From Hoax to Reality: Deepfake-driven Misinformation and the Death of Ozzy Osbourne

Alexander Pfeiffer[1], Nanditha Krishna[3], Thomas Wernbacher[2] and Walter Seböck[1]
[1]Department for Security Studies, University for Continuing Education Krems, Austria
[2]Department for Arts and Culture, University for Continuing Education Krems, Austria
[3] Independent Researcher

alexander.pfeiffer@donau-uni.ac.at
walter.seboeck@donau-uni.ac.at
Thomas.wernbacher@donau-uni.ac.at
Nandithakrish4@gmail.com

Abstract: Deepfakes are AI-generated images, videos, and texts that convincingly mimic real individuals. In recent years, such forgeries have proliferated across social media, and their role in fraud cases has increased markedly. However, detection tools often fail in real-world conditions; open-source detectors typically perform only half as well on "in-the-wild" content compared to curated test sets. This performance gap heightens the risk that fabricated content will undermine public trust and foster a climate of suspicion in which even authentic recordings are questioned—a phenomenon known as the "liar's dividend." In this case study, we examine how the July 2025 death of rock icon Ozzy Osbourne became a focal point for deepfake-driven misinformation and public speculation. Following a seated farewell concert in Birmingham, multiple synthetic videos surfaced, including one in which a digitally recreated Osbourne claimed he knew he was about to die. The clips sparked speculation about assisted suicide, prompting his daughter Kelly to publicly denounce the videos as fake and criticise those who shared them. When Osbourne died two weeks later, some commentators treated the deepfake as prophetic, fuelling conspiracy theories and amplifying public grief. This case illustrates broader ethical and governance challenges related to generative AI. Voice cloning and face-swapping services can create convincing media from minimal training data, yet developers rarely address issues of consent or privacy when sourcing material. Psychological factors such as fear of missing out (FOMO) encourage the viral spread of sensational content without verification. This paper's primary contribution is a theoretical synthesis, which integrates existing technical, psychological, and governance perspectives on deepfake-driven misinformation through a single illustrative case study. Effective countermeasures must combine technical innovations—such as blockchain-based provenance tracking and robust detection—with clear policy frameworks that regulate data use and require transparent labelling of synthetic media. Public education remains essential to help individuals recognise deepfakes and preserve trust in authentic digital communication.

Keywords: Deepfakes, Misinformation, Fake news, Liar's dividend, Digital trust, FOMO

1. Introduction

The proliferation of generative artificial intelligence (GenAI) has democratised the creation of synthetic media, enabling anyone with an internet connection to produce highly realistic yet fabricated images, audio, and video content, commonly known as deepfakes (Lundberg and Mozelius, 2025). Whilst this technology offers creative and educational affordances, its potential for malicious use has become a significant societal concern. Deepfakes are increasingly implicated in a range of harmful activities, including the spread of disinformation, character defamation, financial fraud, and the manipulation of democratic processes (Lundberg and Mozelius, 2025; Brennan Center for Justice, 2023; Pfeiffer and Krishna, 2025). The core challenge lies in the technology's ability to erode public trust in digital communication, creating an environment where the authenticity of any piece of media can be called into question. This erosion of trust is not merely a theoretical risk; it is a tangible threat that has been observed in numerous real-world incidents, where manipulated content has been used to mislead the public and sow discord. Recent advances in generative AI, particularly Google's Veo 3 model announced in May 2025, have dramatically escalated the sophistication and accessibility of synthetic media creation, enabling the generation of photorealistic videos with synchronised audio that are nearly indistinguishable from authentic content (Pfeiffer and Krishna, 2025). In its first week of release, users posted fake news segments in multiple languages, including fabricated death announcements and political news conferences. One commentator described Veo 3 as a potential "death knell for truth on the internet" (PCMag, 2025).

The Ozzy Osbourne case of July 2025 serves as a stark and poignant example of the real-world impact of deepfake technology. The incident, which unfolded in the weeks leading up to the rock icon's death, illustrates the complex interplay between synthetic media, public speculation, and the psychological vulnerabilities that drive the spread of misinformation. Multiple deepfake videos, appearing to show a frail Osbourne prophesying his own demise, were widely circulated on social media platforms, triggering a maelstrom of rumours and conspiracy theories (Amelinckx, 2025). The videos' apparent prescience, following Osbourne's actual death on

