a personal level such as in cases of non-consensual pornography and AI-generated misinformation demonstrates its alarming psychological and social toll, for instance, the tragic story of Molly Russell, a 14-year-old who ended her life following exposure to harmful online content (North London Coroner's Service 2022). Another example of children being targeted with AI-generated content involves Isabel, a Spanish teenage girl, who discovered a nude photo of herself being circulated among her classmates. The young girl was terrified and depressed. She was not the only victim of deepfakes in her city; more than 20 teenage girls were targeted by AI-generated content in a city with a population of around 30,000, which is alarming (VIEJO 2023).

In response to these challenges, existing literature has not fully addressed the disconnect between technological advances and their societal implications. While legislative efforts, such as the UK's Online Safety Bill and the EU's Artificial Intelligence Act, represent progress, there is a need for studies that bridge the gap between regulatory efforts and the broader societal need for education and awareness (EC 2024; Loughran 2024). Social media platforms, despite their key role, also lack robust systems to mitigate the spread of deepfakes proactively (Alanazi et al. 2024).

This paper adopts a mixed-methods approach that combines a dichotomous question survey and in-depth interviews conducted with experts in law, ethics, AI, cybersecurity, and social sciences. By synthesizing perspectives from these diverse fields, this study seeks to bridge the research gap between the technological advancements of deepfake technology and the corresponding regulatory and societal responses required to mitigate its risks. This article examines the necessity of regulatory frameworks and educational initiatives to mitigate the risks of deepfake misuse while balancing innovation and individual rights. By proposing actionable recommendations, this study contributes to addressing the gap in aligning global regulations, ethical considerations (Appendix 1), and public education with the challenges and opportunities posed by deepfakes. Through an interdisciplinary analysis, this study formulates concrete steps to foster a safer and more trustworthy digital ecosystem, addressing the urgent gaps in the current research landscape.

2 Methods

2.1 Assessing expert opinions on the societal impact of deepfake technology

This method was presented during the International Robotics Showcase 2023, held in July in Bristol. This renowned event brings together leading companies and top universities at the forefront of robotics and autonomous systems. The showcase featured participation from industry giants like Ocado Technology, Dyson, ABB Robotics, and Thales, as well as research institutions such as the University of Oxford, Cambridge University, Imperial College London, and Bristol Robotics Laboratory. The event provided an opportunity to engage with AI experts from academia and industry, offering a unique platform to gather informed perspectives. As participation was limited to attendees of this robotics event, the number of participants could not be controlled. In total, 25 experts participated, comprising 8 women and 17 men, representing diverse expertise in AI, ethics, and robotics.

Before engaging participants in the voting process, we presented an overview of deepfake technology, including its creation, detection methods, and social implications. This background set the stage for a more informed voting session. The methodology for gauging expert opinions was straightforward yet effective. A closed yes/no question was posed: 'Will Deep Fake have a significant societal impact?' The simplicity of the binary question aimed to encourage straightforward responses and reduce ambiguity in interpretation. To collect responses, a voting system involving two boxes and colored tokens was set up: red tokens represented a 'No' vote, and yellow tokens signified a 'Yes,' as shown in Fig. 1. The system's simplicity was designed to encourage participation and provide a clear visual representation of the consensus.

2.2 Expert interviews and policy recommendations

2.2.1 Study setting and data collection

Our study took place in a digital environment where boundaries blur, and knowledge goes beyond geographical limitations. The central element of our research, the interviews, was exclusively conducted online through Microsoft Teams. This platform enabled smooth and secure communication,

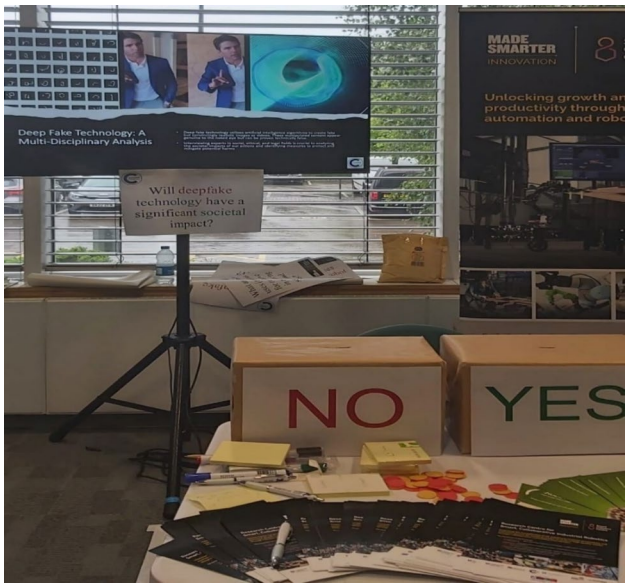


Fig. 1 Experts' voting system on deepfake's societal impact

offering a digital avenue that transcended physical limitations. Embracing this digital methodology allowed us to engage with experts from diverse locations, each contributing a unique perspective to enhance the discourse on deepfake technology.

Before each interview, participants were carefully chosen based on their expertise and achievements in their specific domains. Our objective was to assemble a group of experts capable of offering not only in-depth insights into deepfake technology but also a comprehensive understanding spanning legal, ethical, and psychological dimensions. The involvement of a chartered psychologist and a member of the British Psychological Society enhanced our study, providing profound insights into the psychological impacts of deepfake technology. Additionally, the inclusion of a magistrate and barrister added real-world perspectives, enriching our discussion with insights into the legal challenges posed by this technology in the judicial system.

Gathering data in a digital realm demanded careful consideration of ethical standards. Before each interview, participants were presented with a clear consent form outlining the purpose of the study, the voluntary nature of their participation, and the assurance of confidentiality. With their informed consent secured, interviews were recorded to ensure accurate and comprehensive analysis, respecting the nuances of each expert's perspective.

2.2.2 Participants

Our study comprised 14 diverse and accomplished experts, each selected for their rich background and expertise across various disciplines. The participants, consisting of nine women and five men, brought a wealth of knowledge that enhanced the depth and breadth of our exploration into the social impacts of deepfake technology. This included a senior researcher in applied psychology and cognitive ergonomics, a researcher specializing in dementia and ageing, a lecturer in human–computer interaction, a chartered psychologist and crisis resilience specialist, a cognitive psychologist with expertise in individual differences, a chartered occupational psychologist and course director, a business psychologist and strategic advisor, an experienced legal advisor, a professor of law & technology, a researcher in social informatics, a law lecturer with a background in theology, a researcher in human-technology interaction, and a senior research fellow in psychology. Table 1 shows the characteristics of participants. Their backgrounds spanned areas such as cybersecurity, law, human–robot interaction, and cognitive psychology. This diverse array of participants ensured a comprehensive exploration of the social impacts and legislation of deepfake technology, enriching our study with a multifaceted understanding derived from their collective knowledge and real-world experiences.

Our study engaged 14 experts from diverse professional backgrounds, carefully selected for their expertise in fields related to deepfake technology and its societal implications. The participants included nine women and five men, with specialties spanning applied psychology, law, human-technology interaction, social informatics, and human–robot collaboration. Their inclusion brought a wide range of perspectives, ensuring a comprehensive and interdisciplinary exploration of the topic. Table 1 provides an overview of the participants' roles, affiliations, and areas of expertise.

To delve into the societal, psychological, and ethical dimensions of deepfake technology, we posed a series of carefully designed questions during the interviews (Appendix 2). These questions focused on the participants' understanding of deepfake technology, its societal impact, and its psychological and ethical implications. For instance, participants were asked about the potential for deepfake technology to exacerbate disinformation and manipulation, its influence on social norms and cultural values, and the specific populations most affected by this technology. In addition, the discussions explored potential regulatory frameworks, ethical guidelines, and strategies for promoting media literacy and mitigating risks associated with deepfakes.

