International Journal of Social Science Research (IJSSR)
Volume- 2, Issue- 2 | March - April 2025 ISSN: 3048-9490

Deepfake and the Crisis of Trust in Digital Public Spheres

Vipin Gupta
PhD Scholar, Department of Sociology, Banaras Hindu University, Varanasi
Email: gvipin763@gmail.com

Abstract

The rapid rise of deepfake technology is changing how we experience digital media and raising serious questions about what we can trust online. As synthetic audio and video content becomes more lifelike and easier for anyone to create, it is beginning to influence how people understand truth, credibility, and public communication. This study looks at how deepfakes are fueling a wider crisis of trust by blurring the line between real and fabricated content, spreading misinformation, and increasing public scepticism across digital platforms. Using qualitative methods, particularly social media discourse analysis and case studies of major deepfake incidents, the research explores how people make sense of and respond to synthetic media in their everyday online lives. The findings show that deepfakes create a sense of epistemic uncertainty, weaken trust in both traditional and digital news sources, and shape civic behaviour by encouraging doubt, confusion, and polarisation. Ultimately, the paper argues that this crisis of trust is not just about technology; it is deeply social, emerging from a post-truth environment and fragmented digital publics. By examining the cultural meanings and social consequences of deepfakes, the study adds to ongoing discussions on misinformation, digital ethics, and the future of democratic communication.

Introduction

The rapid advancement of digital technologies has dramatically altered the landscape of communication, knowledge production, and social interaction. Among the most consequential developments within this arena is the emergence of deepfake technology, a sophisticated form of synthetic media generated primarily through machine learning techniques such as Generative Adversarial Networks (GANs) and diffusion-based models. Deepfakes manipulate or fabricate video, audio, and images with a level of precision that frequently renders them indistinguishable from authentic content. In less than a decade, what began as an innovative experiment within niche technological communities has evolved into a pervasive and destabilising global phenomenon. As deepfake technology becomes increasingly accessible, its cultural, political, and ethical implications continue to expand, prompting critical questions about truth, trust, and power in digital societies.

The impetus for studying deepfakes emerges from a growing recognition that digital media is not merely a platform for communication, but a constitutive element of contemporary social reality. The proliferation of synthetic media alters the epistemic foundations upon which individuals and societies form beliefs, engage in civic action, and negotiate shared meaning. Analysts predict that by 2030, the majority of online content may involve some degree of AI-generated manipulation or fabrication. The speed and scale of this technological diffusion present profound sociological challenges, particularly as deepfakes infiltrate political communication, journalism, entertainment, interpersonal relationships, and identity politics.

From deepfake pornography targeting women to fabricated political speeches that influence public opinion, deepfake culture is recalibrating how people perceive reality and how societies determine what can be trusted.

This dramatic shift must be understood within the broader context of the post-truth era, a period characterised by the declining influence of objective facts in public life and the increasing prominence of emotions, opinions, and personal beliefs in shaping collective perceptions. Deepfakes do not merely contribute to misinformation; they symbolise a deeper epistemic rupture. The foundational principle that seeing is believing, once central to the authority of visual media, has been fundamentally undermined. As synthetic media becomes more realistic, individuals face difficulty differentiating truth from fabrication, thereby eroding trust not only in specific pieces of content but also in the systems, institutions, and actors responsible for producing and circulating information.

The digital public sphere, a conceptual evolution of Jürgen Habermas's classic formulation of the public sphere, provides a valuable lens through which to analyse these transformations. Habermas envisioned the public sphere as a domain of rational-critical debate where citizens engage in democratic deliberation. Yet, in the contemporary era, this idealized model is challenged by the fragmentation, algorithmic curation, and emotionalised discourse inherent in digital communication environments. Scholars such as Nancy Fraser have critiqued the universalist assumptions embedded in Habermas's model, arguing instead for the existence of multiple, competing, and sometimes conflicting publics. Meanwhile, Zizi Papacharissi conceptualises digital public spheres as "affective publics," wherein emotions, affect, and personal narratives shape collective action. Deepfake culture accentuates these complexities, introducing an additional layer of uncertainty and mistrust that further destabilises the communicative foundations of the public sphere.

