AI-Driven Media Manipulation: Public Awareness, Trust, and the Role of Detection Frameworks in Addressing Deepfake Technologies

Yapay Zeka Destekli Medya Manipülasyonu: Kamuoyu Farkındalığı, Güven ve Deepfake Teknolojilerini Ele Almada Algılama Çerçevelerinin Rolü

Kareem Mohamed[1]
Prof.Dr. Bahire Efe Ozad[2]

Abstract

The proliferation of AI-driven media manipulation technologies,such as deepfakes, voice cloning, and face-swapping,has significantly disrupted public trust and the perceived authenticity of digital media content. This study investigates how affluent European citizens aged 40 and above perceive and respond to such technologies, using a mixed-method design that includes structured surveys and empirical testing via BioID detection software. The study sampled purposively, 51 participants primarily from high-income digital communities in countries including Monaco, France, Germany, Switzerland, and Austria, offering insight into a demographic often targeted by sophisticated AI scams. Findings reveal high exposure to synthetic media (76.5%), widespread distrust in digital content (86.3%), and strong correlations between AI awareness, identity theft, and declining media trust ($x^2 = 25.548$, $p < 0.001$; Cronbach's $\alpha = 0.783$). BioID deepfake detection results (scores = 0.00053 and 0.06223 for fakes; 0.92363 for authentic selfies) confirm the system's reliability in distinguishing manipulated content.

The study uniquely integrates the Theory of Planned Behavior and McLuhan's Medium Theory to analyze both micro-level behavioral responses and macro-level media transformations. TPB explains vulnerability to manipulation through attitudes, social norms, and perceived control, while McLuhan's theory reframes detection tools like BioID as new media forms that mediate perception and redefine authenticity. This dual-theoretical framework reveals how individual cognition and media environments interact to shape trust in an era of AI-generated deepfake media contents. The findings highlight the urgent

[1]Ph.D. Candidate, Faculty of Communication and Media Studies, Eastern Mediterranean University
kareemmohamed25@gmail.com, Orcid ID: 0000-0003-1706-7802
[2]Chair of Radio, Cinema & TV department, Faculty of Communication and Media Studies, Eastern Mediterranean University
bahire.ozad@emu.edu.tr, Orcid ID: 0000-0003-3615-5090

need for targeted digital literacy programs, robust detection frameworks, and regulatory strategies to restore public confidence in digital communication. Limitations include the small and demographically specific sample, indicating a need for broader, cross-cultural research.

Keywords: Deepfakes, Artificial Intelligence, BioID, Media Trust, Theory of Planned Behavior, McLuhan's Medium Theory, Digital Media Literacy, Voice Cloning, AI Detection, Epistemology

Öz

Deepfake'ler, ses klonlama ve yüz değiştirme gibi yapay zekâ destekli medya manipülasyon teknolojilerinin yaygınlaşması, kamuoyunun güvenini ve dijital medya içeriklerinin algılanan özgünlüğünü ciddi şekilde sarsmıştır. Bu çalışma, Avrupa'nın yüksek gelirli dijital topluluklarında yer alan, 40 yaş ve üzeri varlıklı bireylerin bu tür teknolojileri nasıl algıladığını ve onlara nasıl tepki verdiğini incelemektedir. Araştırma, yapılandırılmış anketler ve BioID tespit yazılımı ile gerçekleştirilen deneysel analizleri içeren karma yöntemli bir tasarım kullanmaktadır. Çalışmaya Monako, Fransa, Almanya, İsviçre ve Avusturya gibi ülkelerden gelen toplam 51 katılımcı dâhil edilmiştir. Bulgular, katılımcıların %76,5'inin sentetik medya ile karşılaştığını, %86,3'ünün dijital içeriklere güven duymadığını ve yapay zekâ farkındalığı, kimlik hırsızlığı ve medya güvenindeki azalma arasında güçlü korelasyonlar olduğunu ortaya koymuştur ($x^2 = 25.548$, $p < 0.001$; Cronbach's $\alpha = 0.783$). BioID sistemiyle yapılan deepfake tespiti sonuçları (sahte içeriklerde 0.00053 ve 0.06223; özgün özçekimlerde 0.92363 puanı) yazılımın güvenilirliğini doğrulamaktadır.

Çalışma, bireysel düzeydeki davranışsal tepkileri ve medya sistemindeki yapısal dönüşümleri analiz etmek üzere Planlanmış Davranış Teorisi ile McLuhan'ın Ortam Teorisini birlikte kullanmaktadır. Planlanmış Davranış Teorisi; tutumlar, sosyal normlar ve algılanan kontrol üzerinden manipülasyona karşı bireysel kırılganlığı açıklarken; McLuhan'ın teorisi, BioID gibi tespit araçlarını, algıyı şekillendiren ve özgünlük kavramını yeniden tanımlayan yeni medya biçimleri olarak yeniden konumlandırmaktadır. Bu çift teorik çerçeve, bireysel biliş ile medya ortamlarının etkileşiminin, yapay zekâ tarafından üretilen dezenformasyon çağında güveni nasıl şekillendirdiğini ortaya koymaktadır. Bulgular, hedefe yönelik dijital okuryazarlık programları, sağlam tespit sistemleri ve düzenleyici stratejilerin acil gerekliliğini vurgulamaktadır. Çalışmanın sınırlılıkları arasında küçük ve demografik olarak özgül bir örneklem yer almakta olup, daha geniş ve kültürlerarası araştırmalara ihtiyaç duyulmaktadır.

Anahtar Kelimeler: Deepfake, Yapay Zekâ, BioID, Medya Güveni, Planlanmış Davranış Teorisi, McLuhan'ın Ortam Teorisi, Dijital Medya Okuryazarlığı, Ses Klonlama, AI Tespiti,

Epistemoloji

1. Introduction

Artificial intelligence technologies, techniques, and tools have revolutionized the communication and media studies field. Innovations like voice cloning, Deepfake videos, instant face swapping, their potential for manipulation and even creative expressions have gathered significant attention. Voice cloning is where we can create a synthetic voice that accurately resembles a specific vocal characteristic of an individual through tools like iSpeech and Descript, facilitating the development and accessibility. Machine learning (ML) algorithms can be utilized to create media content that can manipulate reality, often without the knowledge of users, viewers, and audiences, as deepfake technologies (Ng, 2024) and (Al-Khazraji, 2023). Software such as Reface aligns as an instant face-swapping tool that seamlessly interchanges faces in videos and photos, blurring the lines between fabrication and reality. The evolving landscape of AI-driven media manipulation necessitates situating this research within established academic discourse on media epistemology, trust dynamics, and digital publics. Recent scholarship on media epistemology has explored how emerging technologies fundamentally alter the ways in which knowledge claims are constructed, evaluated, and verified in digital environments. (Carlson, 2020) examines how deepfake technologies disrupt traditional epistemological frameworks by challenging the evidentiary value of audiovisual media that has historically served as a cornerstone of truth verification. His analysis demonstrates how the proliferation of synthetic media technologies creates what he terms "epistemic instability," where audiences must develop new literacy practices for navigating increasingly uncertain information landscapes. Similarly, (Fallis, 2021) explores the epistemological implications of deepfakes through the lens of testimonial knowledge, arguing that the democratization of media manipulation technologies fundamentally alters the relationship between seeing and believing in digital contexts, concluding that emerging verification frameworks must address not only technical detection capabilities but also the social practices through which audiences assign credibility to mediated evidence.

The erosion of trust in media institutions represents a central concern within media sociology, particularly as AI manipulation technologies become more sophisticated and accessible. Couldry and (Mejias, 2019) frame deepfake technologies within their broader analysis of "data colonialism," positioning synthetic media as an extension of computational logics that increasingly colonize social relationships and undermine traditional trust structures. Their work emphasizes how the technical capabilities for media manipulation intersect with existing power asymmetries, potentially amplifying distrust among already marginalized populations. (Livingstone & Lunt, 2023) offer complementary insights through their examination of how different demographic groups develop "trust calibration" strategies in response to potential misinformation, finding that older adults often rely on

heuristic judgments that may be particularly vulnerable to sophisticated manipulation. Their research suggests that age-specific media literacy interventions may be necessary as synthetic media becomes increasingly prevalent across communication channels that seniors regularly engage with. The conceptualization of digital publics provides essential context for understanding how AI manipulation technologies reshape collective engagement with mediated content. (Papacharissi, 2021) examines how deepfakes and synthetic media contribute to what she terms "affective publics," wherein emotional resonance often supersedes factual accuracy in determining content circulation and belief. Her analysis suggests that effective interventions must address not only cognitive verification but also the affective dimensions of engagement with potentially manipulated content. Building on this work, (Shah et al., 2023) investigate how awareness of manipulation capabilities influences civic participation across different demographic groups, finding that knowledge of deepfake technologies correlates with decreased political engagement among older adults specifically. This research highlights the potential democratic implications of trust erosion resulting from emerging AI technologies. Similarly, (Gillespie, 2022) examines how platform governance approaches to synthetic media create new forms of "platform publics" with distinct verification norms and trust practices, offering valuable insights into how institutional responses shape public understanding of and engagement with potentially manipulated content across different demographic segments.

The proliferation of accessible AI-driven media manipulation technologies presents significant implications for public trust in digital content. Tools for synthetic voice generation (such as ElevenLabs and Play.ht), facial replacement (exemplified by various face-swapping applications), video dubbing (like Rask.AI), and lip synchronization technologies (SyncLabs) collectively represent a new frontier in media manipulation capabilities. These platforms operate through similar mechanisms—utilizing neural networks trained on extensive datasets to analyze, decompose, and reconstruct media elements—while requiring minimal technical expertise from users. The democratization of these once-specialized capabilities raises important questions regarding media literacy and trust in an increasingly synthetic information environment.