22 July 2025, lent them a veneer of credibility in the eyes of some, demonstrating how deepfakes can exploit and amplify public grief and confusion. The incident is documented in the OECD AI Incidents Monitor as "AI-Generated Deepfake Video Falsely Claims Ozzy Osbourne Is Dying" (incident date: 12 July 2025), classified as an AI Incident with reputational, psychological, and public interest harms (OECD AI Policy Observatory, 2025). This case study provides a valuable lens through which to examine the multifaceted challenges posed by deepfakes, from the technical limitations of detection tools to the ethical and governance vacuums in which this technology currently operates. By dissecting the Osbourne case, we can gain a deeper understanding of the mechanisms by which deepfakes exert their influence and explore potential countermeasures to mitigate their harmful effects. This paper's primary contribution is a theoretical synthesis that integrates existing technical, psychological, and governance perspectives on deepfake-driven misinformation. Rather than advancing a new empirical study or formal theory, we use the Ozzy Osbourne case as a single, illustrative example to put together these disparate threads of analysis. This approach allows for a broader understanding of the deepfake phenomenon, guiding how technological capabilities, human psychology, and regulatory frameworks interact in a real-world context

This paper will proceed as follows. First, it will provide a brief overview of deepfake technology and the current state of detection capabilities, highlighting the significant performance gap between academic benchmarks and real-world applications. Second, it will present a detailed analysis of the Ozzy Osbourne case study, tracing the timeline of events and examining the public's reaction to the deepfake videos. Third, it will look into the psychological factors, such as the 'fear of missing out' (FOMO), that contribute to the viral spread of sensationalist and unverified content. Finally, it will explore the broader ethical and governance challenges associated with generative AI, including the issues of consent in data training and the need for robust policy frameworks. Effective countermeasures must combine technical innovations such as blockchain-based provenance tracking with robust detection frameworks and clear labelling of synthetic media.

2. The Technological Challenge: Deepfakes and the Detection Arms Race

Deepfakes are the product of sophisticated machine learning techniques, specifically generative adversarial networks (GANs), which involve two neural networks—a generator and a discriminator—competing against each other to produce increasingly realistic synthetic media. The generator creates the fake content, whilst the discriminator attempts to distinguish it from authentic content. This adversarial process results in a rapid improvement in the quality and believability of deepfakes, making them increasingly difficult to detect with the naked eye. The accessibility of deepfake creation tools has also grown exponentially, with numerous online services and open-source software allowing users to create convincing forgeries with minimal technical expertise (Lundberg and Mozelius, 2025). This democratisation of deepfake technology, whilst having some positive applications in areas like entertainment and education, has also lowered the barrier to entry for malicious actors seeking to exploit it for nefarious purposes. The landscape has become dramatically more challenging with recent technological advances that can create clips that are "nearly indistinguishable from real ones," complete with dialogue, soundtracks, sound effects, and adherence to real-world physics (Pfeiffer and Krishna, 2025)[1].

The rapid advancement of deepfake generation has triggered a corresponding arms race in detection technology. Researchers and technology companies are developing a variety of methods to identify synthetic media, ranging from the analysis of digital artefacts and inconsistencies in the generated content to the use of blockchain-based provenance tracking to verify the origin and history of a piece of media (Content Authenticity Initiative, 2025; Tech Policy Press, 2025; Pfeiffer and Krishna, 2025). However, the effectiveness of these detection tools in real-world scenarios remains a significant concern. A recent study by Chandra et al. (2025), which introduced a new benchmark of in-the-wild deepfakes, found that the performance of open-source state-of-the-art detection models exhibited a marked performance degradation when evaluated on real-world content, with a decrease in the area under the curve (AUC) of approximately 50% for video models, 48% for audio models, and 45% for image models compared to their performance on academic datasets. This performance gap is a critical vulnerability, as it means that many of the deepfakes circulating on social media and other online platforms are likely to evade detection by automated systems. Even commercial detection models, whilst outperforming their open-source counterparts, still fall short of the accuracy achieved by human forensic analysts, highlighting the ongoing challenge of developing robust and reliable detection methods

[1]One of the authors can confirm this from personal experience. After a playdate, the mother of the other child sent a photo of both children to her father. He uploaded it to GROG without any hesitation and created an extremely lifelike video showing the two children dancing and falling over. This shows that the tools are easy to use even for computer-savvy people aged 80+, but that media literacy was completely lacking, at least in this case. In addition to the cyberbullying aspect, the overarching issue is, of course, the provision of training data from image material that does not belong to one person. This also applies to images of minors and even small children.

(Chandra et al., 2025). The study's findings underscore a troubling reality: existing approaches, primarily focused on deepfake detection, fall short due to the evolving sophistication of artificial intelligence methods and the emergence of what researchers term the "liar's dividend"—the ability for bad actors not only to spread fake content but also to dismiss authentic content as potentially fake. This dynamic, systematically analysed by Chesney and Citron (2019) in the California Law Review, highlights the erosion of epistemic trust when authenticity itself becomes contestable (Chesney and Citron, 2019; Pfeiffer and Krishna, 2025).

The limitations of current detection technologies are further compounded by the sheer volume and velocity of content being generated and shared online. Social media platforms, in particular, are fertile ground for the spread of deepfakes, as their algorithms are often designed to prioritise engagement and virality over accuracy and authenticity. The psychological phenomenon of FOMO, a pervasive anxiety that one might miss out on exciting or interesting events that others are experiencing, can further accelerate the spread of sensationalist content, as users are more likely to share information without verifying its authenticity in an effort to stay connected and engaged (Gong and Ren, 2023). The interplay between rapid deepfake evolution, limited detection capacity, and social amplification mechanisms such as FOMO creates fertile ground for misinformation, as illustrated by the Ozzy Osbourne case.