A central component of this destabilisation is the crisis of trust, a phenomenon that has become increasingly pervasive across political, social, and economic domains. Trust is not merely an interpersonal attribute; it is a sociological mechanism that stabilises social relations, facilitates cooperation, and underpins institutional legitimacy. Deepfake culture disrupts this mechanism by introducing epistemic ambiguity and cognitive dissonance on a mass scale. When individuals cannot ascertain the authenticity of digital content, mistrust extends from specific media artefacts to the broader ecosystem of digital communication, including news outlets, political actors, governmental institutions, and even personal acquaintances. In this sense, deepfakes do not simply deceive; they diminish the confidence of individuals in the very possibility of knowing what is true.

The erosion of trust caused by deepfakes can be analysed through several influential theoretical frameworks. Bauman's Liquid Modernity offers a compelling interpretation of deepfake culture by suggesting that contemporary life is characterised by fluidity, uncertainty, and instability. In a world where identities, information flows, and social structures are constantly shifting, deepfakes represent the epitome of liquid modern conditions, where the boundaries between real and fake dissolve. Bauman emphasises that individuals in liquid modern societies often struggle to anchor themselves in stable truths or institutions, resulting in heightened

anxiety and insecurity. Deepfakes exacerbate this condition by making even the most concrete forms of evidence, audio and video, susceptible to manipulation.

Another relevant theoretical perspective comes from Ulrich Beck's Risk Society, which posits that modern societies are increasingly preoccupied with managing manufactured risks created by technological and industrial advancement. Deepfakes epitomise such manufactured risks, as they generate new forms of harm that are difficult to predict, regulate, or control. In Beck's framework, risks are often global, invisible, and distributed unevenly across social groups. Similarly, deepfake harms disproportionately affect women (through non-consensual pornography), political minorities (through targeted manipulation), and individuals with public visibility (such as activists and journalists). Beck's theory thus helps illuminate how deepfakes produce asymmetric vulnerabilities and contribute to structural inequalities within digital environments.

In addition, Jean Baudrillard's concept of hyperreality provides a powerful interpretive lens for understanding deepfake culture. Hyperreality refers to a condition in which the distinction between reality and simulation becomes blurred, with simulations becoming more compelling, persuasive, or desirable than the realities they imitate. Deepfakes embody hyperreality by creating synthetic content that can overshadow authentic events, distort public memory, or fabricate alternative narratives. In the hyperreal digital ecosystem, authenticity becomes a contested and unstable category, and individuals must navigate a world where simulations may carry greater emotional or political weight than factual content.

Alongside these macro-level theories, micro-sociological frameworks such as Erving Goffman's dramaturgical analysis offer insight into the performative aspects of identity in digital spaces. Goffman posits that individuals manage their self-presentations through front-stage and back-stage behaviours. Deepfakes disrupt this process by allowing others to produce deceptive representations of individuals without their consent. This not only undermines personal autonomy and self-identity but also creates a culture of suspicion whereby individuals must constantly question whether the digital representations of others are genuine.

Empirical studies further illuminate the scale and impact of deepfake proliferation. According to an EU Commission survey conducted in 2023, approximately 70% of citizens in major democracies expressed concern that deepfakes could be used to manipulate elections, while 64% of Indian social media users reported difficulty distinguishing between genuine and manipulated digital content. High-profile incidents, such as the deepfake of U.S. Speaker Nancy Pelosi, the fabricated video of Ukrainian President Volodymyr Zelenskyy announcing his country's surrender, and various AI-generated political campaign videos circulated during elections in India, demonstrate the real-world consequences of synthetic media. These examples highlight deepfakes' capacity to distort public perception, fuel misinformation, and erode democratic engagement.