3 Analysis process

3.1 Overview of voting results

The voting boxes were securely sealed. Upon opening them, we discovered that all 25 voters had voted 'Yes.' This led to the identification of a unanimous concern among experts:

Table 1 Participants' characteristics

Participant	Role and affiliation	Expertise and highlights
P1	Senior researcher	PhD in applied psychology and cognitive ergonomics. International experience. Awards for interdisciplinary research
P2	Researcher in dementia and ageing	Postdoctoral position contributing to Sustainable Development Goals. Expertise in dementia and ageing
P3	Lecturer in human–computer interaction	PhD in HCP. Expertise in human–robot interaction and computational psychology. Module coordinator for user interface design and HCI
P4	Chartered psychologist, magistrate, and crisis resilience specialist	Founder of a company that specializes in cybersecurity and crisis management
P5	Cognitive psychologist with expertise in individual differences	Diverse experience in driving simulation, psychophysiology, and human factors. Proficient in mixed methods approaches
P6	Chartered occupational psychologist and course director	Lecturer in occupational psychology. Expertise in cyber security, defence training, and human performance
P7	Business psychologist and strategic advisor	Chartered psychologist with experience in organizational and leadership development. Recognized for work in occupational psychology
P8	Senior human factors consultant	Human factors in industrial human–robot collaboration, rail industry, aerospace manufacturing, advanced research skills, cutting-edge research, and practical applications in human factors and automation
P9	Experienced legal advisor	Holds a PhD in international commercial law. Skilled in legal writing, litigation, and legal advice
P10	Professor of law and technology	International leader in regulation and governance of new technologies, knowledge exchange, internet law and policy
P11	Researcher with expertise in social informatics	Holds a PhD in social informatics. Expertise in securing research funding and working on EU-funded projects
P12	Law lecturer	Solicitor and barrister with teaching experience. Involved in legal education, and curriculum development
P13	Researcher in human-technology interaction	Cognitive psychology, human-technology interaction, user experience, technology integration
P14	Senior research fellow	Human factors in industrial human–robot collaboration, aerospace manufacturing, advanced research skills and cutting-edge research

3.1.1 Unanimous concern among experts

The voting results revealed a unanimous 'Yes' vote from all participating experts. This unanimous decision draws attention to a widespread concern among AI professionals regarding the societal implications of deepfakes. The data demonstrates a robust consensus across different fields concerning the potential repercussions of this technology.

3.1.2 Industry and academic consensus

The participants, representing both industrial and academic sectors, demonstrated a cohesive stance on the impact of deepfakes. This broad agreement underscores the shared concerns across various sectors about the potential risks and implications associated with deepfake technology.

3.2 Qualitative coding of expert interviews

In our research, we adopted the thematic analysis framework proposed by Braun and Clarke (2023), which advocates for reflexivity and theoretical sensitivity as fundamental to conducting rigorous research. This approach, underscored by the definitions of themes by Saunders et al. (2016) and expanded upon by King and Horrocks (2010), enabled us to scrutinize the complex narratives surrounding deepfakes through a critical lens. Utilizing NVivo, a software designed for qualitative data analysis, we enhanced our methodological rigor by incorporating features for text illuminating, note-taking, and conceptual linking, as described by Jackson and Bazeley (2019). NVivo's capacity to manage large volumes of data proved invaluable, especially in coding and segmenting interview

transcripts for a more nuanced analysis (Welsh 2002). Drawing parallels with the work of Callari et al. (2024) in their exploration of ethical frameworks for human–robot collaboration in manufacturing, we found NVivo to be an essential tool in organizing and refining our thematic analysis. Similar to their study, which employed NVivo for clustering and refining ethical themes within a collaborative design process, we utilized the software to enhance our understanding of deepfake technology's societal impacts. This comparative approach underscores the value of rigorous qualitative methods in analysing complex ethical issues across different technological domains. By implementing inter-rater reliability measures, we maintained coding consistency, and NVivo's visualization tools helped us construct diagrams that visually represented interconnected themes, offering a deeper understanding of the data. The themes we identified, reflecting profound ethical dilemmas and societal concerns associated with deepfakes, were vital for illustrating our analytical process and were substantiated by participant quotes, shedding light on their perspectives (Appendix 3) (Levitt et al. 2018).

4 Result

Through our detailed analysis, we derived insights from two complementary studies: the Integrated Analysis of Expert Opinions based on voting results from the International Robotics Showcase and the Expert Interviews capturing in-depth qualitative insights. These findings together illuminate the multifaceted nature of deepfake technology and its societal implications.

4.1 Integrated analysis of expert opinions

The integrated analysis of specialists' opinions reveals a unanimous concern about the dangers posed by deepfakes, as evident from voting results and interviews, emphasizing the need for societal awareness and policy intervention. This shared worry among experts underscores the critical nature of the threats deepfakes pose, ranging from violations of individual privacy to broader issues of societal trust. The consensus extends across industry and academia, indicating the wide-reaching implications of deepfakes that go beyond theoretical discussions and require collaborative efforts for effective solutions. Both sectors acknowledge the potential of deepfakes to affect politics, media, and individual rights, bringing to the forefront the necessity for a unified approach to addressing these challenges.

The unanimous agreement among experts stresses the urgency for proactive policy and regulatory measures that balance technological innovation with ethical considerations and societal welfare. The discussions advocate for adaptable regulations to keep pace with the evolving nature of deepfake technology. Furthermore, the findings point to the importance of interdisciplinary research to explore the societal, legal, ethical, psychological, and social dimensions of deepfakes, aiming to find effective risk mitigation strategies and understand potential benefits.

Moreover, the consensus on ethical considerations in the development of AI technologies like deepfakes stresses the importance of incorporating ethical principles throughout the lifecycle of these technologies. It calls for a conscientious approach that prioritizes the societal impacts of technological advancements. This collective insight from experts across various fields highlights a critical path forward in navigating the challenges and opportunities presented by deepfake technology.

4.2 Detailed discussion of expert interviews with visual representations and quotes

Our analysis revealed seven primary themes, each accompanied by sub-themes that delve into the complexities of deepfake technology. To provide a richer understanding of the findings, we have incorporated direct quotes from participants, visual representations, and detailed explanations of the sub-themes. These additions aim to clearly link the insights to the overarching themes, offering a nuanced perspective on the topic.

4.2.1 Understanding deepfake technology

This theme explores the core capabilities of deepfake technology and its far-reaching implications for digital media. Participants described how deepfakes manipulate audio and visual content to produce highly realistic yet fabricated representations, blurring the line between reality and illusion. Subthemes such as voice manipulation illustrate the technology's ability to replicate an individual's speech with minimal input, while another key concern involves the deliberate targeting of groups like celebrities or political figures for harmful purposes.

A participant's reflection on the evolution of trust in visual content lays the foundation for understanding the transformative impact of deepfakes. P10, a female professor in law and international leader in the regulation and governance of new technologies, said, 'In the past, when we encountered a photograph or video, we tended to accept it as truth… However, this perspective has evolved'. This evolution underscores a growing skepticism towards digital content, amplified by the capabilities of deepfake technology to convincingly imitate reality.

Deepfake technology, described by one participant as 'a technology powered by artificial intelligence to manipulate or create realistic-looking audio, video, or image content' (P11), represents a significant leap in the realm of digital manipulation. The ability to create content that blurs the line between real and

fabricated not only challenges our ability to discern truth but also raises important questions about the integrity of digital media.

The comparison of deepfakes to traditional image manipulation techniques, such as Photoshop, was a recurring theme. However, participants noted the ease with which deepfakes can be created and their potential to alter reality more convincingly: 'It's in its own way… We are creating a fake image from what we already have. So it feels like it's the extension by which the tools are being used' (P3). It does not require an expert to make a highly realistic video due to the availability of software and apps that can do this specific purpose (Alanazi & Asif 2024). This observation highlights the accessibility of deepfake technology and its capacity to extend beyond conventional image editing, offering both creative opportunities and challenges in discerning authenticity.

The dialogue with participants illuminated the technological complexity required to discern deepfakes, revealing the continuous struggle between their creation and detection. Traditional methods of image verification and forgery detection, such as watermarking and statistical analysis, often fall short when applied to sophisticated deepfake images (Hsu et al. 2020). Initial efforts in detecting deepfakes relied heavily on manual feature extraction and biological cues, but these approaches have proven inadequate against the rapidly advancing capabilities of deepfake technology (Güera & Delp 2018). These limitations underscore the urgent need for more sophisticated detection strategies capable of keeping pace with evolving threats. This complexity, necessitating advanced methods to differentiate real from manufactured content, mirrors concerns about digital content's trustworthiness and misuse potential. Furthermore, apprehensions extend to the generation of fake documents and the adaptation of conventional image manipulation tools. The rising occurrence of deepfake videos, the generation of image-based models, and a growing familiarity with these technologies suggest an expanding and dynamic field. Additionally, fears regarding specific groups being targeted and voice manipulation through deepfake techniques underscore the technology's broadening and rapidly progressing capabilities Voice fraud, a particular worry associated with deepfakes, exemplifies the ease of impersonating individuals' voices, as demonstrated by a notable incident involving a live Kabatznik in Florida. His bank thwarted a significant fraudulent money transfer requested in his voice, thanks to their phone system's detection capabilities (Flitter & Cowley 2023), highlighting the critical need for advanced detection and prevention strategies in this evolving landscape.