The field of deepfake detection has evolved rapidly, with researchers exploring diverse approaches ranging from traditional forensic techniques to advanced deep learning methods. A detailed review by Heidari et al. (2024) categorises detection methods into four primary domains: video detection, image detection, audio detection, and hybrid multimedia detection, each presenting unique challenges and opportunities. Verdoliva (2020) provides a foundational overview of media forensics in the deepfake era, noting that whilst traditional manipulation techniques such as splicing and copy-move forgery can be detected through pixel-level analysis, deepfakes generated by sophisticated neural networks present fundamentally different challenges. One promising avenue involves exploiting biological signals that are difficult for generative models to replicate. Ciftci, Demir, and Yin (2020) demonstrated that photoplethysmography (PPG) signals—subtle colour changes in facial skin caused by blood flow—can serve as an authenticity marker, as these physiological patterns are neither spatially nor temporally preserved in synthetic content. Their 'FakeCatcher' system achieved accuracies of 96% on Face Forensics and 91.07% on in-the-wild datasets by analysing biological signal maps through convolutional neural networks, suggesting that nature's own 'stamps' may provide more robust detection than artefact-based methods alone.

3. Case Study: The Deepfake Prophecy of Ozzy Osbourne's Death

The case of Ozzy Osbourne in July 2025 provides a compelling and tragic illustration of how deepfake technology can be weaponised to exploit public figures and manipulate public sentiment. In the weeks preceding his death, the 76-year-old rock legend, who had been publicly battling Parkinson's disease, became the subject of multiple viral deepfake videos that purported to show him speaking about his own impending demise. The timeline of events unfolded as follows: on 5 July 2025, Osbourne performed his "Back to the Beginning" farewell concert at Villa Park in Birmingham, England, reuniting with Black Sabbath for what would be his final show (Entertainment Weekly, 2025). Just four days later, on 9 July 2025, a TikTok clip surfaced featuring a synthetic voice-over claiming to be Osbourne discussing his final performance, stating, "This was my final performance. Uh, I have to say goodbye to this stage I've loved all my life. I'm truly sorry. I could only do this show strapped into a custom chair 'cause my body just couldn't handle it anymore. My Parkinson's has reached stage 5" (Lead Stories, 2025).[2] The video further claimed that Osbourne had donated all proceeds from the concert to Parkinson's research, a statement that was never corroborated by fact-based reporting.

The deepfake videos immediately fuelled speculation and conspiracy theories, with many social media users accepting the content as authentic. The clips were often presented in the context of a rumoured suicide pact between Osbourne and his wife, Sharon, a narrative that had its origins in historical interviews where Sharon Osbourne had discussed their shared belief in euthanasia under certain circumstances. In 2023, Sharon confirmed on the family's podcast that they still maintained an assisted suicide pact, stating they would go to Switzerland for euthanasia if either developed Alzheimer's disease (Yahoo Entertainment, 2025). The deepfake videos, therefore, did not emerge in a vacuum but rather tapped into and amplified existing public narratives and anxieties surrounding Osbourne's health. The situation prompted a swift and angry response from

[2] Archived here: https://web.archive.org/web/20250725165520/https://www.tiktok.com/@fametalks.01/video/7525036503290973454?lang=en

Osbourne's daughter, Kelly, who took to Instagram on 11 July 2025 to denounce the videos as malicious fabrications. In her public statement, she vehemently denied the rumours of her father's impending death and criticised the creators and distributors of the deepfakes, asking, "What is wrong with you?" (People Magazine, 2025). The incident was covered by multiple international news outlets, including People Magazine, Entertainment Weekly, and various social media platforms, indicating the widespread circulation of the deepfakes before Osbourne's actual death.

However, the deepfake phenomenon surrounding Osbourne did not end with Kelly's denial. The TikTok video posted on 9 July 2025 by the account @fametalks.01 continued to circulate, and was subsequently debunked by Lead Stories, a fact-checking organisation, on 26 July 2025. Lead Stories used VERA.ai detection software to analyse the video, and the analysis confirmed AI-generated segments with greater than 99% certainty and non-matching voice profiles, and the voice-over did not match any actual statements issued by Osbourne or corroborated by fact-based reporting (Lead Stories, 2025). The narration was identified as characteristic of AI text-to-speech programs, with no effort made to synchronise the audio with mouth movements in the video clips, further confirming its synthetic nature. The fact-check article noted that the video was a narrated PowerPoint-style slideshow with a series of clips and still images from the Villa Park show, and that the producer had no access to video of Osbourne saying these things.