The gendered dimensions of deepfake culture further underscore its sociological significance. Research indicates that more than 95% of deepfake videos online are non-consensual sexual content, overwhelmingly targeting women. This form of digital violence perpetuates patriarchal control over women's bodies, exacerbates gender inequalities, and creates long-

lasting psychological, reputational, and social harm. The weaponisation of deepfake pornography illustrates how technology intersects with gendered power structures in ways that reinforce existing forms of oppression.

Within the context of everyday digital life, deepfakes contribute to epistemic anxiety, a condition marked by confusion, uncertainty, and scepticism about the reliability of information sources. As deepfake incidents become more frequent, individuals increasingly question not only the authenticity of specific media artefacts but also the broader credibility of digital communication systems. This phenomenon aligns with the notion of the liar's dividend, a term used to describe how the existence of deepfake technology enables wrongdoers to deny the authenticity of legitimate evidence. In other words, even truthful content can be dismissed as fake, allowing powerful actors to evade accountability and manipulate public perception.

The architecture of contemporary digital platforms further amplifies the dangers associated with deepfakes. Social media platforms, driven by attention economies, are designed to prioritise engagement, sensation, and virality over accuracy. A 2024 MIT analysis found that false visual content spreads six times faster on social media than verified information. This structural bias toward speed and spectacle creates fertile ground for deepfakes to achieve rapid visibility, often before fact-checking mechanisms can intervene. Moreover, algorithmic curation fosters echo chambers and filter bubbles that intensify polarisation, making individuals more susceptible to misinformation aligned with their ideological leanings.

Deepfake culture also raises critical questions about the future of journalism and media institutions. Journalistic authority has historically been anchored in the ability to verify facts, produce credible narratives, and serve as a check on power. However, the rise of deepfakes jeopardises these functions by weakening the epistemic foundations of journalism. When visual evidence becomes unreliable, journalists face challenges in authenticating sources, reporting breaking news, and maintaining public trust. This crisis of credibility is compounded by the decline of traditional media institutions, the rise of citizen journalism, and the proliferation of user-generated content.

The implications of deepfake culture extend to political communication, electoral integrity, and civic participation. In increasingly polarised societies, deepfakes serve as tools for propaganda, disinformation, and psychological warfare. The circulation of deepfake political speeches, manipulated campaign materials, and fabricated evidence can influence public opinion, create confusion among voters, and undermine democratic processes. Deepfakes also pose national security risks by enabling foreign interference, cyberattacks, and geopolitical manipulation. Against this backdrop, the role of the state, regulatory bodies, and international institutions becomes crucial in developing frameworks to mitigate the harmful effects of synthetic media.

Yet, the crisis of trust associated with deepfake culture is not solely a technological or regulatory problem; it is fundamentally a sociological problem. Trust is embedded in social relations, cultural norms, institutional structures, and collective histories. Thus, the breakdown of trust cannot be addressed merely through technological solutions such as deepfake detection algorithms or digital watermarking. Instead, a comprehensive understanding requires attention to how individuals interpret, negotiate, and make sense of deepfakes within their daily lives.

The meanings people attach to synthetic media are shaped by their social backgrounds, political beliefs, educational levels, media literacy capacities, generational experiences, and digital practices.

Deepfake culture represents a profound challenge to the integrity of digital public spheres and the stability of democratic life. The crisis of trust triggered by deepfakes is not confined to isolated incidents but reflects deeper transformations in how societies produce, evaluate, and negotiate truth. As digital technologies continue to evolve, understanding the sociological dimensions of deepfake culture becomes essential for addressing the epistemic, political, and ethical challenges of the 21st century.