4.2.2 Navigating the complexities of deepfake legislation enforcement

This theme examines the regulatory and enforcement challenges associated with deepfakes. Participants pointed to issues such as the variation in applicable laws across jurisdictions and the difficulty of source identification, which complicates efforts to trace the origins of deepfake content. Furthermore, experts noted a significant gap in technical and legal expertise among lawmakers and enforcers, which impedes the creation of effective legal frameworks and complicates the enforcement of existing regulations.

As one interviewee insightfully notes, "Enforcing measures to regulate deepfake technology presents several significant challenges… there is a substantial gap in understanding the intricacies of this technology among many law and policymakers" (P10). This points out the essential need for legislators to either possess a deep understanding of these technologies or to seek advice from technical experts to formulate laws that are both effective and enforceable. The complexity also arises from the existence of law enforcement agencies alongside a legal system. Concerns are growing about the potential use of deepfakes as evidence in court, which poses challenges in distinguishing between genuine evidence and deepfakes. Breen (2021) contends that efforts should primarily aim to reduce skepticism about the validity of video and image evidence. Measures to achieve this could involve creating protocols to check the veracity and authenticity of evidence, alongside defining criteria for the acceptance of digital evidence in legal proceedings (Cover 2022).

Additionally, the challenge of detecting deepfakes underlines the critical need for substantial investment in the research and development of technologies capable of identifying and mitigating their effects. The jurisdictional complexities introduced by the internet's global reach further complicate regulatory efforts, necessitating international cooperation to address the cross-border nature of digital crimes effectively. As P12, law lecturer, solicitor and barrister, points out, "Achieving effective regulation becomes more challenging without a universal agreement," underscoring the need for collaborative international frameworks.

The swift pace of technological advancement in deepfake creation poses another hurdle for legal systems, which traditionally respond more slowly. For example, generative adversarial networks (GANs) are causing concerns about privacy and trust among online users due to their ability to produce extremely realistic deepfakes(Alanazi & Asif 2023). Moreover, ethical considerations around accountability, especially for those unknowingly sharing deepfake content, demand a balanced and fair approach in legal responses. Public resistance and the damage to credibility caused by misinformation necessitate a dual approach of legal measures and public education to enhance critical media literacy.

Enforcing legislation against the misuse of deepfake technology requires bridging the gap in understanding between technical capabilities and legal frameworks, fostering international collaboration, staying ahead of technological advancements, and addressing ethical and public perception challenges. Collaborative efforts among lawmakers,

technical experts, and the international community are crucial for crafting effective, responsive measures against the escalating challenge of deepfakes.

4.2.3 Long-term consequences

The advent and normalization of deepfake technology present multifaceted long-term consequences that affect various aspects of societal, psychological, and political life. At the heart of these concerns is the technology's capacity to spread misinformation and damage reputations. Participants consistently underscored the significant role of deepfakes in amplifying geopolitical conflicts and influencing democratic processes. For example, deepfakes were cited as exacerbating tensions during the Ukrainian and Russian war, where their capacity to manipulate public perception fueled specific political agendas. This aligns with P7's observation that "deepfakes can manipulate public perception, fueling political agendas and undermining democratic processes."

The rapid dissemination of fabricated content on social media platforms was highlighted as causing irreversible damage, particularly in cases where governments and politicians use deepfakes to manipulate public opinion or discredit opponents. This misuse represents a direct threat to the integrity of democratic processes and the authenticity of political discourse. Participants also noted how deepfakes reinforce societal divisions by enhancing confirmation bias and influencing public opinion, further contributing to societal unrest. The role of deepfakes in democratic processes, such as their potential to be used for electoral manipulation in the US elections, underscores the risk of reputational harm as opponents leverage fake videos to tarnish each other's image (ASHISH JAIMAN 2020).

The psychological repercussions of deepfakes are equally troubling. Participants identified emotional distress caused by targeted fake content, the exacerbation of mental health issues like body dysmorphia, and the unique vulnerabilities of certain populations, such as teenagers. This was particularly emphasized in discussions about how the digital environment perpetuates unrealistic ideals and deepens societal anxieties. These insights point to an underlying concern about the erosion of foundational trust in media, government, and interpersonal relationships, compounded by the widespread acceptance of deepfakes as part of the digital landscape.

Participants further highlighted the challenges deepfakes pose to education and research. The difficulty of teaching critical media literacy in an era where distinguishing real from fabricated content becomes increasingly complex was underscored. This challenge emphasizes the necessity for a reevaluation of how educational systems prepare individuals to critically assess digital information.

As illustrated in Fig. 2, ten key insights from our participants shed light on the broader consequences of deepfakes. These include their role in exacerbating misinformation, eroding societal trust, and influencing political processes. Addressing these challenges demands a concerted effort that spans legal, technological, and educational interventions. The urgency to counteract the normalization of deepfakes and restore eroded trust underscores the critical need for a comprehensive approach that integrates policymaking, technological advancements, and societal engagement to safeguard the pillars of truth and authenticity in the digital age.

4.2.4 The role of social and cultural factors in deepfake acceptance

This section delves into how societal norms and cultural values shape the acceptance or rejection of deepfake technology. Legal and policy considerations play a crucial role, as observed by expert 3, who noted a potential trend toward 'tired or passive acceptance' of deepfakes, moderated by legislative actions similar to those seen in privacy laws concerning photography. This observation suggests a form of resigned acceptance that is heavily influenced by the existing legal frameworks.

Participants highlighted significant regional differences in how societies perceive and respond to deepfakes. While some cultures exhibit passive acceptance, viewing deepfakes as innovative tools, others implement stricter policies against their misuse. "Cultural background and societal attitudes shape how people respond to deepfakes," remarked P8, emphasizing the role of societal perception and cultural dependencies in shaping attitudes toward the technology. For instance, some cultures perceive deepfake technology as a reflection of technological advancement, while others view it as a significant ethical and societal threat.

The influence of factors such as geographical location, personality traits, and social media engagement was also discussed extensively. Participants remarked that these factors contribute to varying societal attitudes towards technology, authenticity, and truth, leading to diverse receptions of deepfake technology across different cultural contexts. Such disparities underline the profound impact that cultural values and social norms have on shaping perceptions of deepfakes on a global scale.

The broader societal perception of deepfakes, as discussed by participants 2, 3, 5, 6, 7, 10, 11, 12, 13, and 14, is also of paramount importance. This encompasses a range of factors, from the general acceptance and understanding of technology to the extent to which deepfake content aligns with a society's prevailing beliefs, values, and experiences. Although the direct influence on individual attitudes may appear marginal, broader societal influences—particularly educational initiatives aimed at fostering critical thinking—are instrumental in shaping perceptions of deepfakes. For example, educating children on technology and promoting

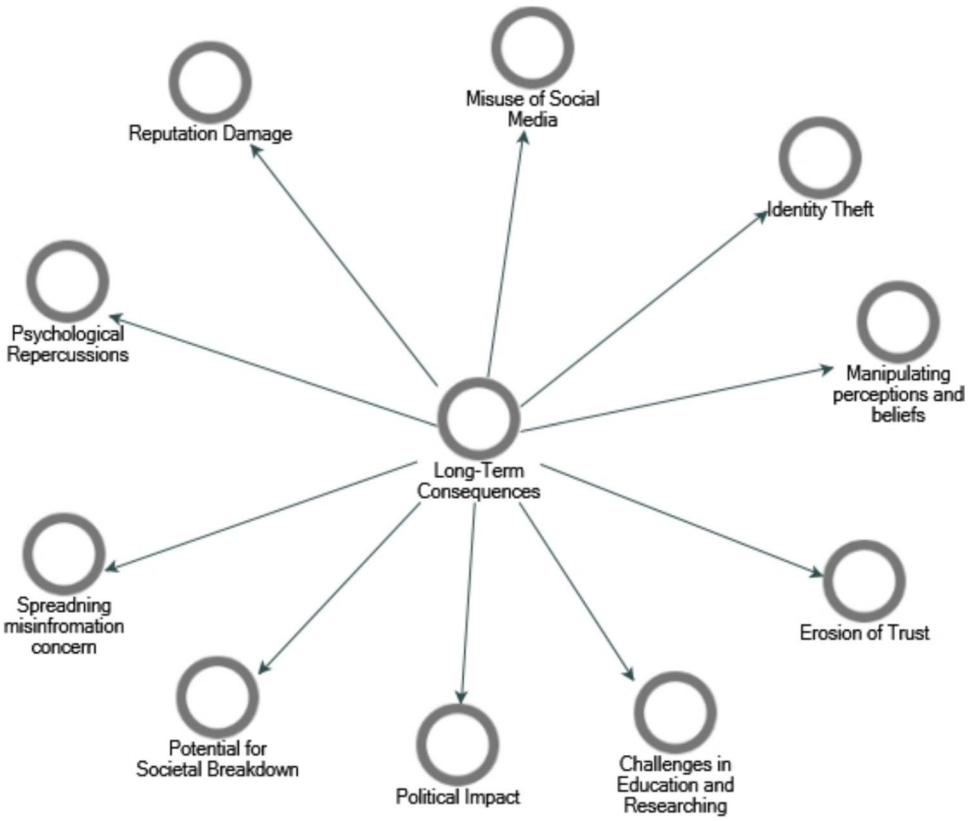


Fig. 2 Participant insights on the societal impact of deepfake normalization

critical assessment skills can help them better understand the difference between real and fabricated content, mitigating the negative effects of deepfakes in both physical and digital environments.