Tragically, when Ozzy Osbourne passed away on 22 July 2025 (BBC News, 2025), just eleven days after Kelly's denial and thirteen days after the TikTok deepfake video was posted, the fabricated content took on a new and even more insidious dimension. Following Osbourne's death, the fabricated clips were retrospectively interpreted as prophetic, demonstrating deepfakes' capacity to manipulate grief and blur truth boundaries. The family was subsequently confronted with "disgusting" conspiracy theories, including false suicide claims that family friends publicly denounced as "vile" rumours (Watson.de, 2025). This phenomenon highlights a particularly dangerous aspect of deepfakes: their ability to be re-contextualised by real-world events, making them even more potent tools of misinformation. Social media discussions following Osbourne's death revealed widespread exposure to AI-generated tributes and memorial content. One such video, with 5.8 million views, exemplified the problem of "AI slop"—synthetic content designed solely for engagement, featuring an uncanny rendering of Osbourne with inappropriate music (Barlow, 2025). The Osbourne case demonstrates how deepfakes can not only create false narratives but also hijack and distort real ones, amplifying public grief and confusion in the process. It serves as a powerful reminder of the urgent need for more effective strategies to counter the spread of synthetic media, from improved detection and provenance tools to greater public awareness and critical media literacy.

Whilst this case study provides valuable insights into the real-world dynamics of deepfake misinformation, it is important to acknowledge certain limitations. As a single-case analysis focused on a celebrity figure with pre-existing health concerns and public narratives, the findings may not generalise to all deepfake scenarios, particularly those targeting private individuals or operating in different cultural contexts. Additionally, the retrospective nature of this analysis means that we cannot fully quantify the reach and impact of the deepfakes or measure the effectiveness of counter-narratives in real-time. Nevertheless, the Osbourne case offers a rich illustration of the temporal dynamics, psychological vulnerabilities, and detection challenges that characterise deepfake incidents more broadly.

4. The Human Element: Psychological Drivers of Misinformation

The viral spread of the Ozzy Osbourne deepfakes cannot be attributed solely to the sophistication of the technology itself. The incident also underscores the critical role of human psychology in the dissemination of misinformation. One of the key psychological drivers at play is FOMO, which is characterised by a persistent anxiety that one might miss out on rewarding experiences that others are having (Gong and Ren, 2023). In the context of social media, FOMO often manifests as a compulsive need to stay connected and up-to-date with the latest news and trends, leading to a higher propensity to share content without adequate verification. Research has shown that FOMO mediates the relationship between psychological distress and susceptibility to misinformation, with individuals experiencing higher levels of FOMO being more likely to believe fake news (Gong and Ren, 2023). FOMO interacts with the liar's dividend (discussed later in detail) (Pfeiffer and Krishna, 2025) to create a feedback loop between affective arousal and misinformation sharing. Social media platforms, with their constant stream of updates and notifications, provide an "efficient and low friction path" for information consumption, which can exacerbate FOMO and make users more susceptible to manipulation (Gong and Ren, 2023). The sensational and emotionally charged nature of the Osbourne deepfakes made them

particularly well-suited to exploit this psychological vulnerability, as users were driven to share the content to be part of the unfolding drama and to signal their awareness of the 'breaking news.'

Beyond FOMO, other cognitive and emotional factors contribute to the public's susceptibility to fake news. Research has shown that older individuals may be particularly vulnerable to misinformation, as they are often less likely to scrutinise suspicious content and may have lower levels of digital literacy (Gong and Ren, 2023). A cross-sectional study of earthquake survivors found that post-traumatic stress disorder (PTSD) was directly associated with believing fake news, and that this relationship was mediated by FOMO, with effects more pronounced in older people and present in females but not males (Gong and Ren, 2023). Whilst the Osbourne case does not involve a traumatic event in the traditional sense, the public's emotional investment in the rock star's health and the anticipation of his death created a heightened state of anxiety and vigilance that may have made individuals more susceptible to believing and sharing the deepfakes. The emotional content of misinformation also plays a significant role; people are more likely to believe and share false statements that appeal to their existing beliefs, biases, and emotions. In the Osbourne case, the deepfakes tapped into a pre-existing narrative of the rock star's declining health and the public's affection for him, creating a potent emotional cocktail that made the content more believable and shareable. The fact that the deepfakes were presented as 'prophecies' that were later 'fulfilled' by Osbourne's death further reinforced their emotional impact, making it even more difficult for some to dismiss them as fabrications.

The challenge of combating deepfake-driven misinformation is therefore not just a technical one; it is also a profoundly human one. It requires a deeper understanding of the psychological mechanisms that make us vulnerable to manipulation and the development of strategies to foster greater critical thinking and media literacy. This includes educating the public about the existence and capabilities of deepfake technology, teaching them how to identify the signs of a potential fake, and encouraging a culture of verification before sharing. As the line between reality and artifice becomes increasingly blurred, the ability to critically evaluate and question the information we encounter online will be an essential skill for navigating the digital world responsibly.