Figure 1: AI Deepfake Media Softwares

Advancements in artificial intelligence have revolutionized digital media production, enabling the seamless generation and modification of audiovisual content through sophisticated deep learning models. These technologies allow users to create highly realistic voice clones, dubbed videos, lip-synced animations, and face-swapped visuals, offering significant applications in media, entertainment, and accessibility while raising ethical concerns regarding misinformation and digital identity manipulation (Chesney & Citron, 2019). Voice cloning technologies enable users to replicate speech patterns by selecting a predefined voice model or uploading a recorded sample, refining parameters such as speed and pitch to enhance the output. Similarly, video dubbing tools facilitate the translation of audiovisual content by allowing users to upload a video, select an original audio track, specify a target language, and adjust voice styles before finalizing the dubbed version (Vaccari & Chadwick, 2020). Lip synchronization software refines audiovisual realism by aligning lip movements with an audio track, ensuring accurate speech synchronization through algorithmic adjustments. Additionally, face-swapping tools enable users to merge facial features from one individual onto another in a video, requiring precise positioning for a seamless, natural appearance (Mirsky and Lee, 2021). While these technologies enhance creative possibilities and cross-lingual communication, their increasing accessibility underscores the importance of rigorous analysis regarding their impact on media authenticity, audience perception, and ethical considerations (Zellers et al., 2019).

These technologies, while innovative, introduce complex challenges to established verification frameworks and public trust dynamics. Voice synthesis tools can generate realistic speech mimicking specific individuals from limited audio samples, while facial manipulation technologies enable the convincing transposition of identities within video content. Similarly, dubbing and lip synchronization tools facilitate the alteration of spoken content while maintaining visual coherence, effectively circumventing traditional markers of media authenticity. The technical capabilities of these tools continue to advance rapidly, progressively narrowing the perceptible gap between authentic and synthetic media. This technological evolution necessitates corresponding advancements in both detection frameworks and public awareness to maintain functional information ecosystems in which veracity can be reasonably assessed and trust appropriately calibrated.

While all these advanced technologies and innovations present many potential possibilities, they also pose significant limitations, ethical considerations, challenges, and risks to public trust and awareness in the ongoing development of digital media content by media companies, professionals, and organizations. Linking to various malicious activities, including misinformation campaigns, reputational harm, and identity theft through the misuse of deepfake technologies, where the potential to influence public opinion and ruin democratic processes leads to media trust erosion and a crisis of trust in generated digital media contents through different platforms and associations (Karnouskos, 2020) and

(Sharma, 2022). The main ramifications of such manipulations in distorting the reality perception and eroding the trust in legitimate media sources and contents where the public awareness of the demonstration or knowledge of AI techniques and technologies remains alarmingly low as a study in 2020 by the Pew Research Center found that most of the participants there in the research often lack the knowledge to identify the manipulated content even though they are aware of deepfakes (Vizoso, 2021) and (Nowroozi, 2022).

Extending beyond individual experiences In various sectors such as politics, media, journalism and personal business relationships where, the implications of AI deception technologies influencing the perceptions of reality, trust and authenticity reverberating media and society through manipulating audio and visual contents raises ethical considerations about the information integrity. The misuse of deepfake videos in creating misleading political campaigns and ads swaying voter and public opinions based on fabricated and manipulated content poses a significant threat to democracy as misinformation often surpasses efforts to expose false narratives that possibly spread rapidly through different digital media platforms (Amerini, 2024), (Temir, 2020) and (Oza, 2024).

Underestimating the psychological impact of deepfake-generated media content cannot last long as research by (George, 2023) and (Nasar, 2020) has shown the leading of increased scepticism towards legitimate media sources of information by the use of manipulation and fabrication fostering the media trust erosion and the long-lasting effects on societal cohesion, functioning of media associations, professionals and democratic institutions in addition to the ethical concerns in personal contexts through reported cases of identity theft to impersonate individuals authenticity for financial gains. All of those scenarios underscore the necessity and urgent need for individuals to understand the capabilities of AI deception technologies and their misuse, as well as the potential risks and challenges associated with public education and awareness campaigns. It has been crucial to develop effective, accurate detection software and detection frameworks to identify fabricated media contents through leveraging ML Algorithms and AI such as Deep-ware scanners and Sensity AI detecting inconsistencies in audio-visual contents in deepfakes from authentic media (Shakil, 2024) and (Chapagain, 2024). The effectiveness of those techniques depends heavily on public exposure, understanding, education, knowledge and awareness of AI technology's functionality and existence. Expected growth of 35% to the CAGR (compound annual growth rate) of the global broad market for artificial intelligence-generated media content driven by computer vision, machine learning algorithms, and natural language processing applications and techniques as retrieved by (Future, 2023), where a survey conducted in 2023 by the Digital Media Research Group showed that only 30% of the participants could identify deepfake videos, as public awareness remains low among older demographics who are less familiar with digital media fabricated content. In

particular, those aged between 18 and 24 accurately identified 55% of manipulated content, demonstrating higher awareness and recognition. This highlights the urgent need for targeted educational campaigns to enhance and increase awareness across all age groups by policymakers, technologies, stakeholders, media professionals, and ethicists through establishing guidelines balancing innovations and AI-driven technologies with digital literacy, fraud prevention, and public trust (Tuysuz, 2023), (Fabuyi, 2024) and (Esezoobo, 2023).

In light of these concerns, the present study aims to critically examine the evolving relationship between AI-driven media manipulation technologies and public trust, with a particular focus on affluent European citizens aged 40 and above. Central to this investigation is an exploration of how individuals within this demographic perceive, comprehend, and react to emerging deception tools such as voice cloning, deepfake videos, video dubbing, and face-swapping. Using three structured questionnaires; each targeting awareness and exposure, trust in digital content, and experiences with identity theft, the research seeks to assess how varying degrees of familiarity with synthetic media technologies influence shifts in perception, skepticism, and media engagement. Grounded in the Theory of Planned Behavior and McLuhan's Medium Theory, the study also evaluates the role of AI detection tools like BioID in supporting verification practices and restoring digital confidence. By combining empirical evidence with theoretical insights from media studies and behavioral science, this research contributes to a deeper understanding of the complex interplay between technological innovation, media credibility, and public literacy, ultimately offering guidance for educational initiatives, regulatory policies, and platform governance aimed at safeguarding trust in the digital information ecosystem.

1. Introduction

The rapid advancement and democratization of AI-driven media manipulation technologies present multifaceted challenges to public trust and information integrity. As technologies for voice cloning, deepfake video generation, video dubbing, and face-swapping become increasingly accessible and sophisticated, they fundamentally alter the relationship between media consumers and digital content. This study addresses three critical problems: First, there exists a significant knowledge gap regarding awareness and understanding of AI deception technologies among older demographic groups, particularly affluent European citizens aged 40 and above, who may be specifically targeted for financial scams using these technologies. Despite being potential high-value targets for voice cloning scams and deepfake-related identity theft, this demographic has received limited research attention regarding their vulnerability to and awareness of these emerging threats. Second, there is insufficient empirical evidence documenting the concrete impact of exposure to deepfake technologies on trust erosion in digital media. While theoretical concerns about declining trust are widespread, quantitative measurements establishing the correlation between

awareness of manipulation capabilities and resulting changes in media trust remain underdeveloped, particularly among this demographic cohort. Third, current detection frameworks and educational approaches face substantial implementation challenges. Although technical solutions for identifying synthetic media exist, their practical integration into everyday media consumption practices remains problematic. This study examines the accessibility and effectiveness of detection tools like BioID software while addressing the disconnect between technical capabilities and practical application in protecting vulnerable populations.

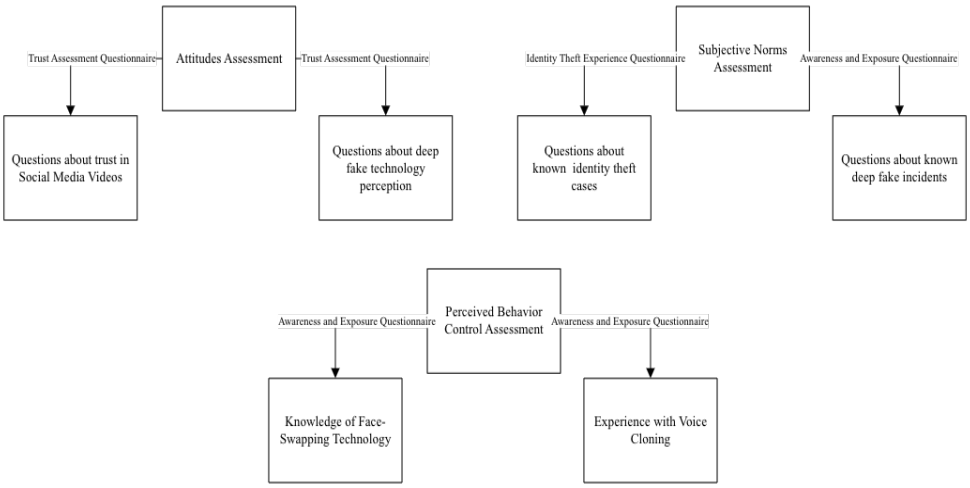
1.2. Aim of the Study

The present study aims to critically examine the evolving relationship between AI-driven media manipulation technologies and public trust, with a particular focus on affluent European citizens aged 40 and above. As central to the research is an exploration of how individuals within this demographic perceive, understand, and respond to emerging AI deception tools such as voice cloning, deepfake videos, video dubbing, and face-swapping. The study is designed to evaluate the extent of public awareness, the psychological and behavioral implications of exposure to synthetic media, and the broader societal impact of AI-driven manipulation. The reserach implemented three structured questionnaires, each consisting of five questions; focused on measuring awareness and exposure, trust in digital media content, and experiences with identity theft. Through these instruments, the research seeks to uncover how different levels of familiarity with AI technologies correlate with shifts in perception, trust, and media engagement through the lens of theory of planned behavior. Additionally, the study evaluates the role of AI detection tools, particularly BioID software, in supporting public efforts to authenticate digital content and restore confidence in online communication environments through the lens of Mcluhan's medium theory. By integrating empirical data with theoretical perspectives from media and behavioral studies, the study contributes to a deeper understanding of the intersection between technological innovation and public trust, aiming to inform educational, technological, and regulatory strategies for mitigating the risks posed by synthetic media in the digital age.

1.3. Aim of the Study

1.3.1. Aim of the Study1.2. Aim of the Study

Figure 2: Application of TPB to the survey



The Theory of Planned Behavior (TPB) provides a valuable framework for interpreting the results of this study on AI deception technologies. TPB (Ajzen, 2011) posits that the individual's intentions determine their behaviours, which are influenced by subjective norms, attitudes, and perceived control of behaviour.