Methodology

This study adopts a qualitative research design to critically examine how deepfake culture contributes to the crisis of trust within digital public spheres. The purpose of using qualitative methods is to capture the interpretive meanings, subjective experiences, and social complexities associated with individuals' engagement with synthetic media. The research relies on two complementary methodological strategies: social media discourse analysis and case study examination of prominent deepfake incidents. First, social media discourse analysis was conducted across platforms such as X (formerly Twitter), Instagram, Facebook, and YouTube to explore public reactions, user-generated interpretations, comment threads, and viral discussions surrounding deepfake content. Hashtags related to deepfakes, misinformation, political manipulation, and AI-generated media were systematically identified, and posts were analysed for themes including scepticism, confusion, fear, humour, moral panic, and political distrust. This method enabled the researcher to trace how narratives about deepfakes circulate, evolve, and acquire meaning within online publics. Second, the study employed a case study approach to examine a set of high-impact deepfake incidents, such as the fabricated video of Ukrainian President Volodymyr Zelenskyy "surrendering," deepfake political campaign materials circulated in India, and non-consensual deepfake pornography cases targeting women, to understand their sociopolitical consequences, media responses, and effects on public trust. These cases were selected purposively based on their visibility, documented influence, and relevance to contemporary public discourse. Data from news articles, fact-checking reports, digital forensics analyses, and platform responses were triangulated to ensure depth and credibility. The interpretive analysis followed a thematic framework, integrating both inductive coding (emerging themes from the data) and deductive coding (themes derived from theoretical concepts such as hyperreality, risk society, and epistemic trust). Ethical considerations were carefully maintained by focusing only on publicly available content, anonymising user identities wherever necessary, and critically reflecting on the implications of studying digitally manipulated media. This methodological approach allows for a comprehensive, nuanced understanding of how deepfakes are produced, circulated, interpreted, and contested, providing rich insights into the broader sociological crisis of trust in digital public spheres.

Findings and Analysis

The qualitative analysis generated through social media discourse, platform-level interactions,

and case studies reveals a multilayered and deeply unsettling transformation occurring within digital public spheres as a result of deepfake culture. The findings highlight how synthetic media not only deceives viewers, but also shapes digital anxieties, polarisation, identity politics, moral interpretations, and institutional distrust. Rather than functioning as isolated incidents, deepfakes interact with pre-existing social tensions, power structures, and informational inequalities. Five major themes emerged from the analysis: (1) Epistemic Anxiety and the Collapse of Visual Trust, (2) Algorithmic Amplification and the Viral Ecology of Suspicion, (3) Deepfakes as Political Weapons and the Reconfiguration of Democratic Discourse, (4) Gendered Harms and the Normalisation of Digital Violence, and (5) The Liar's Dividend and the Crisis of Institutional Legitimacy. Together, these themes demonstrate that deepfake culture is not merely a technological disruption but a profound sociological shift affecting how societies understand truth, credibility, and public life.

1. Epistemic Anxiety and the Collapse of Visual Trust

One of the most prominent findings was the pervasive sense of epistemic anxiety expressed by users across digital platforms. Participants in online discussions frequently articulated confusion, doubt, and emotional unease when encountering suspicious videos or news clips. This anxiety was not limited to deepfake content itself; rather, deepfakes appeared to contaminate the trustworthiness of all digital visuals. Comments such as "How do we even know what is real anymore?" or "You can never trust videos now" emerged repeatedly across platforms during events involving manipulated content.

This phenomenon reflects a deeper sociological rupture: the erosion of the long-standing assumption that images and videos serve as objective evidence. Historically, visual media held a privileged epistemic status, what Charles Peirce described as the "indexical guarantee" of photographic realism. Deepfakes have dissolved this guarantee. Many social media users expressed the sense that digital reality had become ontologically unstable. Even when content was verified as authentic, users often responded with suspicion, suggesting a diffusion of scepticism beyond the fake itself.

Such reactions align closely with Baudrillard's concept of hyperreality, wherein the distinction between simulation and reality collapses. Social media users often struggled to categorically place content in either realm. They referred to deepfakes as "real fakes," "fake reals," or "something in-between," demonstrating a cultural discomfort with media that appears genuine but lacks an authentic referent. This confusion also mirrors Zygmunt Bauman's Liquid Modernity, where certainty dissolves, and individuals must navigate increasingly ambiguous social contexts. Deepfakes exacerbate this liquid condition by transforming truth into something fluid, contestable, and unstable.