5. Ethical and Governance Challenges: The Consent Gap and the Liar's Dividend

The Ozzy Osbourne case study is a microcosm of the broader ethical and governance challenges posed by the rapid advancement of generative AI. At the heart of these challenges is the issue of consent. The creation of deepfakes, particularly those that impersonate real individuals, raises profound questions about the right to one's own likeness and voice. Current legal and ethical frameworks, which are largely built on a traditional understanding of consent, are proving to be inadequate in the age of AI (Pistilli and Trevelin, 2025). As researchers have pointed out, whilst an individual might consent to their data being used for a specific purpose, they cannot meaningfully consent to the myriad potential outputs that their data could be used to generate, a phenomenon described as the "consent gap" (Pistilli and Trevelin, 2025). This is particularly problematic when AI-generated content creates new and unforeseen forms of personal representation, often without the individual's knowledge or permission, and can be distributed globally in an instant. The question of whether AI can be truly consentful remains a pressing ethical concern, as the technology's capacity to generate synthetic representations far exceeds the scope of traditional consent mechanisms (Pistilli and Trevelin, 2025).

The sourcing of training data for AI models is another area fraught with ethical and legal ambiguity. Many generative models are trained on vast datasets scraped from the internet, which often include personal images, videos, and writings that are used without the explicit consent of the individuals concerned. This practice raises significant privacy concerns and has led to calls for greater transparency and accountability in the AI development pipeline. The development of robust governance frameworks is therefore essential to address these challenges. This includes the establishment of clear regulations regarding the use of personal data in AI training, as well as the implementation of policies that mandate the transparent labelling of synthetic media. In response to the growing threat of deepfakes, some jurisdictions are taking pioneering legislative action. Denmark, for instance, announced a proposal in June 2025 to amend its copyright laws, granting individuals rights over their biometric likenesses and positioning facial and vocal features as intellectual property to provide stronger legal recourse against AI-generated impersonations (The Guardian, 2025). This proposal has not yet been enacted. Similarly, Northern Ireland has taken steps towards criminalising the creation and sharing of deepfake pornography, reflecting a critical step towards addressing the severe psychological harm and gendered abuse associated with such content. Justice Minister Naomi Long stated in July 2025 that she wants deepfakes to become a criminal offence "sooner rather than later" (BBC News, 2025).

Beyond the issue of consent, the deepfake phenomenon has given rise to what researchers term the "liar's dividend"—the ability for bad actors not only to spread fake content but also to dismiss real footage as fake (Chesney and Citron, 2019; Pfeiffer and Krishna, 2025). This erosion of trust in authentic media represents a fundamental threat to democratic discourse and accountability. If seeing and hearing can no longer be believed, the very notion of truth is undermined, and public figures can evade accountability by falsely claiming that damaging but authentic content is fabricated (Chesney and Citron, 2019; Pfeiffer and Krishna, 2025). The Osbourne case illustrates this dynamic in reverse: whilst the deepfakes were fabrications, their retrospective 'confirmation' by Osbourne's actual death created a situation where some members of the public may have been more inclined to believe future synthetic content, further eroding the boundary between truth and falsehood. Policymakers must therefore prevent manipulated media from being used to undermine elections, disenfranchise voters, and distort public discourse (Brennan Center for Justice, 2023).

Initiatives like the Content Authenticity Initiative (CAI) and the Coalition for Content Provenance and Authenticity (C2PA) are pioneering technical solutions to address these challenges. The CAI, a cross-industry community including civil society, media, and technology companies, is promoting the adoption of Content Credentials, which function like a nutrition label for digital content (Content Authenticity Initiative, 2025). These credentials are based on open-source tools for verifiably recording the provenance of any digital media, including content made with generative AI, and focus on verifying the origin, integrity, and history of digital content. The C2PA specification reached version 2.1, reinforcing interoperability in provenance metadata (C2PA, 2025). Building upon C2PA's open standard, Pfeiffer and Krishna (2025) propose cryptographic hashing and metadata anchoring integrated into capture devices, optionally linked to eIDAS or SSI credentials, and immutably anchored on public blockchains (Ethereum Layer 2, Solana, Ardor) to form a provenance ledger capable of rapid verification and tamper resistance. Such blockchain-based media authentication offers multiple advantages, including strengthened content integrity, clear provenance, rapid verification capabilities, deterrence of misinformation, protection of intellectual property, and resilience across decentralised platforms.