1.3.1.1 Attitudes

The Trust Assessment questionnaire in this study effectively gauged participants' attitudes towards AI technologies and digital media content. For instance, "Do you trust videos shared on social media?" directly assesses attitudes that could influence behaviour related to engaging with or sharing digital content. Also, another question, "Do you believe deepfake technology can be used for positive purposes?" reflects attitudes towards the technology itself, which may impact how individuals interact with or respond to deepfake content. TPB states that these attitudes are crucial in shaping behavioural intentions related to AI deception technologies.

1.3.1.2 Subjective Norms

The study addressed the subjective norms component of TPB through questions about participants' social circles and their experiences with AI technologies. For example, the question "Do you know someone who has experienced identity theft due to AI technologies?" from the Identity Theft Experiences questionnaire relates to social norms that can influence an individual's perception and behaviour. Another question, "Have you heard of someone being misled by a deepfake or voice clone?" from the Awareness and Exposure questionnaire, also taps into social norms. As posited by TPB, these social influences can significantly impact an individual's intentions to protect themselves from or engage with AI deception technologies.

1.3.1.3 Perceived Behavioral Control

The Awareness and Exposure questionnaire assessed participants' knowledge of AI technologies related to their perceived ability to control or respond to these technologies. Questions such as "Do you know what face-swapping technology is and how to generate deepfake videos?" and "Have you seen a voice cloning demonstration or learned how to generate it?" According to TPB, these items gauge the participants' perceived knowledge and skills, contributing to their perceived behavioural control in dealing with AI deception technologies.

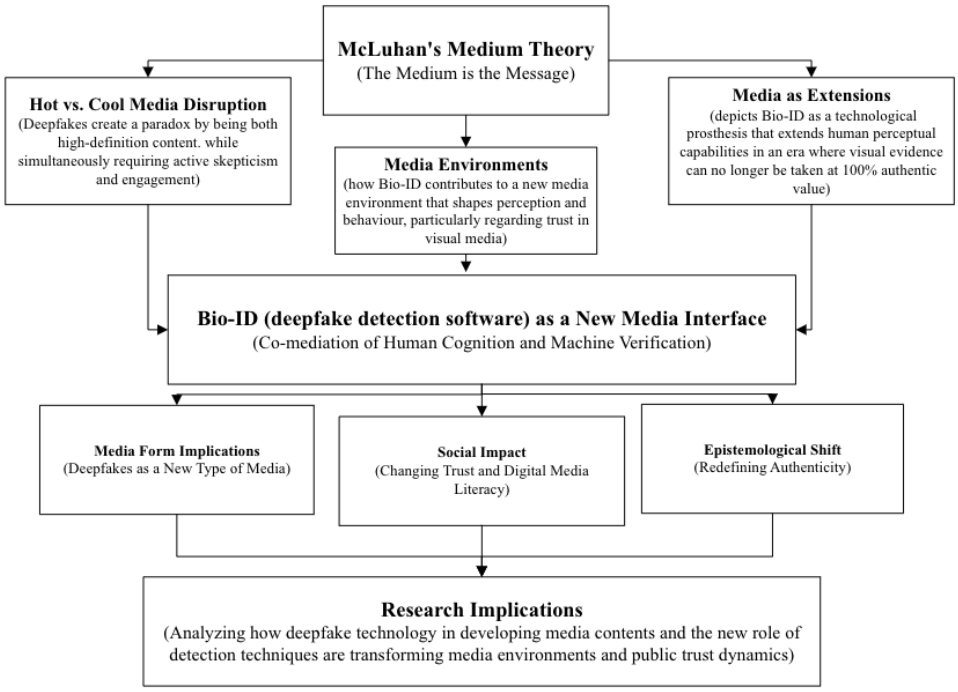
1.3.1.4 Behavioural Intention and Actual Behavior

The combination of attitudes, norms, and perceived control assessed in the questionnaires can predict intentions towards engaging with or protecting oneself from AI deception technologies. Moreover, some questions provided insights into actual behaviours: "Have you ever shared a video that you later discovered was a deepfake?" and "Have you ever reported a scam related to deepfake technology?" According to TPB, these questions offer a glimpse into the participants' actual behaviours, which are the outcome of attitudes, norms, perceived control, and intentions.

1.3.2 Mcluhan's Medium Theory

Marshall McLuhan developed Medium Theory in 1964 in his book Understanding Media: The Extensions of Man, which offers a foundational approach to understanding how the form of media, not merely its content; affects human perception, behavior, and social structures. McLuhan's idea: "the medium is the message," suggests that the characteristics of a medium have a more significant impact than the information it conveys. In the context of synthetic media, particularly AI-generated video and audio content; Medium Theory provides a crucial lens through which to interpret how these emerging technologies reshape communication, trust, and audience engagement. McLuhan's categorization of media into "hot" and "cool" helps frame the experiential dimensions of deepfake content. Hot media, such as cinema or high-resolution video, require less audience participation due to their high sensory data, while cool media, like telephone or low-definition visuals, demand more active involvement. However, with the rise of deepfake technologies, this boundary becomes increasingly ambiguous. AI-generated content often mimics high-definition, emotionally resonant audiovisual formats; appearing "hot", yet paradoxically demands heightened viewer skepticism and cognitive involvement, characteristics of "cool" media. As such, synthetic media challenge McLuhan's original classification, suggesting the need for new interpretations within Medium Theory. The Trust Assessment questionnaire in

Figure 3: Application of Mcluhan's Medium Theory



this study Central to this research is the role of Bio-ID, a biometric-based detection tool used to analyze facial authenticity in video content. From a Medium Theory perspective, Bio-ID does not simply function as a detection mechanism; it represents a new layer of mediation between the viewer and the content. Its integration reflects a broader shift in the media environment, where artificial intelligence is not only producing content but also serving as a gatekeeper for verifying its legitimacy. This aligns with McLuhan's view that media evolve into environments that shape our sensory experiences and epistemological frameworks. By extending the sensory boundaries of human observation, tools like Bio-ID serve as technological prostheses; extensions of perception in a media landscape where visual and auditory evidence can no longer be taken at face value. McLuhan emphasized that every new medium introduces a new environment and displaces older ways of knowing and communicating. In this regard, AI-driven detection tools such as Bio-ID mark a transformation in how society navigates truth, authenticity, and digital literacy. Thus, this theoretical framework positions deepfake and AI-generated content not merely as technological phenomena, but as cultural and epistemological shifts. Medium Theory provides a conceptual structure to explore these shifts, while the incorporation of biometric detection systems such as Bio-ID signals the emergence of a new kind of media interface; where perception is co-mediated by human cognition and machine verification.

By integrating McLuhan's Medium Theory with the Theory of Planned Behavior, this research bridges the gap between individual cognitive responses such as awareness, trust, and verification efforts, and the structural conditions shaped by evolving media

technologies. This theoretical combination enables a holistic understanding of both the micro-level behavioral dynamics and the macro-level transformations within the digital information ecosystem, offering a nuanced view of how AI-generated media is reshaping both perception and society.

1.4 Research Questions

Based on the identified research problem, aims, and theoretical framework, this study addresses the following research questions:

RQ1: To what extent are affluent European citizens aged 40 and above aware of AI-driven media manipulation technologies, and how does this awareness correlate with their levels of trust in digital media content?

RQ2: How do experiences with AI deception technologies influence media consumption behaviors among affluent European citizens, and how can these behavioral adaptations be understood through the Theory of Planned Behavior framework?

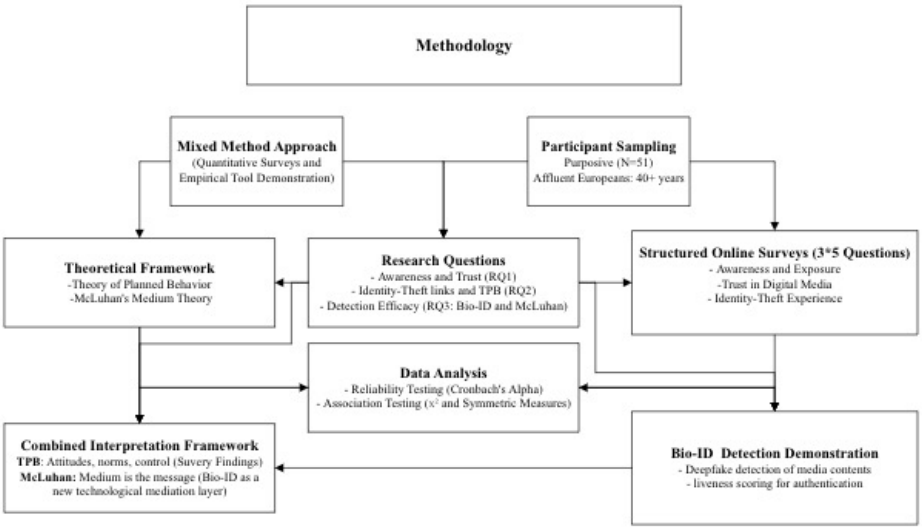
RQ3: How effective is BioID software in detecting synthetic media content, and what insights does McLuhan's Medium Theory provide about the role of detection technologies in reshaping trust and perception in digital environments?

2. Methodology

2.1. Research Design

This study employed a mixed-method approach to examine the relationship between AI-driven media manipulation and public trust among affluent European citizens aged 40 and above. The primary data source consisted of three structured quantitative questionnaires designed to assess participants' attitudes, perceived behavioral control, and social influences, as guided by the Theory of Planned Behavior (TPB). Statistical analysis was conducted using IBM SPSS, applying tests such as Cronbach's Alpha for reliability and Chi-Square for association. Additionally, the BioID detection tool was used to demonstrate and validate the identification of manipulated media, integrating McLuhan's Medium Theory to assess the broader media environment and cognitive impact.

Figure 4: A Flowchart of The Present Methodology of The Study



2.2. Research Design

The present study employed purposive sampling to recruit 51 senior European citizens aged 40 and above from Austria, Switzerland, Netherlands, Croatia, Germany, France, The United Kingdom, Luxembourg, Liechtenstein, Belgium, Norway, Sweden, Denmark, Iceland, and Monaco. Participants were selected from online communities and interest groups associated with luxury lifestyles, including golf equipment, high-end automobile parts (e.g., Porsche and Ferrari), diamonds, gold, and casino gambling. This demographic was intentionally chosen due to their higher likelihood of being targeted by sophisticated AI-driven scams, such as voice cloning fraud and deepfake impersonation. Affluent individuals are often viewed as high-value targets for financial exploitation, making them more susceptible to scams that leverage advanced manipulation techniques. By focusing on this group, the study aims to investigate how digital deception technologies influence trust, awareness, and perceived vulnerability in populations that are both financially exposed and active within digitally mediated environments.