Acceptance or rejection of deepfake technology is not a uniform response but rather a complex interplay of legal, cultural, and societal factors. These elements collectively create a nuanced landscape where acceptance, resistance, and passive acquiescence to deepfakes vary significantly across cultural contexts and societal norms. The evolving nature of technology and media literacy further underscores the need for a tailored, culturally sensitive approach to addressing the challenges posed by deepfakes.

4.2.5 Identifying deepfake vulnerabilities across populations

Groups most vulnerable to the adverse effects of deepfake technology, such as women targeted in non-consensual pornography, political figures, individuals unable to recognize AI-generated content, minorities, the elderly, celebrities, and businesses, are identified in this theme, which indicates their heightened risk of emotional, psychological, and reputational damage due to the multifaceted impact of deepfakes. Notably, non-consensual pornography and abusive applications are the most widespread and harmful uses of deepfake technology, inflicting severe personal harm (Ajder et al. 2019a, b).

To illustrate the extent of these impacts, consider the following examples: Businesses face significant financial and reputational risks, as demonstrated by a recent deepfake scam targeting a prominent CEO(BBC 2023). Another instance involves a deepfake video of the popular YouTube personality MrBeast, used in a fraudulent investment scheme (Gerken 2023). A particularly troubling case occurred in Hong Kong, where scammers utilized deepfake technology to impersonate senior executives of a multinational company, deceiving an employee into transferring US$25.6 million into fraudulent accounts. This elaborate fraud involved sending a phishing email to a finance employee, followed by a series of video calls featuring convincingly deepfaked executives (McKay 2024). Such incidents reveal the urgent need for businesses to take a proactive approach to cybersecurity and prioritize comprehensive employee training to recognize and respond to these deceptive tactics.

Participants call attention to the profound psychological and societal consequences of deepfakes. They discuss the distressing potential for deepfakes to cause emotional harm and stigmatization, especially among younger generations who may experience identity crises or even suicidal tendencies due to manipulated identities (P1).

The discussion also extends to financial security concerns, where deepfakes pose a threat to individuals' life savings through identity theft. This illustrates the wide-ranging harms deepfakes can cause, affecting both personal and financial well-being (P11).

Examples of deepfake fraud targeting individuals and businesses further demonstrate the mechanisms of impact, showing the technology's ability to deceive and manipulate public opinion (P7, P12). Such deception not only creates an environment conducive to fraud but also amplifies existing societal vulnerabilities, such as body dysmorphia among teenagers and young adults who struggle with unrealistic ideals perpetuated online (P13).

Table 2, presented below, categorizes these vulnerable groups alongside specific examples of how they can be targeted by deepfake technology. This table serves as a valuable resource for understanding the direct and indirect risks various populations may face due to deepfake-related threats. It illustrates the importance of targeted awareness and protective strategies to mitigate these vulnerabilities. Analyzing the table helps identify the unique challenges each group encounters and highlights the necessity for comprehensive measures to safeguard against exploitation and harm.

These examples underscore the critical need for targeted awareness and protective strategies to address the diverse risks posed by deepfakes to these vulnerable groups. Implementing comprehensive measures to mitigate these vulnerabilities is essential for safeguarding individuals and organizations from exploitation and harm.

4.2.6 Comprehensive deepfake mitigation and responsible use strategies

The swift development and widespread increase of deepfake technology call for a comprehensive strategy for risk reduction, focusing on the integration of detection technologies, legal frameworks, and educational efforts to promote responsible use and minimize harm. Participants proposed several approaches for addressing deepfake risks, focusing on the critical need for advanced detection algorithms, AI-driven verification tools, and community-based educational initiatives to combat malicious uses of deepfakes.

Expert opinions from diverse sectors underscore the importance of combining technological solutions, legal policies, and public education to effectively address these challenges. For instance, P11 proposed a tripartite strategy: "Develop better tools for detection, enforce legal actions to establish boundaries, and educate the public to recognize deepfakes." This insight spotlights the need for a balanced approach that integrates technical, legal, and societal measures.

Contributors indicate the vital importance of sophisticated detection technologies, particularly AI and machine learning, for the identification of deepfakes. The implementation of such advanced technologies is regarded as a fundamental defense against the proliferation of disinformation and malevolent content, acting as a technical conduit to a resolution. However, technology alone is not sufficient; accessible reporting mechanisms are also vital, creating platforms for the public to report suspected deepfakes and enhancing community involvement in detection efforts. This dual approach underlines the synergy between technology and societal action, aiming to protect potential victims through a combination of rules, regulations, and technological solutions.

Discussions among specialists bring to light the crucial need for definitive punitive measures for the creation and dissemination of malicious deepfakes, especially concerning

Table 2 Vulnerabilities and targeting mechanisms of populations affected by deepfake technology

Vulnerable group	How they can be targeted
Women targeted in pornography	Used in non-consensual deepfake pornography, leading to psychological trauma and social stigma (P11, P8, P3)
Political figures	Deepfakes can fabricate statements or actions to tarnish reputations or influence political outcomes. Furthermore, the spread of misinformation through deepfakes can lead to political instability, social unrest, or distrust in institutions, significantly disrupting the political and social spheres (P11, P7)
People lacking knowledge in AI	Vulnerable due to limited awareness and knowledge in identifying manipulated content, leading to the acceptance of misinformation and fraud (P4, P5)
Minorities and younger generations	Deepfakes exacerbate identity crises and suicidal attempts, with voice simulators and personal data used for fraud. This particularly threatens teenagers and young adults by fostering body dysmorphia and depression due to unrealistic online ideals (P1, P13)
Celebrities	Targeted for deepfake revenge porn and other manipulations, often disproportionately affecting women celebrities, leading to reputational damage and emotional distress (P11)
Businesses	Financial fraud and reputation damage through deepfake impersonations of executives or misleading information (P7)
The elderly	Elderly individuals are more susceptible to scams and manipulative tactics due to factors like trust, vulnerability, and less familiarity with emerging technologies. Their potential lack of awareness about deepfakes' nature can lead to direct and severe manipulation (P14, P5, P8)

crimes like deepfake revenge pornography. Furthermore, there is a call for international unity in establishing and enforcing legal standards, recognizing the borderless realm of digital content and the widespread challenges it introduces. Participants emphasized mandatory disclosures when content has been altered and supported comprehensive legal and ethical standards to guide the responsible development of deepfake technology.

Education and digital literacy campaigns are pivotal in combating the influence of deepfakes. Incorporating lessons on digital media verification and the implications of deepfakes into school programs is advocated as a proactive measure to prepare younger generations for the digital landscape they inhabit. Moreover, public initiatives aimed at improving the general population's understanding of digital content's authenticity and the technology behind deepfakes are deemed necessary for fostering a discerning and critical audience. For instance, P11 remarked that public awareness campaigns should focus on both recognizing and mitigating the risks associated with deepfakes. P8 highlighted the importance of 'modernizing the school curriculum' to include elements related to technology and digital literacy. These strategies aim to build societal understanding and resilience against deepfakes, equipping individuals with the skills to critically assess and navigate digital content.

Further insights underscore the necessity of peer-to-peer knowledge sharing and collaboration with regulatory bodies to strengthen media literacy programs. Innovative approaches, such as "evidence cafes," facilitate inclusive dialogue, enabling communities to engage with the ethical and legal aspects of deepfake technology. Additionally, developing practical tools for verifying authenticity empowers individuals to independently assess digital content.

Experts accentuate the joint effort from society and organizations at the individual level to foster training and detection capabilities, creating a world where critical awareness becomes a routine part of digital interaction. This approach is likened to anti-smoking campaigns, reinforcing the need for a similar societal commitment to combating deepfake misuse.