Furthermore, the prevalence of epistemic anxiety contributed to a form of digital fatigue, with users indicating that they felt overwhelmed by the constant need to verify or cross-check information. Many described a shift from active engagement to passive consumption, stating that they had "given up" on determining what is authentic. This finding suggests that deepfake culture contributes not only to mistrust but also to disengagement from public discourse, thereby weakening civic participation.

2. Algorithmic Amplification and the Viral Ecology of Suspicion

A second major finding relates to how digital platforms, especially algorithm-driven ones like X, Instagram, and YouTube, amplify deepfakes in ways that accelerate distrust. Deepfake content often spreads faster, farther, and with more emotional resonance than fact-checked or debunked material. Users frequently shared manipulated content reflexively, sometimes out of outrage, humour, confusion, or fear. This created what can be described as a viral ecology of suspicion, where deepfakes circulated not only as misinformation but also as symbols of broader cultural anxieties.

The analysis showed that algorithms tended to reward content that triggered strong emotional reactions, anger, shock, and amusement being the most common responses to deepfakes. This aligns with Papacharissi's concept of affective publics, where emotions guide engagement and collective meaning-making. Deepfakes, by virtue of their sensational nature, act as affective triggers that mobilise users more quickly than factual or mundane content. This emotional virality creates environments where misinformation thrives and truth is often an afterthought.

Users often accused each other of sharing fake or manipulated content, even when such content was legitimate. In comment threads, predictions of "This is a deepfake" or "AI did this" appeared even in response to genuine videos, demonstrating how deepfakes generate suspicion not only toward fabricated media but toward reality itself. This self-reinforcing cycle of doubt is further amplified by platform features like reposting, sharing, memeification, and algorithmic recommendation.

Additionally, social media discourse reflected confusion about how platforms themselves classify or label deepfakes. Some users complained that platforms were inconsistent: certain AI-generated videos were labelled as "manipulated," while others circulated freely without warnings. This inconsistency undermined the credibility of platform governance and contributed to the perception that digital ecosystems lack reliable safeguards.

This theme intersects strongly with Ulrich Beck's Risk Society, which emphasises how modern technologies produce new forms of uncertainty and system-generated risks. Deepfakes exemplify such risks by creating unpredictable impacts that scale rapidly across networks. Users expressed a sense that they were constantly on the verge of being misled, manipulated, or deceived by forces beyond their control.

3. Deepfakes as Political Weapons and the Reconfiguration of Democratic Discourse

A third major finding concerns the weaponisation of deepfakes in political contexts. Case studies revealed that deepfakes have increasingly been deployed to shape political narratives, manipulate public opinion, and deepen polarisation. The analysis of political deepfakes, such as fabricated speeches, out-of-context videos, or AI-generated "scandals," showed that synthetic media can influence voters even after being debunked. Users often reported believing manipulated content initially and only later realising that it was falsified, yet the emotional residue of the initial exposure persisted.

Deepfake political content generated intense debate on social media, but often in polarised forms. Supporters of one political faction would accept or circulate deepfakes targeting

opponents, while dismissing or condemning deepfakes targeting their own group. This selective scepticism reflects a form of motivated reasoning, where deepfakes function as political ammunition rather than informational artefacts.

The Zelenskyy surrender deepfake, for instance, triggered widespread confusion and fear during the Russia-Ukraine war. Social media users described feeling "momentarily shocked" or "conflicted" before fact-checks emerged. Indian political deepfakes, particularly those portraying party leaders making inflammatory remarks, generated extensive polarisation, with users interpreting them through ideological lenses rather than factual assessments. This aligns with theories of post-truth politics, where truth becomes secondary to identity-based narratives.