2.3. Data Collection

Data was collected between November 27 and December 26, 2024, through three structured online questionnaires, each consisting of five questions. The questionnaires were distributed via Facebook, Gmail, and Instagram in English. The study specifically targeted participants from the previously identified luxury-interest groups, ensuring relevance to the research objective of examining trust erosion and awareness in the context of high-risk, high-income populations. These communities were selected not only for their demographic characteristics but also for their increased exposure to AI scams that exploit personal voice samples, social media presence, and lifestyle cues. By focusing on these online spaces, the study ensures a practical context in which AI-driven manipulation has tangible consequences. Ethical

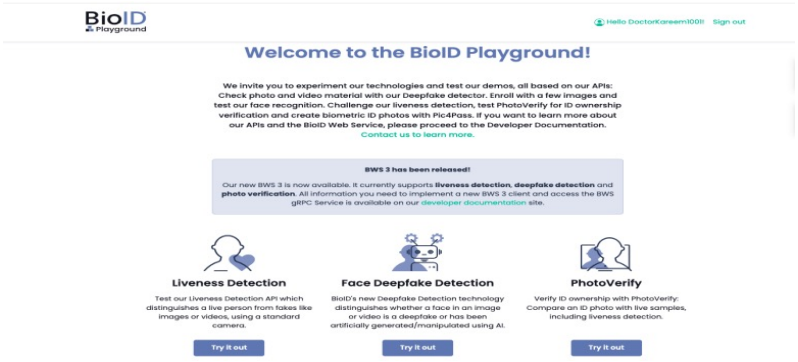
clearance was granted by Eastern Mediterranean University, and informed consent was obtained from all participants, with full assurances of confidentiality and anonymity.

2.4. Data Analysis

The research used a mixed-methodology approach to analyze data collected from 51 senior European citizens. The analysis focused on three key areas: trust assessment in digital media, experiences of identity theft, and awareness and exposure to AI deception technologies. Quantitative data was analyzed using IBM SPSS software (Field, 2024), applying descriptive statistical methods such as reliability analysis with Cronbach's Alpha and association testing through Chi-Square and symmetric measures. The process included data normalization, cleaning, and cross-verification to ensure accuracy and consistency. In addition to the statistical analysis, the study was framed by two theoretical perspectives. The Theory of Planned Behavior (TPB) was used to interpret participant's attitudes, subjective norms, and perceived behavioral control in relation to AI technologies and trust in digital media. Meanwhile, McLuhan's Medium Theory guided the analysis of how media technologies, independent of content; reshape perception and communication environments. This framework was particularly applied in evaluating the use of BioID detection software, not only as a technical tool for identifying synthetic media but as a medium in itself that extends perceptual faculties and influences how individuals engage with authenticity in digital content. By combining quantitative tools with these theoretical lenses, the study offers both empirical rigor and a deeper understanding of how AI deception technologies affect human behavior and media trust.

2.5. Bio-ID as Detection Solution of Deepfake Media Contents

Figure 5: Bio-ID Dashboard

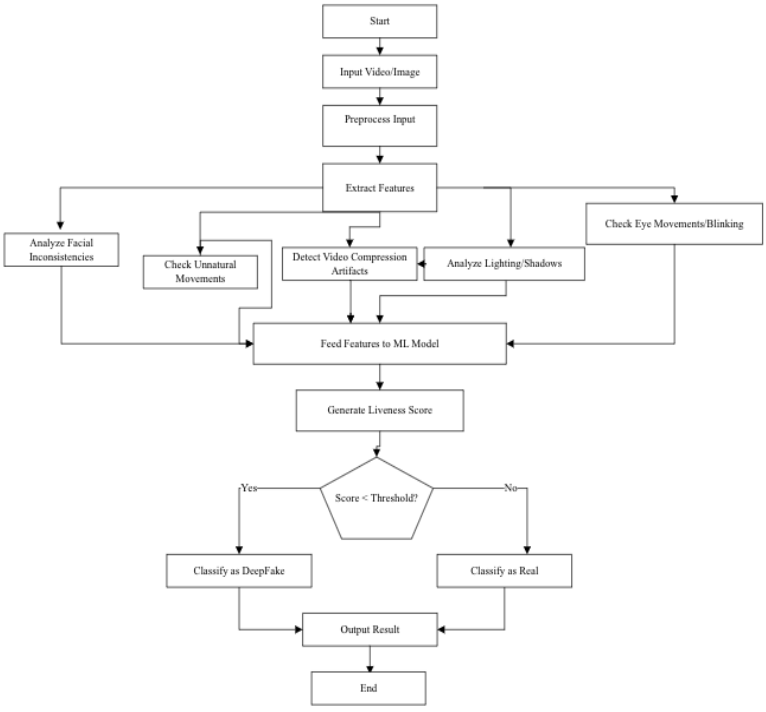


This present research section discusses the procedural framework for using BioID detection software (playground.bioid.com), focusing on its role in biometric authentication and deepfake detection. BioID is an advanced software incorporating state-of-the-art algorithms for liveness detection, image-based recognition, and video deepfake identification (Software,

2024). For image detection and video fake detection, users must upload a file (either an image or a video) for analysis. The system will then process the media using its built-in algorithms to determine whether the content is authentic or manipulated. The analysis typically takes place in real-time, and results are provided within seconds, allowing users to proceed with further steps, if necessary, swiftly. Users must take two live selfies for liveness detection, which the system will process in real time. Alternatively, users can connect with a virtual camera on WhatsApp, Microsoft Teams, Google Meet, or Zoom. The software will analyze the live input instantaneously, using algorithmic techniques to confirm whether the individual in front of the camera is a real, live person or a digital imitation.

The BioID platform (playground.bioid.com) was used to demonstrate real-time AI detection capabilities. Participants or researchers could upload videos or images for analysis. The system processed the content to detect deepfakes by examining features like unnatural facial movements, lighting inconsistencies, and compression artifacts. It assigned a liveness score to determine whether the content was real or synthetic. The tool also supported live video detection through virtual meeting platforms. This tecnique helped in illustrating the potential of biometric AI tools in mitigating trust erosion in digital environments and aligned with McLuhan's concept of media shaping perception regardless of content.

Figure 6: Bio-ID Deepfake Detection: A ML-Based Approach for Videos and Images Analysis



The Bio-ID deepfake detection process begins by receiving a video or image input, which is then preprocessed for optimal feature extraction. Key features such as facial inconsistencies, unnatural movements, video compression artefacts, lighting and shadow anomalies, and irregular eye movements are analyzed. These features are fed into a machine learning model that generates a liveness score, indicating the likelihood of authenticity. If the score falls below a set threshold, the content is classified as a deepfake; otherwise, it is considered authentic. The final result is then outputted, completing the detection process.

3. Results and Discussion

3.1. Survey's Findings

The Awareness and Exposure survey of 51 respondents found that 76.5% had encountered a deepfake video, suggesting the widespread infiltration of manipulated content into everyday digital experiences. This high exposure rate implies not only technical advancement but also insufficient media literacy defenses, particularly among older adults and individuals from non-technical backgrounds, who may lack the tools or training to question audiovisual content critically (Survey, 2024). Significantly, 23.5% had unknowingly shared a deepfake, emphasizing the ease with which misinformation propagates in echo chambers of digital platforms. This finding points to a sociotechnical vulnerability: even informed users may become unintentional vectors of disinformation, raising questions about the adequacy of current awareness campaigns and the usability of content authentication tools. Only 36.3% were aware of face-swapping technology, and an even smaller group (25.5%) had seen a voice cloning demonstration. This suggests a knowledge asymmetry, where technical tools outpace public understanding. Such gaps are particularly dangerous for older or less digitally fluent populations, who may not recognize subtle manipulations; making them more susceptible to impersonation scams and emotional manipulation.

The Trust Assessment Survey provides further evidence of public unease: only 13.7% trusted videos on social media, and 11.8% believed most news videos were genuine. An overwhelming 98% of respondents doubted the authenticity of online videos and voice messages. This erosion of trust marks a cognitive and emotional shift; moving from a passive belief in digital content toward a pervasive skepticism that may result in both paralysis (inaction) or overcorrection (false positives in distrust) (TrustAssessmentSurvey, 2024). Interestingly, 72.5% of participants did not believe deepfake technology could be used positively, reflecting a narrative dominated by fear and fraud. This sentiment likely emerges from repeated exposure to alarming use-cases; scams, defamation, and political manipulation; rather than exposure to its constructive potential in accessibility, historical restoration, or entertainment. (TheftExperienceSurvey, 2024). Socioeconomic implications are also relevant, those in higher-income brackets or luxury groups, often targets of financial fraud, may be more vulnerable to AI deception due to the targeted sophistication of voice cloning and impersonation attacks. Scammers tailor their approaches based on available

financial data, making affluent individuals prime victims. Conversely, less affluent populations may lack access to detection tools and training, reinforcing digital inequality in navigating AI risks.

3.2. Statistical Analysis

3.2.1. Reliability Analysis

The reliability analysis of the questionnaires involved 54 cases, with 51 valid responses (94.4%) and three excluded due to missing data (5.6%). This exclusion is a common practice to ensure the analysis is based on complete data, maintaining the integrity of the results.

Figure 7: Reliability Analysis Cronbach's Alpha Results

Reliability

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	51	94.4
	Excluded[a]	3	5.6
	Total	54	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.783	15

A value above 0.7 is acceptable, reinforcing that the questionnaire reliably captures the intended dimensions. The Cronbach's alpha value of 0.783 indicates good internal consistency among the 15 questions across the three surveys. This statistical measure suggests the items are well-correlated in assessing awareness, trust, and identity theft experience constructs. The Cronbach's alpha value of 0.783 indicates good internal consistency among the 15 questions across the three surveys. This statistical measure suggests that the items are well-correlated in assessing awareness, trust, and identity theft experiences. A value above 0.7 is acceptable, reinforcing that the questionnaire reliably captures the intended dimensions. The results provide a robust foundation for interpreting the responses, allowing the study to draw meaningful conclusions about senior citizens' perceptions and experiences with AI deception technologies.