The insights provided by participants elucidate the complex nature of deepfake technology and its varied societal implications. The need for a comprehensive strategy that integrates technological, legal, and educational measures is evident. Such a strategy requires cross-disciplinary collaboration, drawing on the expertise of software developers, psychologists, legal professionals, and the creative community to navigate the ethical challenges posed by deepfakes.

Furthermore, the discussion highlights the importance of societal efforts in raising awareness, conducting information campaigns, and promoting media literacy to empower individuals against the potential harms of deepfakes. By fostering a culture of critical thinking and skepticism, society can more effectively navigate the challenges presented by deepfakes, thereby minimizing their negative impact while exploring their potential for positive applications.

Addressing these challenges necessitates a comprehensive and integrated approach, encompassing advanced detection technologies, international collaboration on regulatory frameworks, and educational initiatives aimed at digital literacy. Through collaborative efforts and a steadfast commitment to ethical standards, it is possible to mitigate the risks associated with deepfake technology while harnessing its potential for beneficial uses. This holistic approach is illustrated in Fig. 3, capturing the integral components necessary for responsible deepfake mitigation and use.

4.2.7 Ethical and constructive uses of deepfake technology

This theme delves into the potential for deepfakes to revolutionize various industries while scrutinizing the delicate balance between technological innovation and social responsibility. A nuanced perspective emerges from expert insights, demonstrating how deepfake technology can be harnessed for beneficial purposes while adhering to ethical standards and legal boundaries.

Experts commend the innovative impact of deepfake technology in art and design, where it has been used to reinterpret existing works and transform old music samples into new compositions. However, a recent court ruling presents a significant challenge to the advancement of AI-generated content by affirming that such creations, including those produced using deepfake technology, do not qualify for copyright protection due to the requirement for human authorship (CHO 2023). This legal constraint poses a complex dilemma, while deepfakes offer immense potential to enrich cultural and artistic expression, they also face substantial ethical and legal challenges, particularly in industries like Hollywood, where copyright issues are highly sensitive. The need for responsible use becomes even more pronounced as the creative sector navigates the intricate intersection of innovation and intellectual property rights, suggesting a cautious yet potentially groundbreaking integration of AI into creative practices.

Further illustrating the broad applicability of deepfake technology, the Deep Nostalgia app is noted for its potential therapeutic benefits, particularly for individuals with dementia. By animating lifelike images of family members or familiar settings, deepfakes have the capacity to alleviate anxiety and enhance the quality of life for those struggling with memory-related issues (P13). This example underscores the technology's potential as a therapeutic tool in addressing complex emotional and psychological challenges.

The entertainment industry, in particular, stands to benefit from the ethical application of deepfakes, as evidenced by a notable commercial featuring a digitally recreated Audrey

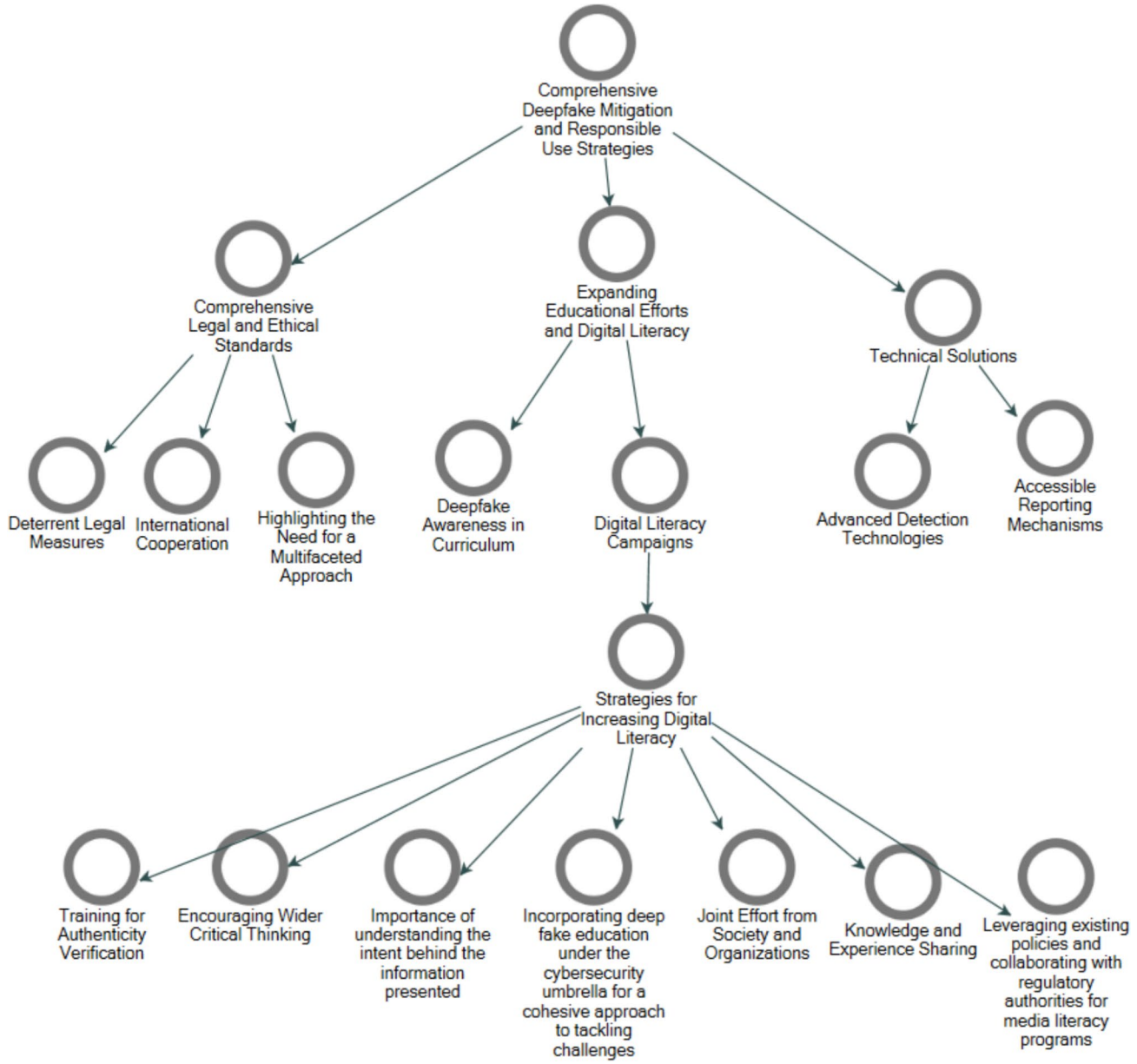


Fig. 3 Deepfake mitigation and responsible use strategies

Hepburn. Such applications not only revive beloved characters and recreate cherished scenes but also offer significant marketing and promotional value for brands (P14). This capacity to breathe new life into fictional characters or resurrect iconic figures emphasizes the value of deepfake technology in enriching entertainment experiences and captivating audiences.

In educational and creative contexts, deepfakes present opportunities to make learning more engaging and immersive. Whether used in art galleries to animate historical figures or in educational materials to create interactive experiences, the technology's potential to enhance engagement and understanding is clear (P6). This emphasizes the need for transparency and ethical awareness, ensuring that these manipulations are identified as artificial constructs rather than misleading representations.

Despite the promising outlook, experts caution that the ethical use of deepfakes hinges on the creators' intentions and the clarity with which manipulated content is presented. It is crucial to distinguish between creative expression and deceptive intent, as ethical applications require clear communication to avoid misleading audiences (P3, P5).

The ethical and constructive use of deepfake technology presents substantial opportunities for innovation in art,

design, entertainment, and education. By employing deepfakes with a focus on social responsibility and transparent communication, these applications can enhance cultural experiences, offer therapeutic benefits, and enrich learning environments. Expert insights underscore the importance of carefully considering the ethical implications, ensuring that deepfake technology is utilized in ways that comply with legal standards and contribute positively to society.

5 Discussion

5.1 Comparative analysis with existing literature

The study's findings align closely with prior research, reinforcing key challenges posed by deepfake technology. For instance, Güera and Delp (n.d.) stressed the limitations of traditional detection methods, mirroring participants' emphasis on the need for advanced AI-driven solutions. Similarly, Ajder et al. (2019b) documented the psychological harm caused by non-consensual pornography, particularly among women, underscoring the urgency of protective measures—a concern also raised in this study.

Participants' concerns about the political misuse of deepfakes reflect Breen's (2021) findings on fabricated content influencing elections and geopolitical stability. Furthermore, the influence of cultural norms on the perception of deepfake technology aligns with Herring et al. (2022), who identified societal attitudes as a critical factor in technology adoption.