However, as policy experts have noted, technical solutions alone are not sufficient. They must be complemented by a comprehensive regulatory approach that moves beyond a simplistic "synthetic or authentic" binary and addresses the full spectrum of risks associated with AI-generated media, from its potential to undermine democratic processes to its use in personal harassment and fraud (Tech Policy Press, 2025). Current approaches to regulating synthetic content may face a number of limitations and trade-offs, including concerns related to privacy and security, device integration complexities, scalability, verification infrastructure, and evolving legal and regulatory landscapes (Tech Policy Press, 2025; Pfeiffer and Krishna, 2025). To overcome these barriers, parallel initiatives in user and stakeholder education are crucial, emphasising media literacy, transparency, and global inclusivity (Pfeiffer and Krishna, 2025). Policymakers and the public must move beyond "synthetic or authentic" binaries when considering responses to AI-generated media, recognising that the challenge is not simply to distinguish between real and fake, but to build systems and cultures that can maintain trust and accountability in an age of ubiquitous synthetic media (Tech Policy Press, 2025).

Beyond technological solutions, the legal landscape surrounding deepfakes remains inadequate to address the multifaceted harms they inflict. Ramluckan (2024) argues that existing legal frameworks—including defamation, fraud, and intellectual property law—are ill-equipped to handle the unique characteristics of deepfake-enabled harm, particularly the reputational and psychological damage that occurs even when deepfakes are quickly debunked. The Osbourne case exemplifies this gap: whilst the deepfakes were identified and condemned, the emotional distress to the family and the erosion of public trust in authentic media persisted. Current legislation in most jurisdictions focuses narrowly on pornographic deepfakes or election interference, leaving vast categories of harmful content—such as fabricated health announcements or manipulated celebrity statements—largely unregulated. Ramluckan emphasises that effective legal responses must balance protection against harm with preservation of legitimate creative expression, a challenge that requires nuanced understanding of both the technology and its societal impacts.

6. Conclusion

The proliferation of deepfake technology represents a paradigm shift in the landscape of digital information, posing a formidable threat to public trust and the very notion of objective reality. The case of Ozzy Osbourne's 'death prophecy' serves as a sobering illustration of this threat, demonstrating how synthetic media can be used to manipulate public emotion, amplify misinformation, and exploit personal tragedy. The incident underscores the inadequacy of current detection technologies, which consistently fail to keep pace with the advancements in generative AI, leaving the public vulnerable to increasingly sophisticated and believable forgeries. The

psychological drivers of misinformation, such as FOMO, further exacerbate the problem, creating a fertile ground for sensationalist and unverified content to go viral. The case was widely discussed in policy forums and resembles incidents recorded in the OECD AI Incident Database (OECD AI Policy Observatory, 2025), highlighting the real-world consequences of deepfake technology and the urgent need for comprehensive responses.

The Osbourne case also reveals the temporal dynamics of deepfake misinformation. The first deepfake video circulated in early July 2025, with a TikTok clip appearing on 9 July. Kelly Osbourne's public denial followed on 11 July, but the deepfakes continued to circulate. When Osbourne died on 22 July, the deepfakes gained retrospective credibility, demonstrating how synthetic media can be re-contextualised by real events to become even more potent tools of misinformation. This timeline illustrates the challenge of combating deepfakes in real-time, as the speed of creation and dissemination often outpaces the capacity for detection and debunking. The proliferation of AI-generated tributes and memorial content following Osbourne's death further demonstrates how synthetic media can infiltrate and distort genuine expressions of public grief, creating an environment where authenticity itself becomes suspect.

Addressing the diverse challenges of deepfakes requires a concerted and multi-pronged approach that combines technical innovation, robust governance, and widespread public education. Technical solutions, such as the blockchain-based provenance and authentication standards being developed by the C2PA and proposed by researchers, offer a promising path towards a more transparent and accountable digital ecosystem (Content Authenticity Initiative, 2025; C2PA, 2025; Pfeiffer and Krishna, 2025). The integration of cryptographic hashing, secure metadata anchoring, and digital identity signatures into media-capturing devices, with authenticity markers immutably anchored to public blockchain infrastructures, could create an incorruptible provenance ledger that significantly mitigates the impact of deepfakes (Pfeiffer and Krishna, 2025). However, these technical solutions must be supported by clear and comprehensive policy frameworks that regulate the development and deployment of generative AI, with a particular focus on the ethical sourcing of training data, the mandatory labelling of synthetic media, and the protection of individuals' rights over their own biometric likenesses (Brennan Center for Justice, 2023; Pfeiffer and Krishna, 2025). Legislative initiatives such as Denmark's proposed amendments to copyright laws and Northern Ireland's consultation on criminalising deepfake pornography represent important steps in this direction (The Guardian, 2025; Department of Justice (NI), 2025).

Ultimately, however, the most potent defence against deepfake-driven misinformation lies in a well-informed and critically engaged public. Fostering digital and media literacy on a global scale is therefore an urgent imperative. By equipping individuals with the skills to critically evaluate the information they encounter online, to recognise the psychological vulnerabilities that make them susceptible to manipulation, and to understand the capabilities and limitations of AI technology, we can build a more resilient and discerning public sphere. The Osbourne case demonstrates how synthetic media can exploit personal tragedy to manipulate public sentiment. Addressing this requires the convergence of authentication technologies, legislative foresight, and a critically literate public equipped to navigate truth in a synthetic era

7. Implications for Cybersecurity Professionals

The Osbourne case yields several actionable insights for cybersecurity practitioners and information security teams.