3.2.2. Reliability Analysis

Figure 8: Chi-Square Test Results

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	25.548[a]	1	<.001		
Continuity Correction[b]	21.911	1	<.001		
Likelihood Ratio	24.948	1	<.001		
Fisher's Exact Test				<.001	<.001
Linear-by-Linear Association	25.047	1	<.001		
McNemar Test				.219[c]	
N of Valid Cases	51				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 3.24.
b. Computed only for a 2x2 table
c. Binomial distribution used.

The results of various statistical tests, particularly chi-square tests, reveal significant associations between categorical variables. The Pearson Chi-Square test, with a statistic of 25.548 and a p-value below 0.001, strongly rejects the null hypothesis of no association. Similarly, the continuity correction (Yates' correction) and the likelihood ratio tests yield p-values below 0.001, confirming significant results. Fisher's exact test, suited for small sample sizes, also indicates significant associations with p-values under 0.001. The linear-by-linear association test shows a significant linear relationship with a statistic of 25.047 and a p-value below 0.001. However, the McNemar test, which assesses changes in paired categorical data, shows no significant change, as its p-value is 0.219.

Figure 9: Symmetric Measures Results

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Nominal by Nominal	Phi	.708			<.001
	Cramer's V	.708			<.001
Interval by Interval	Pearson's R	.708	.107	7.013	<.001[c]
Ordinal by Ordinal	Spearman Correlation	.708	.107	7.013	<.001[c]
N of Valid Cases		51			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.
c. Based on normal approximation.

The symmetric measures used in this analysis reveal strong associations between the variables, as evidenced by the Phi coefficient, Cramer's V, Pearson's R, and Spearman's correlation, all of which share a value of 0.708. This value, consistent across various measures, indicates a strong positive relationship between the analyzed variables, with high statistical significance (p-values below 0.001). These findings suggest significant connections between trust assessments, awareness, and exposure to AI technologies, particularly in the context of video manipulation and identity theft. For example, respondents with greater awareness of AI-based video manipulation technologies may express more concern about identity theft risks. In contrast, those with higher exposure to these technologies might report lower trust levels. The non-significant McNemar test result ($p = 0.219$) suggests no notable change in paired responses, highlighting consistency in participant views across different study aspects. Overall, these results emphasize the interrelated nature of trust, exposure to AI technologies, and identity theft experiences, providing a strong foundation for further

research on detecting and preventing fake videos. Additionally, these associations highlight the importance of considering these factors in discussions about trust and emerging AI technologies, particularly regarding their societal implications through the digital media landscape.

3.3. Implications of Utilizing the Theory of Planned Behavior application to the research findings

The Theory of Planned Behavior (TPB) provides a robust framework for interpreting the complex interrelationships revealed in this study between awareness of AI deception technologies, trust in digital content, and experiences with identity theft. The statistically significant correlations observed ($x^2 = 25.548$, df = 1, $p < 0.001$, Phi/Cramer's V = 0.708) can be systematically unpacked through the TPB lens to derive deeper insights into cognitive processes and behavioral patterns.

3.3.1. Attitudes Toward the Behavior

The study's finding that 98% of respondents doubted the authenticity of online videos and voice messages represents a profound negative attitudinal shift toward digital content. Through the TPB framework, this can be interpreted as a defensive cognitive adaptation: respondents are developing increasingly skeptical attitudes as a protective mechanism. This skepticism functions as an attitudinal shield against potential manipulation. The remarkably low trust rates, only 13.7% trusting social media videos and 11.8% believing news videos are genuine; suggest an emerging "presumption of inauthenticity" as the default attitudinal stance. This represents a fundamental reversal of traditional media consumption patterns, where content was historically presumed authentic until proven otherwise. These attitudinal shifts are particularly significant because TPB posits that attitudes are direct determinants of behavioral intentions. The near-universal distrust documented in our survey suggests we may soon observe widespread behavioral changes in how individuals consume, share, and respond to digital content.

3.3.2. Subjective Norms

The finding that 23.5% of respondents had unknowingly shared deepfake content indicates a complex relationship with subjective norms. TPB explains that behavioral decisions are significantly influenced by perceived social pressures and normative beliefs about what others expect. In the context of the research, the high unintentional sharing rate suggests a normative conflict within contemporary digital media practices. This tension manifests between emerging cautionary norms; reflecting a growing collective awareness that digital content warrants critical scrutiny, as evidenced by the substantial proportion of respondents having encountered deepfakes, and persistent sharing norms embedded in established social media behaviors that prioritize immediacy and engagement through rapid information dissemination without verification protocols. This sociocultural contradiction

represents a transitional phase in digital literacy wherein recognition of manipulation risks coexists uncomfortably with behavioral patterns optimized for pre-deepfake information environments, creating cognitive dissonance among users navigating increasingly complex media landscapes.

This normative conflict creates a cognitive dissonance, where individuals may intellectually recognize the risks but behaviorally default to established social media sharing patterns. The strong correlation values (Pearson's R = 0.708) between trust assessments and exposure to AI technologies further supports this interpretation, suggesting that subjective norms are in flux as communities collectively navigate changing digital landscapes.

3.3.3. Perceived Behavioral Control

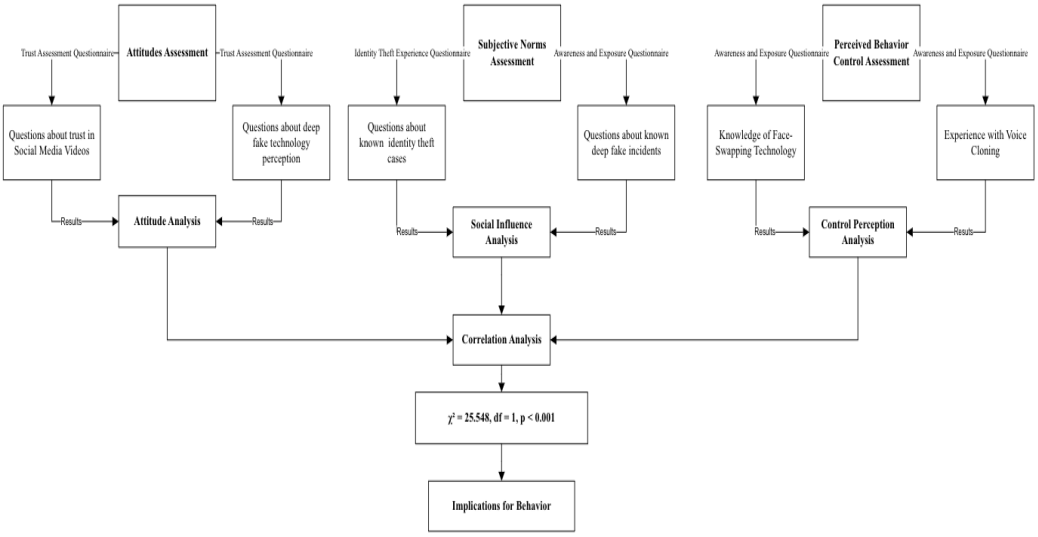
The TPB component of perceived behavioral control is particularly illuminated by our findings on knowledge asymmetry. With only 36.3% aware of face-swapping technology and 25.5% familiar with voice cloning demonstrations, many respondents lack the technical knowledge that would empower them to exercise control over their digital content consumption and verification behaviors. This knowledge deficit directly impacts perceived behavioral control, individuals cannot effectively enact protective behaviors if they lack awareness of what to look for or how to verify authenticity. The strong statistical associations in our symmetric measures (all at 0.708 with $p < 0.001$) suggest that perceived behavioral control may be the critical limiting factor in adaptive responses to AI deception threats. Furthermore, the socioeconomic implications noted in our research align with TPB's emphasis on control factors. Higher-income individuals may have greater resources (representing one form of behavioral control) but simultaneously become targets of more sophisticated attacks, effectively neutralizing their resource advantage.

3.3.4. Perceived Behavioral Control

The TPB framework enables prediction of emergent behavioral patterns arising from the current sociocognitive configuration: negative attitudes manifested as pervasive distrust, conflicted subjective norms characterized by tension between sharing impulses and cautionary imperatives, and limited perceived behavioral control resulting from technical knowledge deficiencies. These factors collectively suggest the probable manifestation of dual behavioral adaptations in response to AI deception technologies. First, a pattern of hyper-vigilance and false positive identification may emerge, wherein the overwhelming distrust expressed by respondents potentially precipitates rejection of even authentic content, an over-correction mechanism that risks further eroding information ecosystem integrity. Simultaneously, a countervailing pattern of resigned vulnerability may develop, particularly among the substantial proportion of individuals lacking technical awareness, wherein behavioral intentions default to a form of digital fatalism characterized by passive acceptance of vulnerability stemming from perceived inability to effectively distinguish

between authentic and algorithmically manipulated content. These predicted behavioral outcomes represent maladaptive responses that could significantly impair healthy information consumption and dissemination practices.

Figure 10: Implications Of TPB



3.4. Bio-ID Detection Results

The results from the deepfake detection system show scores for two different uploaded videos and one result from a live instant captured through selfies.