5.2 Policy implications and ethical considerations

This study demonstrates the significance of establishing globally coordinated regulatory frameworks, reflecting Kasper and Laurits's (2016) argument for standardized criteria in admitting digital evidence within legal systems. Participants also advocated for interdisciplinary collaboration, aligning with the broader consensus that tackling the challenges posed by deepfakes demands a comprehensive approach combining technical expertise, legal insights, and educational initiatives.

5.3 Insights and implications

Comparative analysis reveals significant variation in regulatory approaches across jurisdictions, shaped by differing cultural, political, and technological priorities. Countries like China and South Korea prioritize stringent enforcement, while the EU and Canada integrate innovation and public awareness (Lawson 2023). In contrast, the USA employs a fragmented approach, with state-specific laws addressing deepfake misuse.

This diversity underscores the critical need for globally coordinated frameworks to streamline detection, regulation, and enforcement. Such collaboration would enhance legal recourse, strengthen accountability, and address the transnational nature of deepfake crimes. Future research should explore the effectiveness of these models in fostering international cooperation and technological innovation.

5.4 Comparative analysis and policy recommendations

The regulatory landscape for deepfake technology reflects diverse strategies globally, providing valuable insights into addressing its challenges.

5.4.1 Key international approaches

- China: Comprehensive laws mandate disclaimers on AI-generated content, user consent, and identity verification, enforced by the Cyberspace Administration of China (CAC). This holistic approach ensures accountability throughout the lifecycle of deepfake technology (CAC 2022).
- European Union (EU): The AI Act and Digital Services Act emphasize transparency and platform accountability, penalizing failures to address harmful content. The Code of Practice on Disinformation fosters collaboration between governments and platforms to counter misuse (EC 2024).
- Canada: A multi-faceted approach combines public awareness campaigns, investments in detection technologies, and targeted legislation to address malicious uses, such as non-consensual pornography (Lawson 2023).
- USA: A fragmented system includes state laws tackling specific issues, such as electoral manipulation and revenge pornography. Federal initiatives like the DEEP FAKES Accountability Act aim to establish national standards but face challenges due to decentralized governance (Congress.gov 2019; Kan 2024; Yousif 2024).

5.5 Findings and recommendations

Building on international efforts and the study's findings, this research proposes a multi-tiered regulatory framework guided by the Responsible Innovation Framework:

1. Strengthening detection technologies

   Governments and industries should invest in cutting-edge AI and machine learning tools for deepfake detection. Collaborations with academia can foster innovation in this area.

2. International cooperation

   Unified global standards are essential for addressing the transnational nature of deepfake misuse. International organizations should lead initiatives to harmonize enforcement and streamline regulations.
3. Educational initiatives

   Comprehensive media literacy programs should empower the public to recognize and critically evaluate deepfake content, mitigating misinformation and enhancing digital resilience.
4. Ethical AI certification

   Establishing certification programs for companies adhering to ethical standards would promote responsible innovation and ensure transparency in deepfake creation and use.
5. Transparency and accountability

   Mandating the disclosure of altered content and enforcing clear legal consequences for malicious uses are critical to fostering public trust and maintaining accountability in digital media.

   As depicted in Fig. 4, the workflow visually represents the research process, connecting key stages, findings, and recommendations. It underscores the importance of addressing deepfake technology's societal impact through comprehensive, multi-pronged strategies.

6 Conclusion

Deepfake technology poses significant challenges, threatening societal trust, democratic processes, and individual privacy. Our research spotlights the urgent need for comprehensive strategies to address these risks while fostering ethical innovation. By drawing on international regulatory frameworks, such as the AI Act, and insights from expert interviews, we propose actionable recommendations to navigate the complexities of deepfakes effectively.

Key strategies include strengthening detection technologies through investment in AI-driven solutions and promoting transparency by mandating the disclosure of altered content. These measures should be complemented by interdisciplinary task forces integrating expertise from law, ethics, cybersecurity, and AI to craft adaptable and forward-looking regulations. International collaboration, engaging organizations like the United Nations and the International Telecommunication Union, is critical for harmonizing

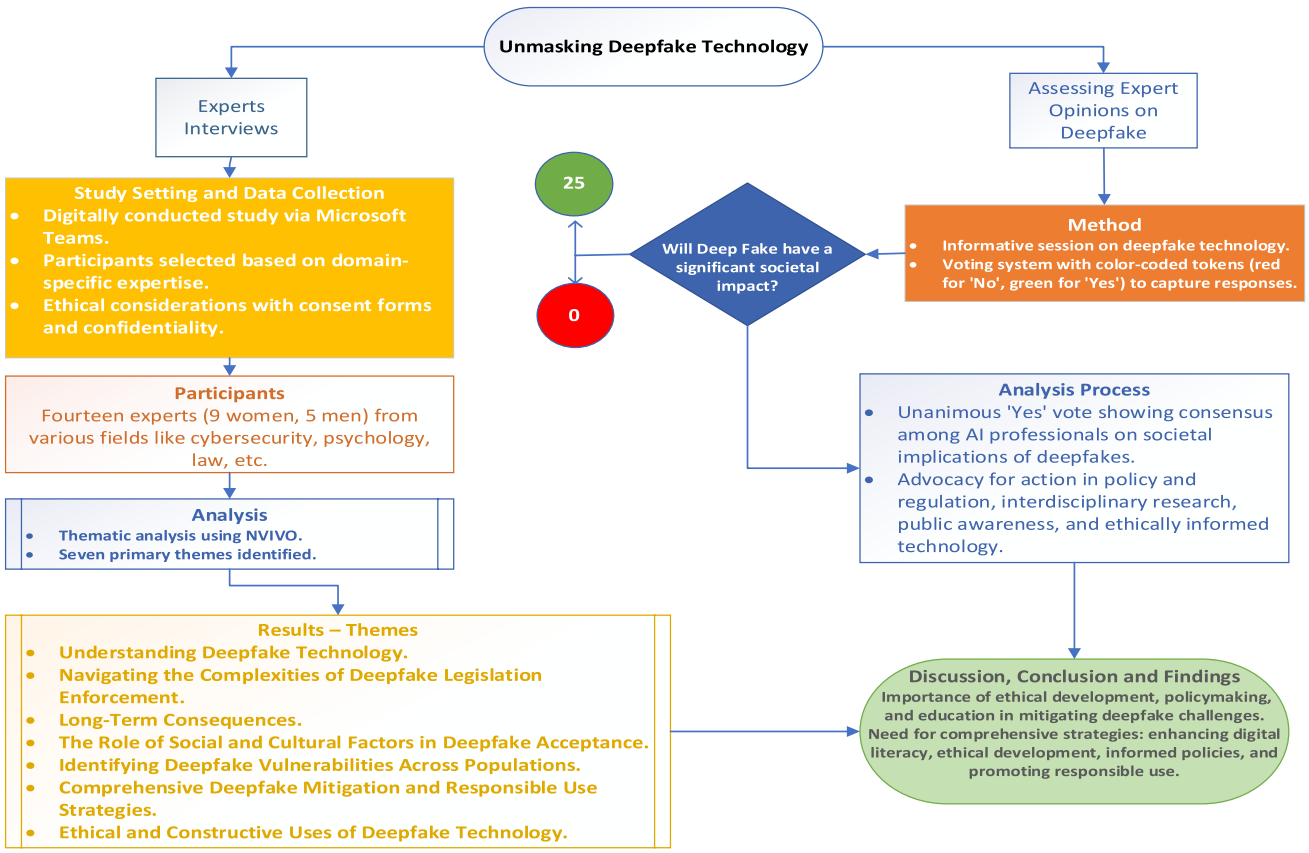


Fig. 4 Unmasking deepfake technology studies workflow

standards and addressing cross-border challenges associated with deepfakes.

Educational initiatives play a pivotal role in building public resilience against deepfake misuse. Partnerships with schools, media organizations, and technology platforms can enhance digital literacy, empowering individuals to critically evaluate digital content. Additionally, an ethical AI certification program for companies adhering to best practices would promote responsible innovation and build public trust in AI technologies.

Our findings emphasize the importance of a multi-tiered regulatory framework, incorporating risk-based assessments, stringent standards for high-risk applications, and proactive policy-making that aligns with technological advancements. This framework should focus on privacy, consent, and transparency, ensuring regulations remain adaptable to the rapid evolution of deepfake capabilities.

Future research should explore the long-term societal impacts of deepfake technology and assess the effectiveness of regulatory and educational interventions. Empirical studies on public perception and the global coordination of regulatory frameworks are essential to inform policy and maintain public trust.