- First, temporal awareness is critical: deepfakes may gain credibility retrospectively when real events appear to validate them, necessitating continuous monitoring even after initial debunking efforts. Security teams should implement event-triggered alert systems that flag resurgent synthetic content following related real-world developments.
- Second, the detection gap identified by Chandra et al. (2025)—where state-of-the-art models experience approximately 50% performance degradation on in-the-wild content—underscores the inadequacy of relying solely on automated detection tools. Organisations should adopt a defence-in-depth approach that combines technical detection with human forensic analysis, provenance verification, and rapid response protocols.
- Third, the psychological dimension of deepfake propagation, particularly FOMO-driven sharing behaviours, suggests that user education and awareness training should be integrated into organisational security programmes. Employees and stakeholders must be equipped to recognise manipulation tactics and resist the impulse to share unverified sensational content. Fourth, the case highlights the value of proactive authentication frameworks: organisations managing public-facing communications should consider implementing C2PA-compliant content credentials and blockchain-

based provenance tracking for official statements, particularly those concerning sensitive topics such as health, security, or financial matters. This creates a verifiable baseline against which deepfakes can be identified.

- Finally, cybersecurity professionals should engage with cross-sector collaboration initiatives such as the Content Authenticity Initiative and contribute to the development of industry standards for synthetic media labelling and detection. The arms race between generation and detection technologies requires collective intelligence sharing and coordinated response mechanisms. Incident response plans should explicitly address deepfake scenarios, including protocols for rapid verification, public communication, legal recourse, and platform coordination to remove malicious content. The convergence of technical sophistication, psychological manipulation, and regulatory gaps demands that cybersecurity practice evolve beyond traditional threat models to encompass the unique challenges posed by synthetic media in the post-truth era.

Ethics Declaration: We did not conduct any research involving user data. We did not receive any funding for writing this paper. Our motivation for writing this academic opinion paper about the Ozzy use case was intrinsic and maybe paranoid in regard to the future development of AI generated deepfakes

AI Declaration: We used the following tools: Consensus APP: for literature research, DeepL Write Pro: for spell-checking and ensuring consistency with British English throughout the text, CoPilot: Not for 'generating', but for challenging our text and thoughts. It was also used for formatting the references list in accordance with the conference guidelines.

This paper is dedicated to Ozzy Osbourne, the Prince of Darkness.

References

Amelinckx, A. (2025) 'The wild death rumors about Ozzy Osbourne that his daughter shut down', *Grunge*, 16 July. Available at: https://www.grunge.com/1914338/wild-death-rumors-ozzy-osbourne-daughter-kelly-osbourne-shut-down/ (Last accessed: 19 October 2025).

Barlow, G. (2025) 'This Ozzy Osbourne tribute video may have 5.8M views, but it just shows everything wrong with AI slop (and what the hell is that music?)', TechRadar, 24 July. Available at: https://www.techradar.com/ai-platforms-assistants/this-ozzy-osbourne-tribute-video-may-have-5-8m-views-but-it-just-shows-everything-wrong-with-ai-slop-and-what-the-hell-is-that-music (Last accessed: 19 October 2025).

BBC News (2025) 'Deepfakes to become criminal offence in NI "sooner rather than later"', *BBC News*, 22 July. Available at: https://www.bbc.com/news/articles/cwyxv9k28evo (Last accessed: 19 October 2025).

BBC News (2025) 'Ozzy Osbourne dies, weeks after farewell show', *BBC News*, 22 July. Available at: https://www.bbc.com/news/articles/cyvjd3p887qo (Last accessed: 19 October 2025).

Brennan Center for Justice (2023) 'Regulating AI deepfakes and synthetic media in the political arena', 5 December. Available at: https://www.brennancenter.org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena (Last accessed: 19 October 2025).

C2PA (2025) *C2PA Specification v2.1*. Available at: https://c2pa.org/specifications/specifications/2.1/specs/C2PA_Specification.html (Last accessed: 19 October 2025).

Chandra, N. A. et al. (2025) 'Deepfake-Eval-2024: A multi-modal in-the-wild benchmark of deepfakes circulated in 2024', *arXiv preprint arXiv:2503.02857*. Available at: https://arxiv.org/abs/2503.02857 (Last accessed: 19 October 2025).

Chesney, R. and Citron, D. K. (2019) 'Deep fakes: a looming challenge for privacy, democracy, and national security', *California Law Review*, 107(6), pp. 1753–1820. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954 (Last accessed: 19 October 2025).

Ciftci, U. A., Demir, I. and Yin, L. (2020) 'FakeCatcher: detection of synthetic portrait videos using biological signals', *IEEE Transactions on Pattern Analysis and Machine Intelligence*. doi: 10.1109/TPAMI.2020.3009287. Available at: https://arxiv.org/abs/1901.02212 (Last accessed: 19 October 2025).