Figure 11: First Detected Deepfake Video

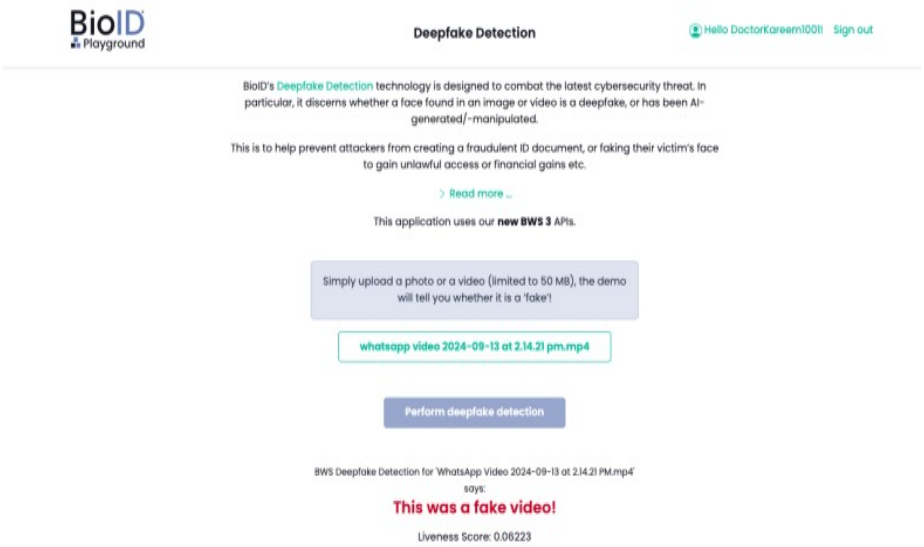
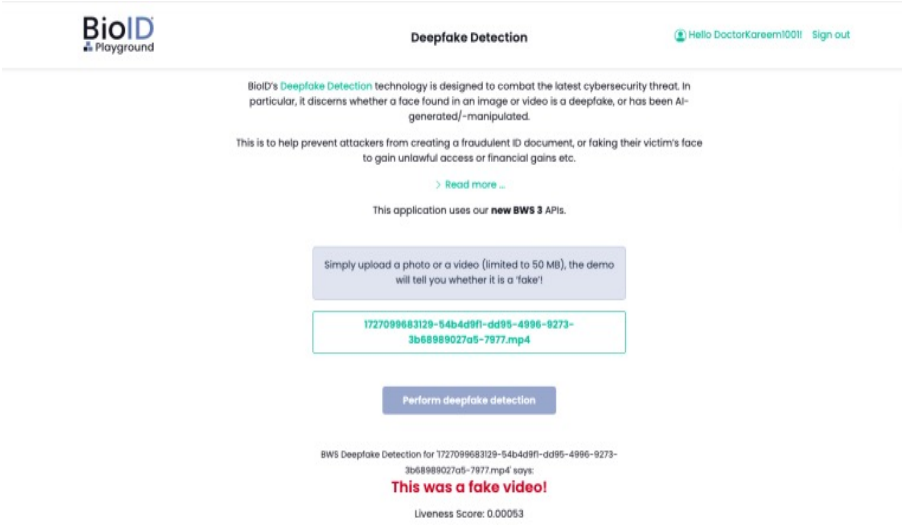


Figure 12: Second Detected Deepfake Video



For the two uploaded videos to be examined on Bio-ID, the scores of 0.00053 and 0.06223 indicate a very high likelihood that they are deep-fakes. These scores are significantly low, suggesting that the features analyzed by the online software such as facial inconsistencies, unnatural movements, and other manipulation indicators point towards the videos being artificially created or altered. In deep-fake detection, lower scores typically correlate with a higher probability of the content being fake, meaning that the system has identified substantial evidence of tampering or synthetic generation.

In contrast, the score of 0.92363 from the liveness detection of two selfies indicates a strong likelihood that this content is genuine. A score of 1 suggests that the selfies exhibit characteristics typical of fundamental, live human interactions, such as natural facial expressions and movements. This high score implies that the system recognized the selfies as being taken in real time, reinforcing their authenticity. These results demonstrate the detection system's effectiveness in distinguishing between manipulated videos and genuine live content based on the assigned liveness scores.

3.4.1. Applying Mcluhan's medium theory to the findings of Bio-ID as a Detection to Deepfake Media Contents

Additionally, The findings of this study centered on public awareness, trust, and exposure to AI-driven deception technologies among affluent European citizens aged 40 and above; can be rigorously interpreted through the lens of Marshall McLuhan's Medium Theory, particularly his central claim that "the medium is the message." This conceptual framework shifts analytical focus from the content of communication to the form and technological nature of the medium itself, offering a deeper understanding of how AI-generated content,

such as deepfakes and voice cloning, reconfigures cognitive, social, and behavioral structures. The study reveals that 76.5% of respondents had encountered deepfakes, while 98% doubted the authenticity of online videos and voice content. According to McLuhan, media technologies do not merely transmit messages but restructure the sensory balance and the epistemological frameworks through which societies interpret reality. Deepfakes, in this regard, are not just deceptive content—they are new media forms that produce an altered media environment where seeing is no longer believing. The trust once placed in audiovisual material is eroded, not by the individual messages, but by the very existence of synthetic media capabilities. This aligns with McLuhan's notion that the societal consequences of a new medium unfold independently of its content.

McLuhan's distinction between hot and cool media further illuminates the paradox revealed in the findings. Deepfakes are hot media in the sense that they provide high-resolution, immersive content that appears credible and complete. However, they simultaneously require intensive cognitive engagement—a trait of cool media, because audiences must now actively assess, verify, and doubt what they consume. The survey's documentation of widespread skepticism (e.g., 86.3% distrust in social media videos and 98% in voice messages) illustrates this cognitive dissonance, where media appear self-evident but provoke high scrutiny. This emergent hybrid space signals a new form of media environment, where technological realism paradoxically generates informational uncertainty.

A core tenet of McLuhan's theory is that media serve as extensions of human faculties, while also rendering certain older modes of perception obsolete. AI technologies such as voice cloning and face-swapping extend human creativity and communication, but at the same time, undermine the reliability of our natural senses, particularly sight and hearing, which were once foundational to trust. The 66.7% of respondents who were victims of voice-cloning scams demonstrate this breakdown: humans can no longer rely solely on auditory recognition as a valid indicator of authenticity.

In McLuhan's terms, the human ear; extended through audio media; becomes vulnerable when superseded by an artificially enhanced version of itself. McLuhan emphasized that new media technologies create new environments that alter the social order. The research highlights the formation of such an environment, wherein trust is no longer passively assumed but must be actively constructed through detection tools like BioID. These tools, as discussed in the study, function as new interpretive technologies, technological media through which users interact with other media. They do not merely detect falsehoods but mediate truth, making them part of the message itself. The rising reliance on such technologies to verify authenticity suggests the emergence of a secondary media system, layered atop traditional media, that is becoming integral to how users assess credibility. The findings support the idea of a media feedback loop, in which the presence

of synthetic media technologies breeds distrust, leading to greater reliance on AI detection tools, which in turn become new media channels that shape audience behavior. This loop is visible in the statistically significant correlation between AI awareness and reduced trust, confirmed by a Chi-square value of 25.548 ($p < 0.001$) and high association scores (e.g., P value = 0.708). McLuhan would interpret this as evidence of how media environments recursively shape both technological use and user psychology, resulting in structural shifts in how individuals experience reality.

The application of McLuhan's Medium Theory to Bio-ID deepfake detection reveals how technology restructures human perception, epistemology, and social reality. Bio-ID's detection scores: 0.00053 and 0.06223 for deepfakes versus 0.92363 for authentic content; support McLuhan's view that technology is not merely a tool but an environmental force that reshapes consciousness and frames of reference. This quantitative divide signifies not just algorithmic output but a deeper ontological shift in how authenticity is cognitively processed within a technologically mediated environment.

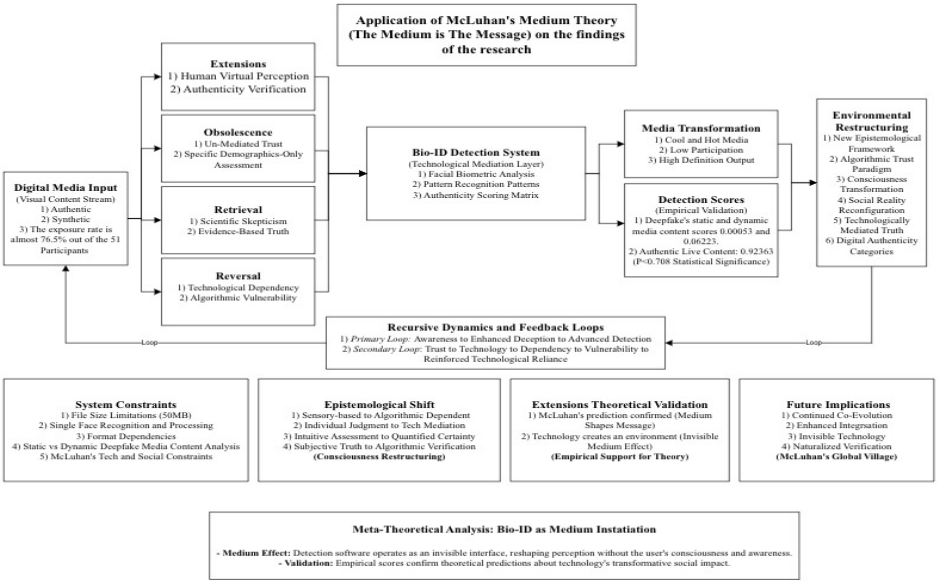
Bio-ID reflects McLuhan's media tetrad by extending human perception beyond its natural limits, rendering imperceptible manipulations visible, while simultaneously obsolescing traditional, intuition-based visual judgment. It revives concerns about media manipulation from the pre-digital era and scientific methods of verification, yet introduces new dependencies that risk undermining human autonomy in judgment; confirming McLuhan's insight that technological extensions eventually induce systemic reliance. The system also exemplifies the shift from "cool" to "hot" media: deepfake detection, which traditionally required interpretive effort, is transformed into low-participation, high-definition outputs through precise detection scores. This shift minimizes human engagement while redefining visual truth via numerical authority. Even with a statistical p-value of < 0.708, the transformation reinforces McLuhan's claim that mediums reshape cognition and social interaction.

Recursive feedback loops further validate McLuhan's theory through improved detection alters deepfake creation, prompting new algorithmic developments. Trust in the system also grows, creating secondary loops of dependence and vulnerability. These loops underscore McLuhan's view of media as ecosystems that evolve through complex, circular influences rather than linear progress. Bio-ID thus constructs a "digital authenticity environment" where unaided perception is no longer sufficient. It creates structural categories of verified/unverified content that reshape public discourse on truth. The 76.5% exposure rate to deepfakes among study participants highlights how these environments become normalized, influencing social consciousness beyond individual awareness.
Epistemologically, Bio-ID replaces sensory-based judgment with algorithmic validation, transforming consciousness through repeated mediation. It becomes not just a detection

tool, but a cognitive agent that redefines how people perceive and evaluate visual reality. However, Bio-ID's limitations; such as file size constraints, single-face detection, and static processing; illustrate how technical constraints evolve into social ones, as McLuhan predicted. These boundaries affect user interaction and perception, reinforcing the way media capabilities dictate expressive and interpretive possibilities. Applying McLuhan's theory to Bio-ID demonstrates how detection systems operate as invisible yet powerful infrastructures that shape perception, agency, and definitions of truth. As both deepfakes and detection technologies advance, the recursive, environment-shaping dynamics of McLuhan's theoretical framework suggests an ongoing, profound transformation of human consciousness and social structure driven by digital mediation.