Habermas's ideal of the public sphere as a site for rational-critical debate is severely compromised in such contexts. Instead of facilitating deliberation, deepfakes amplify affective reactions and reduce political discussions to battles of perception rather than substance. Many users expressed hopelessness over the idea of democratic decision-making in an era where truth is manipulable: "If everything can be faked, how can we vote responsibly?"; "Elections will never be the same."

This theme shows that deepfakes reconfigure democratic discourse by eroding the epistemic foundations upon which democratic legitimacy depends. When citizens cannot fully trust political communications, leaders' speeches, or news reports, the possibility of informed civic engagement diminishes significantly.

4. Gendered Harms and the Normalisation of Digital Violence

Perhaps the most disturbing finding relates to the gendered dimensions of deepfake culture. A substantial portion of deepfake content remains non-consensual sexual imagery, overwhelmingly targeting women. Social media discourse surrounding deepfake pornography reveals patterns of victim blaming, trivialisation, humour, and sexual objectification. Many users dismissed victims' experiences by claiming "it's just AI" or "it's not real," ignoring the severe emotional and reputational harm inflicted upon targeted individuals.

The analysis showed that women, particularly celebrities, journalists, activists, and social media influencers, are disproportionately affected. Case studies demonstrated that deepfake pornography is frequently used for harassment, blackmail, public shaming, and revenge. Victims who spoke publicly described feelings of humiliation, identity violation, and helplessness, illustrating how deepfakes blur the boundaries between bodily autonomy and digital exploitation.

This phenomenon aligns with feminist theories of digital patriarchy, where technology becomes a tool for reproducing gendered power inequalities. Deepfake pornography reinforces male dominance by weaponising women's bodies, violating their consent, and reducing them to sexualised objects. Users frequently internalised these violations as entertainment, thus normalising digital violence.

Additionally, deepfakes compromise the authenticity of women's online self-presentations. According to Goffman's dramaturgical theory, individuals construct identity through controlled self-representation. Deepfakes disrupt this process by allowing others to fabricate identities on

their behalf, stripping women of agency over their own image. This results in heightened self-surveillance among women online, driven by fear that their images might be misappropriated.

The gendered harms of deepfakes therefore represent not only individual-level trauma but structural reproduction of patriarchal power within digital spaces. While platform policies nominally prohibit non-consensual synthetic pornography, enforcement is inconsistent and often ineffective, further deepening mistrust in institutional safeguards.

5. The Liar's Dividend and the Crisis of Institutional Legitimacy

A central analytical theme emerging from the case studies is the concept of the liar's dividend, the ability of wrongdoers to deny authentic evidence by claiming it is fake. In several political and legal cases, individuals confronted with incriminating videos or recordings dismissed them as deepfakes, even when authenticity was later verified. This phenomenon demonstrates how deepfake culture erodes accountability mechanisms.

Users expressed frustration over this dynamic, noting that public figures increasingly evade responsibility by exploiting the ambiguity created by deepfakes. Comments such as "Now every politician can claim videos are fake" or "Deepfake is the new excuse for everything" illustrate how synthetic media undermines institutional credibility.

This erosion of accountability extends to journalism as well. News organisations faced difficulties verifying visual content quickly enough, resulting in delayed reporting or retractions. Users frequently accused media houses of incompetence or complicity, reflecting a wider crisis of institutional trust.

The liar's dividend creates two parallel crises:

1. Genuine evidence is dismissed as fake, enabling wrongdoers to escape scrutiny.
2. Fake evidence is believed as real, enabling manipulation of public perception.

This double-edged crisis represents a profound challenge for democratic systems, legal institutions, and journalistic ethics. Moreover, it contributes to a cultural environment where truth itself becomes negotiable, subject to denial, reinterpretation, or outright rejection.

These findings suggest that deepfake culture amplifies pre-existing structural distrust in institutions, making it increasingly difficult to sustain shared standards of evidence and public accountability.