Content Authenticity Initiative (2025) *About the Content Authenticity Initiative*. Available at: https://contentauthenticity.org (Last accessed: 19 October 2025).

Department of Justice (NI) (2025) 'Proposals to criminalise sexually explicit deepfake images'. Available at: https://www.justice-ni.gov.uk/consultations/proposals-criminalise-sexually-explicit-deepfake-images (Last accessed: 19 October 2025).

Entertainment Weekly (2025) 'Ozzy Osbourne says farewell, plays final concert with Black Sabbath to hometown crowd of 40,000, 6 July'. Available at: https://ew.com/ozzy-osbourne-reunites-with-black-sabbath-for-final-concert-11767172 (Last accessed: 19 October 2025).

Gong, C. and Ren, Y. (2023) 'PTSD, FOMO and fake news beliefs: a cross-sectional study of Wenchuan earthquake survivors', *BMC Public Health*, 23(1), p. 2213. doi: 10.1186/s12889-023-17151-z.

Heidari, A., Jafari Navimipour, N., Dag, H. and Unal, M. (2024) 'Deepfake detection using deep learning methods: a systematic and comprehensive review', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 14(2), e1520. doi: 10.1002/widm.1520.

Lead Stories (2025) 'Fact Check: The viral Ozzy "My final performance… every dime went to Parkinson's research" video is NOT authentic', 26 July. Available at: https://www.yahoo.com/news/articles/fact-check-viral-ozzy-final-032503404.html (Last accessed: 19 October 2025).

Lundberg, E. and Mozelius, P. (2025) 'The potential effects of deepfakes on news media and entertainment', *AI & SOCIETY*, 40, pp. 2159–2170. doi: 10.1007/s00146-024-02072-1.

OECD AI Policy Observatory (2025) 'AI-Generated Deepfake Video Falsely Claims Ozzy Osbourne Is Dying', *OECD AI Incidents Monitor*, 12 July. Available at: https://oecd.ai/en/incidents/2025-07-12-45a3 (Last accessed: 19 October 2025).

PCMag (2025) 'I tested out Google's Veo 3 AI video generator. The internet is not prepared', 10 June. Available at: https://www.pcmag.com/opinions/i-tested-out-googles-veo-3-ai-video-generator-the-internet-is-not-prepared (Last accessed: 19 October 2025).

People (2025) 'Kelly Osbourne debunks rumors about dad Ozzy Osbourne's health and clarifies "He's not dying"', 11 July. Available at: https://people.com/kelly-osbourne-debunks-rumors-about-ozzy-osbourne-health-clarifies-hes-not-dying-11770973 (Last accessed: 19 October 2025).

Pfeiffer, A. and Krishna, N. (2025) 'Restoring trust in the age of deepfakes: a blockchain-based proposal – the global challenge of AI-manipulated media', *MAD Opinions*, University for Continuing Education Krems, 30 July. doi: 10.48341/dx7k-e072.

Pistilli, G. and Trevelin, B. (2025) 'Can AI be consentful?', *arXiv preprint arXiv:2507.01051*. Available at: https://arxiv.org/abs/2507.01051 (Last accessed: 19 October 2025).

Ramluckan, T. (2024) 'Deepfakes: the legal implications', in *Proceedings of the 19th International Conference on Cyber Warfare and Security, ICCWS 2024*, pp. 282–288. Available at: https://papers.academic-conferences.org/index.php/iccws/article/view/2099 (Last accessed: 19 October 2025)

Tech Policy Press (2025) 'Synthetic media policy: provenance and authentication', 1 May. Available at: https://techpolicy.press/synthetic-media-policy-provenance-and-authentication-expert-insights-and-questions (Last accessed: 19 October 2025).

The Guardian (2025) 'Denmark to tackle deepfakes by giving people copyright over their faces', 27 June. Available at: https://www.theguardian.com/technology/2025/jun/27/deepfakes-denmark-copyright-law-artificial-intelligence (Last accessed: 19 October 2025).

Verdoliva, L. (2020) 'Media forensics and deepfakes: an overview', *arXiv preprint arXiv:2001.06564*. Available at: https://arxiv.org/abs/2001.06564 (Last accessed: 19 October 2025).

Watson.de (2025) 'Ozzy Osbourne: Fake-Theorien zu seinem Tod – Familie ist "angewidert"', 29 July. Available at: https://www.watson.de/unterhaltung/stars/788233724-ozzy-osbourne-fake-theorien-zu-seinem-tod-familie-ist-angewidert (Last accessed: 19 October 2025).

Yahoo Entertainment (2025) 'Sharon and Ozzy Osbourne had decades-long assisted suicide pact', 22 July. Available at: https://www.yahoo.com/entertainment/articles/sharon-ozzy-osbourne-had-decades-222826539.html (Last accessed: 19 October 2025).