The Bio-ID deepfake detection system demonstrated its ability to distinguish between manipulated and authentic content using liveness scoring. The two deepfake videos received scores of 0.00053 and 0.06223, signaling a strong indication of manipulation. In contrast, live selfies received a score of 0.92363, indicating high authenticity. This technology is more than a diagnostic tool; it plays a critical role in bridging the awareness gap between detection capability and public understanding. For many users; especially those with low media literacy or older age groups; the mere presence of a technical label like "AI-generated" may be insufficient. Bio-ID, with its visual scoring interface and biometric integration, translates complex assessments into intuitive feedback, offering real-time verification that empowers users. Moreover, Bio-ID can act as a trust-restoration mechanism in digital communications. By providing transparent, explainable scoring, it helps rebuild confidence in media through verification rather than censorship. This distinction is essential in democratic societies, where information control must be balanced against the right to free expression. Tools like Bio-ID could become part of the new media grammar; akin to browser padlocks for secure websites; signaling which content has passed authenticity thresholds. Integrating such markers across social platforms and news aggregators could create a new semiotic layer of trust, visible and verifiable by all users regardless of technical literacy.

Figure 13: Mcluhan's Medium Theory Applied to the findings of the present study



The juxtaposition of the Theory of Planned Behavior with McLuhan's Medium Theory generates a multidimensional interpretive framework that elucidates the profound sociocognitive implications of AI deception technologies. McLuhan's idea that "the medium is the message" acquires heightened significance in the contemporary digital landscape where the medium itself; digital video and audio content, has become inherently questionable rather than merely serving as a neutral conduit for information. This fundamental shift in medium reliability necessitates a reconsideration of how individuals form attitudes, respond to norms, and assess behavioral control within increasingly unstable epistemic environments.

The psychological processes articulated through TPB necessarily operate within a media ecosystem that, as McLuhan theorized, shapes cognitive patterns independent of specific content. Our empirical finding that 72.5% of participants could not envision positive applications for deepfake technology substantiates McLuhan's assertion that the medium's impact transcends its particular implementations or uses. This pervasive negative perception reflects not merely apprehension about specific deceptive content but anxiety regarding the medium's destabilizing effect on established modes of truth verification. The confluence of these complementary theoretical frameworks suggests that AI deception technologies constitute more than mere tools with positive or negative applications; rather, they represent a fundamental reconfiguration of the cognitive environment within which individuals form attitudes, internalize subjective norms, and evaluate their capacity for effective behavioral control. This theoretical synthesis provides an essential foundation for understanding Bio-ID's significance beyond its technical capabilities, positioning it not merely as a detection mechanism but as a transformative mediating layer that fundamentally recalibrates the

relationship between viewers and digital content.
The quantitative findings from the liveness detection system demonstrate this recalibration empirically: the pronounced contrast in liveness scores between authentic content (0.92363) and deepfakes (0.00053 and 0.06223) operationalizes what McLuhan would characterize as an emergent "grammar" of media literacy, one where algorithmic verification becomes integrated into the fundamental architecture of media consumption. The remarkably high levels of distrust observed among respondents (98% doubting online videos and voice content) validate McLuhan's theoretical prediction that emerging media technologies create environmental conditions that reshape cognitive processes and social behaviors independent of content specificity.

Within this theoretical integration, Bio-ID emerges simultaneously as both a response to and an active shaper of this new media environment, fulfilling McLuhan's conceptualization of media technologies as both extensions of human capability and formative influences on perception. The significant statistical associations documented in our study ($x^2 = 25.548$, $p < 0.001$) can thus be interpreted as empirical manifestations of McLuhan's environmental media effects, while the TPB framework provides the psychological mechanisms through which these effects translate into attitudinal, normative, and behavioral responses. As synthetic media technologies continue their evolutionary trajectory, Bio-ID and analogous detection systems will increasingly function not merely as technological countermeasures but as defining elements of the epistemological frameworks through which digital truth is established and negotiated. This research consequently positions Bio-ID at the nexus of an emerging media ecosystem where authentication becomes prerequisite to trust formation, and where McLuhan's conception that "the medium is the message" finds renewed relevance in the complex interplay between deceptive media technologies and the verification systems designed to detect them; a dynamic that fundamentally alters how TPB's components of attitudes, subjective norms, and perceived behavioral control manifest in contemporary digital environments.

4. Conclusion

The research offers insights into the awareness, experiences, and perceptions of AI deception technologies among European citizens over 40 from affluent backgrounds. Employing a combination of survey methodologies and practical demonstrations of AI tools uncovers several key findings. A notable majority of respondents (76.5%) have encountered deepfake videos, illustrating significant exposure to this technology. However, awareness of voice cloning and face-swapping remains relatively lower, pointing to a gap in public knowledge. The data also reveals a deep mistrust of digital media, with 86.3% of respondents expressing distrust in videos shared on social media and 98% doubting the authenticity of online videos. This scepticism extends to voice messages, with nearly all respondents (98%) feeling uncertain about their credibility. The real-world impact of AI deception is underscored by

the fact that 66.7% of respondents have fallen victim to voice cloning scams, and 78.4% know someone affected by identity theft linked to AI technologies. A demonstration of the BioID detection system proved effective in identifying manipulated videos, reinforcing the value of detection frameworks in mitigating such risks.

The statistical analysis conducted within this study further solidifies the relationships between key factors such as AI knowledge, trust in digital media, and personal experiences with identity theft. Chi-square tests revealed significant correlations, with a Pearson Chi-Square value of 25.548 ($df = 1$, $p < 0.001$), alongside identical association measures; Phi, Cramer's V, and Pearson's R—at 0.708 ($p < 0.001$), all pointing to the strong interconnectedness of these variables. Moreover, the reliability of the survey instruments, confirmed by a Cronbach's alpha of 0.783, lends credibility to the results. Despite the findings, the study exposes a divide in opinion regarding the ethical applications of deepfake technologies: while most respondents (72.5%) see no positive use for these technologies, a significant minority (27.5%) believe they hold beneficial potential. This divide suggests the necessity of continued discourse on the ethical and regulatory considerations surrounding AI-driven deception technologies.

McLuhan's Medium Theory provides an essential theoretical foundation for understanding the significance of Bio-ID beyond its technical capabilities. The system represents a paradigmatic shift in how media authenticity is established in the digital age. As this research demonstrates, Bio-ID is not merely a tool for detection but constitutes a new layer of mediation that fundamentally alters the relationship between viewers and content. The stark contrast in liveness scores between authentic content (0.92363) and deepfakes (0.00053 and 0.06223) quantifies what McLuhan would describe as a new "grammar" of media literacy; one where algorithmic verification becomes integrated into the basic structure of media consumption. The high levels of distrust observed among respondents (98% doubting online videos and voice content) confirm McLuhan's prediction that new media technologies create environmental conditions that reshape cognition and social behavior independent of specific content. Bio-ID thus emerges as both a response to and a shaper of this new environment, fulfilling McLuhan's concept of media serving as both technological extensions of human capability and formative influences on perception. As synthetic media technologies continue to evolve, Bio-ID and similar detection systems will not simply function as technological countermeasures but will increasingly define the epistemological frameworks through which digital truth is established. This research thus positions Bio-ID at the center of an emerging media ecosystem where authentication becomes a prerequisite for trust, and where McLuhan's idea; "the medium is the message"; finds renewed relevance in the interplay between deceptive media technologies and the systems designed to detect them. The research presents several possibilities for understanding and addressing the challenges posed by AI deception technologies, particularly deepfakes and voice cloning. The findings

highlight the potential for increasing public awareness through targeted education and training programs that equip individuals with the skills to identify and mitigate misinformation. The strong statistical correlations between awareness, trust, and identity theft experiences suggest that interventions based on the Theory of Planned Behavior (TPB) can effectively shape attitudes and behaviors toward AI-generated media. Additionally, the successful application of deepfake detection technologies, such as Bio-ID, demonstrates the feasibility of developing and integrating AI-driven solutions to enhance digital security and media credibility throught the lens of McLuhan's Medium theory. The study also provides a foundation for further research into the psychological and societal impacts of AI-generated content, offering valuable insights for policymakers, media professionals, and technology developers in regulating and improving the ethical use of AI in digital media.

However, the study acknowledges several limitations. The reliance on self-reported data introduces potential biases, as responses may be influenced by individual perceptions rather than objective assessments. The sample, consisting of European citizens over 40 from affluent backgrounds, limits the generalizability of the findings to diverse demographic groups. Furthermore, while statistical analyses establish significant correlations between awareness, trust, and exposure to AI deception technologies, they do not establish causal relationships. The effectiveness of detection tools also depends on their adaptability to evolving deepfake technologies, necessitating continuous updates and improvements as The BioID Deepfake Detection software, while helpful in identifying manipulated media, presents several limitations that could hinder its academic and professional applications. The 50 MB file size restriction prevents the analysis of longer or high-resolution videos, while its inability to process multiple faces limits its functionality in group or dynamic scenes. The software also struggles with videos where facial features are obscured by poor lighting, low resolution, or occlusions, reducing accuracy. Furthermore, its dependence on specific file formats and lack of real-time detection capabilities make it less versatile for live applications. Advanced deepfake techniques that seamlessly blend features or introduce subtle manipulations may evade detection, especially if the algorithm is not regularly updated. Lastly, its reliance on static evaluations rather than dynamic frame-by-frame analysis and the potential need for file preprocessing create additional user challenges. Addressing these limitations is crucial for improving the software's robustness and adaptability.

While the integration of TPB and McLuhan's Medium Theory offers substantial interpretive value, several theoretical limitations warrant acknowledgment in their application to AI deception technologies. The TPB framework, despite its explanatory power regarding attitude-behavior relationships, inadequately addresses the non-rational, affective responses frequently triggered by deepfake encounters, emotional reactions that may circumvent the rational deliberative processes presumed by the theory; while simultaneously undertheorizing the rapidly evolving nature of technological literacy that characterizes

perceived behavioral control in this domain. Similarly, McLuhan's Medium Theory, though valuable for conceptualizing media environments, provides insufficient analytical granularity for examining individual differences in media reception and interpretation, potentially overlooking significant variations in how diverse demographic groups process and respond to potentially deceptive content. Furthermore, future research studies should account the rapidly rising relationship between technological development and user adaptation and experience, wherein detection technologies and deception technologies coevolve in response to one another, creating dynamic equilibria that resist static theoretical modeling.