Across all themes, the central conclusion is clear: deepfakes destabilise not only content but the entire epistemic ecology of digital societies. They corrode trust horizontally (between individuals), vertically (between citizens and institutions), and internally (within the self, through epistemic anxiety). The analysis indicates that the crisis of trust is not caused solely by deepfakes but is magnified by algorithmic architectures, political polarisation, patriarchal structures, and widespread public scepticism.

Deepfake culture therefore represents a sociological phenomenon that reshapes:

- how people interpret digital reality,

- how political narratives are contested,
- how gendered power operates online,
- how institutions maintain legitimacy, and
- how democratic publics function.

Discussion and Conclusion

The findings of this study illuminate a deepening sociological crisis rooted not simply in the technological capacity of deepfakes but in the broader transformations of trust, authority, and meaning-making in contemporary digital societies. While deepfake technology is often understood as a technical threat, something to be solved through better detection tools or platform regulation, the analysis demonstrates that the implications extend far beyond the technological domain. Deepfakes destabilise the social foundations of credibility, heighten epistemic insecurities, and intensify the fragmentation of public spheres already strained by misinformation, political polarisation, and platform capitalism. In this sense, deepfake culture is not an isolated phenomenon but part of a larger post-truth condition in which truth itself becomes negotiable, contested, and increasingly personalised.

One of the central insights that emerges from the analysis is that deepfakes accelerate the erosion of epistemic trust, the basic confidence individuals place in the information they consume and the institutions that produce it. The public's uncertainty is less about whether a specific video is "real" or "fake" and more about the broader feeling that *anything could be manipulated*. This emotional state of scepticism, confusion, and doubt signifies what scholars describe as epistemic anxiety, a condition shaped by information overload, algorithmic amplification, and declining institutional legitimacy. Deepfake culture intensifies this anxiety by making the manipulation of visual and auditory evidence not only possible but also commonplace. Even the potential for manipulation leads individuals to question content that is, in fact, authentic, reflecting the "liar's dividend," a phenomenon in which culprits exploit the existence of deepfake technology to dismiss real evidence as fabricated. This dual erosion, trust in true content and ambiguity in false content, constitutes a direct threat to democratic communication.

The discussion also indicates that deepfake culture interacts deeply with Habermas's theory of the public sphere, which stresses rational-critical debate as a foundation for democratic society. Digital public spheres, however, operate under very different conditions: they are commercialised, algorithmically structured, and saturated with emotionally charged content. Within this context, deepfakes function as powerful agents of disruption. They inject symbolic uncertainty into processes of deliberation, making it difficult for publics to distinguish credible voices from synthetic representations. In doing so, deepfakes undermine the communicative rationality that Habermas envisioned, replacing evidence-based discussion with spectacle, sensationalism, and suspicion. Rather than fostering open discourse, digital spaces become arenas of distrust where users oscillate between credulity and cynicism.

Similarly, the findings resonate strongly with Baudrillard's concept of hyperreality, which argues that in contemporary society, the boundary between the real and the simulated becomes

increasingly blurred. Deepfakes bring this theoretical concept to its highest expression. They produce simulations indistinguishable from reality, collapsing the distinction between original and imitation. In the hyperreal condition, the question is no longer "Is this real?" but "What does real even mean?" The public's struggle to validate authenticity in digital spaces reflects a hyperreal environment where images and appearances overshadow empirical verification. Deepfake culture thereby exposes the fragility of truth in a world governed by symbolic reproduction.

Sensational deepfake videos, particularly political or scandalous ones, are more likely to be liked, shared, and monetised, revealing how economic logics intersect with epistemic vulnerability. The political deepfakes analysed in this study (such as the Zelenskyy surrender video or AI-generated political campaign clips in India) illustrate how actors strategically exploit the affordances of these platforms to manipulate public opinion. Meanwhile, platforms maintain an ambiguous stance: they publicly condemn misinformation but profit from attention-grabbing content. The crisis of trust, therefore, is not an accidental by-product but a structural outcome of digital economies that privilege virality over veracity.