By integrating robust regulatory measures, fostering global collaboration, and enhancing public education, stakeholders can mitigate the risks of deepfake technology. This comprehensive approach ensures a future where the benefits of deepfakes are maximized ethically, safeguarding societal values and individual rights while minimizing harm.

Appendix 1. Ethical consideration



Participant Information Sheet

DEEPFAKE LAW EXPERT INTERVIEW

Purpose of Research

This study is part of the research project which deals with studying the limitation and detection of DeepFake technology, its societal, ethical and legal implications.

Will the interview be recorded?

The interview will be conducted via MS Teams and will be audio-recorded using a digital recorder. That recording is then transcribed so that it can be analysed.

Do I have to take part?

No. It is completely up to you. You are free to stop taking part at any time during the research without giving a reason. If you decide to take part, you do not have to answer all of the questions or contribute to all of the topics we talk about. You can withdraw your data for a period of up to 10 days after the interview, as after this time it will not be possible to identify your individual data from the aggregated anonymised results. If you are uncertain or uncomfortable about taking part in the one to one interview, you can email the researcher. You can ask her any question you like**.**

What are the possible benefits of taking part?

We hope that with your help and that of other participants, we will get an understanding of how we can review the existing legal and ethical considerations surrounding deep fake technology and what we can propose to the policymakers and other stakeholder on how to mitigate the potential risk of deep fake technology. Also, we hope that you will find the experience fun and interesting.

How will my data be used?

Any information you supply will be treated confidentially in accordance with the *General Data Protection Regulation* (EU *GDPR*) and the UK Data Protection Act 1998 to protect your privacy. Anonymised transcripts will be created by removing or replacing identifiers such as name, age, and location. Under no circumstances will *identifiable* responses be provided to any third party. Your interview recording and transcript will be securely stored on a password-protected hard drive at Cranfield University and will be deleted as soon as they are no longer required. We will not use your name when we write up the results, but we would like to use some quotes of what you say. Findings of the study might be published in academic journal articles, presented at academic conferences, at university seminars, or on social media.

Questions About the Research?

If you have questions about the research in general or about your role in the study, please feel free to contact Sami Alanazi at sami.alanazi@cranfield.ac.uk or Dr Seemal Asif at s.asif@cranfield.ac.uk.

Thank you for reading this information sheet.

If you would like to be involved, please complete the attached consent form

Consent

- I have read and understood the information sheet.
- I confirm that the purpose of the study has been explained to me and that I have had the opportunity to ask questions about the research and have had these answered satisfactorily.
- I understand that my participation is voluntary, and that I am free to withdraw at any time without giving any reason.
- I give permission for the researcher to use any photograph(s) and/or videotape(s) made as part of this research project in presentations, publications, for educational uses, or through websites and social media.
- I understand that images on websites can be viewed throughout the world and that some countries may not provide the same level of protection to the rights of individuals as EU/UK legislation provides.
- I agree that my anonymised research data may be used by others for future research.
- I understand that personal data will be always kept confidential.
- I agree to take part in the study.

Name of respondent	Date	Signature
Name of researcher	Date	Signature

Appendix 2. Interview briefing

Interview questions:

1. Are you familiar with deepfake technology? How would you describe it?

2. How would you evaluate the impact of the Deep Fake technology?

3. Do you think Deep fake technology's acceptance and/or rejection can be shaped by social norms and cultural values?

4. Are there psychological and emotional impacts of consuming or being targeted by deepfake content?
    a. Can you think of specific example in this context?

5. Can deepfake technology exacerbate issues of disinformation and manipulation?
    a. What measures can be taken to mitigate these risks?

6. Are there populations in society that are particularly affected by deepfake technology?
    a. Is there anything that could be done about this?

7. Are there ethical guidelines or regulations that can effectively govern the creation and dissemination of deepfake content?
    a. What challenges exist in enforcing such measures?

8. Can interdisciplinary collaboration between technology experts, ethicists, psychologists, and social scientists contribute to a better understanding of the social impacts of deepfake technology?
    a. How?

9. What are the potential long-term consequences of the normalization of deepfakes on human perception, interpersonal relationships, and societal trust?

10. Do you think deepfake technology can be harnessed to promote responsible or positive use?
    a. How?

11. Can education and media literacy programs help individuals recognize and critically evaluate deepfake content,

a. How?
b. What strategies can be employed to increase digital media literacy in the general population?

Invitation Email

Dear .....,

I hope this email finds you well.

My name is Sami Alanazi, and I am a PhD researcher in Artificial Intelligence at Cranfield University. My research focuses on deepfake image and video detection, particularly its societal implications.

As part of my research, I am seeking to interview experts in law and media regarding deepfake technology. I would be most grateful if you could spare 30-60 minutes for an online interview. During this session, I will provide a brief overview (lasting approximately 10 minutes) of deepfake technology, with a focus on its creation and detection, its social implications, and the existing legal framework governing it. The primary objective is to gather insights from legal professionals about the current laws related to deepfake technology and to discuss the necessity for developing more comprehensive legislation to mitigate its potential risks.

Your expertise and perspective would be highly valuable to my research, and I look forward to the possibility of speaking with you.

Thank you for considering my request.

Best wishes,
Sami Alanazi

Appendix 3. Themes

Themetic framework

Name	Files	References
Understanding Deepfake Technology	13	30
The Role of Social and Cultural Factors in Deepfake Acceptance	12	22
Navigating the Complexities of Deepfake Legislation Enforcement	9	17
Long-Term Consequences	14	75
Identifying Deepfake Vulnerabilities Across Populations	12	43
Ethical and Constructive Uses of Deepfake Technology	7	14
Comprehensive Deepfake Mitigation and Responsible Use Strategies	14	105

Name	Files	References
Understanding Deepfake Technology	13	30
Voice Manipulation	3	3
Targeting Specific Groups	2	2
Recent Familiarity	5	5
Image generated models	4	4
Fake Videos	5	5
Extension of traditional tools for manipulating images	3	3
Creating fake documents	2	2
Challenges in Distinguishing Genuine and Fabricated Content	5	6

The Role of Social and Cultural Factors in Deepfake Acceptance	12	22
Societal Perception	12	15
Influence of Pre existing Beliefs and Values	4	4
Familiarity with Technology Dependence	4	4
Attitude Towards Technology in General	7	7
Geographical Location and Cultural Factors Dependence	5	5
Country Legislation Against Unconsented Use of Images	1	1

Navigating the Complexities of Deepfake Legislation Enforcement	9	17
Variety of Applicable Laws	3	4
Requirement for Dedicated Resources and Expertise	2	2
Challenging Source Identification	4	5
Challenges in understanding deepfake technology	2	3
Challenges in Court Evidence	3	3

Long-Term Consequences	14	75
Spreadning misinfromation concern	8	10
Reputation Damage	3	3
Psychological Repercussions	6	7
Potential for Societal Breakdown	3	4
Political Impact	9	14
Misuse of Social Media	7	9
Manipulating perceptions and beliefs	9	10
Identity Theft	6	6
Erosion of Trust	7	10
Challenges in Education and Researching	2	2

Ethical and Constructive Uses of Deepfake Technology	7	14
Innovative and Artistic Applications of Deepfake Technology	5	7
Entertainment	4	6
Educational	1	1

Identifying Deepfake Vulnerabilities Across Populations	12	43
Vulnerable Populations	12	23
Targeting Women in Pornography	3	3
Politicals	2	2
People who lack the knowledge to identify AI generated content	4	4
Minorities	4	4
Individuals belonging to a specific group	1	1
Every one	7	8
Elderly	3	3
Celebrities	1	1
Businesses	1	1
Mechanisms of Impact	12	20
Targeting Women in Pornography	2	3
Politicals	1	1
People who lack the knowledge to identify AI generated content	2	2
Minorities	2	2
Individuals belonging to a specific group	1	1
Every one	9	13
Elderly	2	2
Celebrities	1	1
Businesses	1	1

Comprehensive Deepfake Mitigation and Responsible Use Strategies	14	105
Technical Solutions	5	6
Advanced Detection Technologies	4	4
Accessible Reporting Mechanisms	1	2
Expanding Educational Efforts and Digital Literacy	14	69
Digital Literacy Campaigns	14	48
Strategies for Increasing Digital Literacy	14	48
Training for Authenticity Verification	2	2
Leveraging existing policies and collaborating with regulatory authorities	2	2
Knowledge and Experience Sharing	5	5
Joint Effort from Society and Organizations	2	2
Incorporating deep fake education under the cybersecurity umbrella for	2	2
Importance of understanding the intent behind the information presente	4	6
Encouraging Wider Critical Thinking	12	29
Deepfake Awareness in Curriculum	11	21
Comprehensive Legal and Ethical Standards	10	30
International Cooperation	4	4
Highlighting the Need for a Multifaceted Approach	7	10
Deterrent Legal Measures	4	15

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019a). Graphic design: Eleanor Winter Cover image: Joel Filipe Cite: The state of deepfakes: Landscape, threats, and impact about deeptrace.