To address the growing ethical concerns surrounding deepfake and AI-generated content, the following principles are proposed for media professionals, technology developers, and policymakers:

1) All AI-enhanced or synthetic content should be clearly labeled to distinguish it from authentic media. Transparency in media production is essential to reduce misinformation and protect public trust (Transparency and Labeling).
2) Educational campaigns should be designed to cultivate critical thinking skills, enabling audiences to assess the credibility of digital content and reduce susceptibility to manipulation (Digital Literacy and Public Awareness).
3) Legal and ethical frameworks must enforce accountability for the misuse of AI tools in media production, especially in cases of identity theft, political manipulation, or fraud (Accountability and Regulation).
4) Developers of AI technologies must integrate detection and authentication systems into their platforms to minimize misuse. These tools should be regularly updated to match the evolving sophistication of deepfake technologies (Technological Safeguards).
5) Policymakers, technologists, educators, and communication scholars should work together to develop balanced, enforceable guidelines that uphold ethical standards while encouraging responsible innovation (Interdisciplinary Collaboration).
6) Institutions such as schools, libraries, and local organizations should play a proactive role in promoting ethical media use and digital safety through public programs and community engagement (Community and Institutional Responsibility).

Building on the findings of this study, the following can be suggest as future recommendations:

1) Future studies should target populations most at risk of digital deception; such as

the elderly, youth, and individuals with limited technological literacy; to understand their specific vulnerabilities and educational needs (focusing on vulnerable demographics).

2) Research should explore the effectiveness of tailored digital literacy programs designed to enhance public resilience against deepfakes, voice cloning scams, and AI-based fraud (to develop targeted educational interventions).

3) There is a need for larger-scale empirical evaluations of detection tools like BioID, including their performance across diverse media formats, quality levels, and real-time environments (evaluate AI detection tools at scale).

4) Ongoing studies should collect longitudinal data on public exposure to AI-generated media and its psychological, social, and political consequences, including trust erosion and behavioral changes (to monitor AI manipulation trends).

5) Future work should analyze the effectiveness of existing and emerging regulatory frameworks in managing AI-driven deception, with attention to enforcement mechanisms and international coordination (assess policy and regulatory impact).

6) Researchers should investigate the role of community forums, mentorship programs, and grassroots initiatives in fostering informed dialogue about ethical media practices and the implications of AI technologies (promote civic engagement and public discourse).

References

Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology & health*, 26(9), 1113-1127.

Al-Khazraji, S. H., Saleh, H. H., KHALID, A. I., & MISHKHAL, I. A. (2023). *Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications*. The Eurasia Proceedings of Science Technology Engineering and Mathematics, 23, 429-441.

Amerini, I., Barni, M., Battiato, S., Bestagini, P., Boato, G., Bonaventura, T. S., ... & Vitulano, D. (2024). *Deepfake Media Forensics: State of the Art and Challenges Ahea*d. arXiv preprint, 2408.00388.

Carlson, M. (2020). Fake news as an informational moral panic: The symbolic deviancy of social media during the 2016 US presidential election. *Information, Communication & Society*, 23(3), 374-388. https://doi.org/10.1080/1369118X.2018.1505934

Chapagain, D., Kshetri, N., & Aryal, B. (2024). *Deepfake Disasters: A Comprehensive Review of Technology, Ethical Concerns, Countermeasures, and Societal Implications.* In 2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), 1-9. (IEEE.)

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press. https://doi.org/10.1515/9781503609754

Esezoobo, S. O., & Braimoh, J. J. (2023). Integrating Legal, Ethical, and Technological Strategies to Mitigate AI Deepfake Risks through Strategic Communication. *International Journal of Scientific Research and Management* (IJSRM), 11(08), 914-928.

Fabuyi, J. A., Olaniyi, O. O., Olateju, O. O., Aideyan, N. T., Selesi-Aina, O., & Olaniyi, F. G. (2024). Deepfake Regulations and Their Impact on Content Creation in the Entertainment Industry. *Archives of Current Research International,* 24(12), 52-74.

Fallis, D. (2021). *The epistemic threat of deepfakes. Philosophy & Technology*, 34(4), 623-643. https://doi.org/10.1007/s13347-020-00419-2

Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. (Sage publications limited.)

Future, M. R. (2023). *AI-Generated Media Market Research Report*.

George, A. S. (2023). *Deepfakes: the evolution of hyper realistic media manipulation.* Partners Universal Innovative Research Publication, 1(2), 58-74.

Gillespie, T. (2022). *Governance of and by platforms*. In J. Burgess, A. Marwick, & T. Poell (Eds.), The SAGE handbook of social media (pp. 254-278). SAGE Publications. https://doi.org/10.4135/9781473984066.n15

Karnouskos, S. (2020). *Artificial intelligence in digital media: The era of deepfakes*. IEEE Transactions on Technology and Society, 1(3), 138-147.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). *Deepfakes: Trick or treat?* Business Horizons, 63(2), 135-146. https://doi.org/10.1016/j.bushor.2019.11.006

Livingstone, S., & Lunt, P. (2023). Trust calibration in older adults: Media literacy interventions for the digital age. *European Journal of Communication*, 38(2), 145-162. https://doi.org/10.1177/02673231221147315

https://doguakdenizmy.sharepoint.com/:u:/r/personal/21602622_emu_edu_tr/_layouts/15/Doc.aspx?sourcedoc=%7B8AFEE0C2-4F72-4ED2-B9B52A357AA4F4A6%7D&file=1st%20Mcluhan%27s%20Flowchart%20(AIDeepfake%20Media).

vsdx&action=default&mobileredirect=true

Malik, K. M. & Baig, R. (2023). *Deepfake voice detection: Techniques, challenges, and future directions.* IEEE Access, 11, 12504-12524. https://doi.org/10.1109/ACCESS.2023.3241711

Mirsky, Y., & Lee, W. (2021). *The creation and detection of deepfakes: A survey.* ACM computing surveys (CSUR), 54(1), 1-41.

Nasar, B. F., Sajini, T., & Lason, E. R. . (2020). *Deepfake detection in media files-audios, images and videos.* In 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 74-79. (IEEE.)

Ng, Y. L. (2024). *A longitudinal model of continued acceptance of conversational artificial intelligence.* Information Technology & People.

Nowroozi, E., Seyedshoari, S., Mohammadi, M., & Jolfaei, A. (2022). *Impact of media forensics and deepfake in society.* In Breakthroughs in Digital Biometrics and Forensics, Cham: Springer International Publishing, 387-410.

Oza, P., Patel, N., & Patel, A. (2024). *Deepfake Technology: Overview and Emerging Trends in Social Media.*

Papacharissi, Z. (2021). *Affective publics: Sentiment, technology, and politics.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199999736.001.0001

Play.ht. (2024). *VoiceCloning.* https://play.ht/studio/files/59d00317-79f6-4092-801b-49d3429ec1da?voice=s3%3A%2F%2Fvoice-cloning-zero-shot%2F14d1b898-f68b-4868-96c2-9192755b5095%2Foriginal%2Fmanifest.json

Rask.AI. (2024). *VideoDubbing.* https://app.rask.ai/project/4d33a4f8-7e8b-4b54-99eb-c23a11b2b182

Shah, D. V., Hiaeshutter-Rice, D., Lukito, J., & Wells, C. (2023). *Deepfakes and democratic participation: How synthetic media awareness affects political engagement across demographic groups.* Political Communication, 40(1), 98-119. https://doi.org/10.1080/10584609.2022.2144679

Shakil, M., & Mekuria, F. (2024). *Balancing the Risks and Rewards of Deepfake and Synthetic Media Technology: A Regulatory Framework for Emerging Economies.* In 2024 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), 114-119. (IEEE.)

Sharma, M., & Kaur, M. . (2022). *A review of Deepfake technology: an emerging AI threat.* Soft Computing for Security Applications, Proceedings of ICSCS 2021, 605-619.

Software, B. (2024). *Deepfake Technologies Detection.* https://www.bioid.com/playground/

Survey, A. (2024). Awareness and Exposure. https://docs.google.com/forms/d/e/1FAIpQLSdGQgrBqq97XPFVr0r6-NLpEK_d6XMXqT9eWZheS1qzFSQ6pg/viewform?usp=sf_link

SwapFace. (2024). *SwapFace technologies.* https://www.swapface.org

SyncLabs. (2024). *Lip Synchronization.* https://app.synclabs.so/share/lip-sync/514ec9c3-f0ab-463b-8f9c-b3a7372bf731

Temir, E. (2020). *Deepfake: new era in the age of disinformation & end of reliable journalism.* Selçuk İletişim, 13(2), 1009-1024.

TheftExperienceSurvey. (2024). *Identity and Theft Experience.* https://docs.google.com/forms/d/e/1FAIpQLSeHfV4eCQaTyi-f9S6X0UFzL5tEjCPhX4uRrmhykBKKeLgwkQ/viewform?usp=sf_link

TrustAssessmentSurvey. (2024). *Trust Assessment Survey.* https://docs.google.com/forms/d/e/1FAIpQLSeQGrXEGKqR1G8KtzfBHBScB7fXfoSaqK4zax0G8DMMjHeKqw/viewform?usp=sf_link

Thies, J., Zollhöfer, M., & Nießner, M. (2020). *Deferred neural rendering: Image synthesis using neural textures.* ACM Transactions on Graphics, 38(4), 1-12. https://doi.org/10.1145/3306346.3323035

Tuysuz, M. K., & Kılıç, A. (2023). *Analyzing the Legal and Ethical Considerations of Deepfake Technology.* Interdisciplinary Studies in Society, Law, and Politics, 2(2), 4-10.

Vaccari, C., & Chadwick, A. (2020). *Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news.* Social media+ society, 6(1), 2056305120903408.

Vizoso, Á., Vaz-Álvarez, M., & López-García, X. (2021). *Fighting deepfakes: Media and internet giants' converging and diverging strategies against hi-tech misinformation.* Media and Communication, 9(1), 291-300.

Westerlund, M. (2019). *The emergence of deepfake technology: A review.* Technology Innovation Management Review, 9(11), 40-53. https://doi.org/10.22215/timreview/1282

Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., & Choi, Y. (2019). *Defending against neural fake news.* Advances in neural information processing systems, 32.