Gendered implications also emerge strongly, particularly in the case of non-consensual deepfake pornography. Women face a distinct form of vulnerability as deepfakes are weaponised to shame, silence, and socially control them. This reflects broader patriarchal structures in digital cultures, where technological tools amplify gendered violence rather than mitigate it. The emotional distress, reputational damage, and fear of social scrutiny described in the discourse analysis underscore how deepfakes can serve as instruments of symbolic violence. These findings expand the scope of deepfake scholarship beyond political manipulation, highlighting the intersections of technology, gender, and power.

Despite the alarming implications, the discussion also points toward evolving forms of adaptation and digital resilience. Many users employ informal verification strategies, such as cross-checking sources, relying on trusted communicators, or using fact-checking websites. While these practices cannot fully counter the structural challenges posed by deepfakes, they demonstrate that publics are not passive victims but active interpreters who develop coping mechanisms to navigate uncertainty. This resilience, however, is uneven across social groups; those with higher digital literacy and media awareness are better equipped to detect manipulation, while others remain vulnerable.

This discussion underscores that solutions to deepfake-driven mistrust must extend beyond technological detection tools. The crisis is rooted in structural forces: platform incentives, declining institutional legitimacy, political polarisation, and the hyperreal media environment. Addressing these crises requires strengthening media literacy, rethinking platform governance, fostering digital ethics, and rebuilding the social foundations of trust. Deepfake culture ultimately exposes the broader vulnerabilities of digital societies, reminding us that the crisis of trust is not simply about images, it is about the fragile architecture of truth upon which democratic life depends.

The crisis of trust triggered by deepfake culture reflects a deeper sociological transformation: the shift from an information-based society to a verification-based society. In earlier media

environments, the primary concern was accessing information; today, the challenge lies in evaluating its authenticity. This transition places new cognitive, emotional, and ethical burdens on individuals and communities. Trust, once anchored in institutions, shared norms, and stable media systems, becomes a personalised responsibility, negotiated individually through fragmented platforms. As a result, public trust becomes fragile, provisional, and easily destabilised.

References:

1. Battista, D. (2024). *Political communication in the age of artificial intelligence: an overview of deepfakes and their implications*. *Society Register, 8*(2), 7–24.
2. Chadwick, A. (2019). *The new crisis of public communication: Challenges and opportunities for future research on digital media and politics*.
3. Eynern, C. (2024). *Olaf Scholz Deepfake: How a Deepfake impacts Public Trust* (Bachelor's thesis, University of Twente).
4. Faragó, T. (2019). *Deep fakes–an emerging risk to individuals and societies alike*.
5. Gehringer, J. (2024). *The postmodern public sphere: ICT, misinformation and simulated civil society*. In *Handbook of Critical Perspectives on Nonprofit Organizing and Voluntary Action* (pp. 149–163). Edward Elgar Publishing.
6. Gregory, S. (2022). *Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism*. *Journalism, 23*(3), 708–729.
7. Gregory, S. (2024). *From social media to Deepfakes: Participatory human rights witnessing and advocacy using audiovisual media, incorporating the emerging impacts of deceptive AI and technologies for authenticity and trust (2007–22)* (Doctoral dissertation, University of Westminster).
8. Morris, K. W. (2024). *Deepfake Sockpuppets: The Toxic "Realities" of a Weaponised Internet*. In *Gothic nostalgia: The uses of toxic memory in 21st century popular culture* (pp. 61–79). Springer International Publishing.
9. Somogyi, A. (2023). *ARE DEEPFAKES A THREAT? REDEFINING DEEPFAKE-AI THROUGH POPULAR CULTURE & THE EVERYDAY* (Doctoral dissertation, Central European University).
10. Verma, N. (2023). *Deepfake technology and the future of public trust in video* (Doctoral dissertation).
11. Yadlin-Segal, A., & Oppenheim, Y. (2021). Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence, 27*(1), 36–51.