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019b). Graphic design: Eleanor Winter Cover image: Joel Filipe Cite: The state of deepfakes: Landscape, threats, and impact about deeptrace.

Alanazi, S., & Asif, S. (2023). Understanding deepfakes: A comprehensive analysis of creation, generation, and detection. https://doi.org/10.54941/ahfe1003290

Alanazi S, Asif S (2024) Exploring deepfake technology: Creation, consequences and countermeasures. Human-Intelligent Systems Integration. https://doi.org/10.1007/s42454-024-00054-8

Alanazi, S., Asif, S., & Moulitsas, I. (2024). Examining the societal impact and legislative requirements of deepfake technology: A comprehensive study. IJSSH. https://doi.org/10.18178/ijssh.2024.14.2.1194

ASHISH JAIMAN. (2020). Debating the ethics of deepfakes. ORF .

BBC. (2023, July 7). Martin Lewis felt "sick" seeing deepfake scam ad on Facebook. BBC. https://www.bbc.co.uk/news/uk-66130785

Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing researcher. In International Journal of Transgender Health (Vol. 24, Issue 1, pp. 1–6). Routledge. https://doi.org/10.1080/26895269.2022.2129597

Breen, D. (2021). Silent no more: How deepfakes will force courts to reconsider video admission standards. Journal of High Technology Law. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/jhtl21§ion=6

CAC. (2022, December 12). The Cyberspace Administration of China and other three departments issued the "Provisions on the Administration of Deep Synthesis of Internet Information Services." Cyberspace Administration of China. http://www.cac.gov.cn/2022-12/11/c_1672221949318230.htm

Callari TC, Vecellio Segate R, Hubbard E-M, Daly A, Lohse N (2024) An ethical framework for human-robot collaboration for the future people-centric manufacturing: A collaborative endeavour with European subject-matter experts in ethics. Technol Soc 78:102680. https://doi.org/10.1016/j.techsoc.2024.102680

Chesney B, Citron D (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. Calif Law Rev 107(6), 1753–1820. https://doi.org/10.15779/Z38RV0D15J

CHO, W. (2023, August 18). AI-created art isn't copyrightable, judge says in ruling that could give hollywood studios pause. https://www.hollywoodreporter.com/business/business-news/ai-works-not-copyrightable-studios-1235570316/

Congress.gov. (2019). H.R.3230 - DEEP FAKES Accountability Act: 116th Congress. https://www.congress.gov/bill/116th-congress/house-bill/3230

Cover, R. (2022). Deepfake culture: The emergence of audio-video deception as an object of social anxiety and regulation. Journal of Media & Cultural Studies, 36(4). https://doi.org/10.1080/10304312.2022.2084039

EC. (2024). AI Act. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Flitter, E., & Cowley, S. (2023, August 30). Voice deepfakes are coming for your bank balance. The New York Times. https://www.nytimes.com/2023/08/30/business/voice-deepfakes-bank-scams.html

Gerken, T. (2023, October 4). MrBeast and BBC stars used in deepfake scam videos. BBC. https://www.bbc.co.uk/news/technology-66993651

Güera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks; Deepfake video detection using recurrent neural networks.

Herring, S. C., Dedema, M., Rodriguez, E., & Yang, L. (2022, November 25). Gender and culture differences in perception of deceptive video filter use. HCI International 2022 - Late Breaking Papers. Interaction in New Media, Learning and Games.

Hsu, C. C., Zhuang, Y. X., & Lee, C. Y. (2020). Deep fake image detection based on pairwise learning. Applied Sciences (Switzerland), 10(1). https://doi.org/10.3390/app10010370

Jackson, K., & Bazeley, P. (2019). Qualitative data analysis with NVivo (3rd ed.). SAGE Publications Ltd.

Kan, M. (2024, March 8). Biden calls for a ban on AI voice impersonation. https://uk.pcmag.com/ai/151364/biden-calls-for-a-ban-on-ai-voice-impersonation

Kasper, A., & Laurits, E. (2016). Challenges in collecting digital evidence: A legal perspective.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? In Business horizons (Vol. 63, Issue 2, pp. 135–146). Elsevier Ltd. https://doi.org/10.1016/j.bushor.2019.11.006

King, N., & Horrocks, C. (2010). Interviews in qualitative research. SAGE Publications Ltd.

Lawson, A. (2023, April 24). A look at global deepfake regulation approaches. Responsible AI Institute. https://www.responsible.ai/a-look-at-global-deepfake-regulation-approaches/

Levitt HM, Bamberg M, Creswell JW, Frost DM, Josselson R, Suárez-Orozco C (2018) Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA publications and communications board task force report. Am Psychol 73(1):26–46. https://doi.org/10.1037/amp0000151

Loughran, J. (2024, March 14). EU signs law to crack down on 'high risk' AI. Eandt. https://eandt.theiet.org/2024/03/14/eu-signs-legislation-crack-down-high-risk-ai?utm_campaign=E%2BT%20News%20-%20Template%20Redesign%2014%20Mar%20%28Split%20test%29&utm_content=E%26T%20News%20-%20Members&utm_medium=email&utm_source=Adestra&utm_term=3477864

McKay, C. (2024, February 4). Beware: Multi-national company loses HK$200 million in elaborate deepfake scam. The AI Literacy Platform. https://www.maginative.com/article/multi-national-company-loses-hk-200-million-in-elaborate-deepfake-scam/#:~:text=AI%20Safety%20China-,Beware%3A%20Multi%2DNational%20Company%20Loses%20HK%24200%20Million%20in%20Elaborate,for%20a%20confidential%20corporate%20transaction.

Mirsky, Y., & Lee, W. (2020). Creation and detection of deepfakes: A survey. In ACM Comput. Surv. 1, 1, Article (Vol. 1). www.forbes.com/sites/forbestechcouncil/2019/05/21/gans-and-deepfakes-could-revolutionize-the-fashion-industry/

Navarro Martínez O, Fernández-García D, Cuartero Monteagudo N, Forero-Rincón O (2024) Possible health benefits and risks of deepfake videos: A qualitative study in nursing students. Nursing Reports 14(4):2746–2757. https://doi.org/10.3390/nursrep14040203

North London Coroner's Service. (2022). Regulation 28 report to prevent future deaths. https://www.judiciary.uk/wp-content/uploads/2022/10/Molly-Russell-Prevention-of-future-deaths-report-2022-0315_Published.pdf

Patterson, D. (2023). Deepfakes for good? How synthetic media is transforming business. https://techinformed.com/deepfakes-for-good-how-synthetic-media-is-transforming-business/

Saunders, M. N., LEWIS, P., & THORNHILL, A. (2016). Research methods for business students (7th ed.). Pearson.

Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning lip sync from audio. ACM Transactions on Graphics, 36(4). https://doi.org/10.1145/3072959.3073640

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media and Society, 6(1). https://doi.org/10.1177/2056305120903408

VIEJO, M. (2023, September 18). In Spain, dozens of girls are reporting AI-generated nude photos of them being circulated at school: 'My heart skipped a beat.' https://english.elpais.com/international/2023-09-18/in-spain-dozens-of-girls-are-reporting-ai-generated-nude-photos-of-them-being-circulated-at-school-my-heart-skipped-a-beat.html

Wakefield, J. (2022, March 18). Deepfake presidents used in Russia-Ukraine war. https://www.bbc.co.uk/news/technology-60780142

Welsh, E. (2002). Dealing with data: Using NVivo in the qualitative data analysis process. http://www.qualitative-research.net/fqs/

Yousif, N. (2024, February 8). US FCC makes AI-generated robocalls illegal. https://www.bbc.co.uk/news/world-us-canada-68240887

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



https://dspace.org/

Unmasking deepfakes: a multidisciplinary examination of social impacts and regulatory responses

Alanazi, Sami; Asif, Seemal; Caird-Daley, Antoinette; Moulitsas, Irene

2025-12
Springer
https://doi.org/10.1007/s42454-025-00060-4
Alanazi S, Asif S, Caird-Daley A, Moulitsas I. (2025) Unmasking deepfakes: a multidisciplinary examination of social impacts and regulatory responses. Human-Intelligent Systems Integration, Volume 7, December 2025, pp. 131-153

Downloaded from a DSpace repository. https://dspace.org/