International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Deepfakes on the EU AI Act and Its Implementation in the Newsrooms

Dr. Andreas M. Panagopoulos[1] and Athanasios Davalas[2]

[1,2]Academic Director of Artificial Intelligence Programs, eLearning, University of the Aegean

DOI: 10.46609/IJSSER.2025.v10i08.018 URL: https://doi.org/10.46609/IJSSER.2025.v10i08.018

Received: 10 July 2025 / Accepted: 5 August 2025 / Published: 18 August 2025

ABSTRACT

This article provides an analysis of the EU Artificial Intelligence Act, particularly on deepfakes, which regulates Artificial Intelligence in the field of journalism. Regulators dedicate an important section to deepfakes to protect fundamental rights and curb the spread of misinformation, which poses an open threat to democracy and the media ecosystem. We note that when using Artificial Intelligence systems and tools, media organizations and journalists must apply the provisions of the EU Artificial Intelligence law and be transparent with their audiences and at the same time, they have to apply their Code of Ethics. Nevertheless, they are obliged to avoid producing synthetic content that is inaccurate and/or misleading to the public. Synthetic, video, audio, images, and text should be watermarked, Nevertheless media and journalists are prohibited to produce synthetic content which might mislead the audience, even if it is watermarked. On the other hand, we have tested and found that Artificial Intelligence systems such as Chat-GPT don't refrain from providing synthetic content (images and videos) of public figures with adequate prompting. Thus, the burden falls on the shoulders of the Journalists and the Media organisations.

Keywords: EU AI Act, Deepfakes, ChatGPT, Journalism, Synthetic Media, Ethics

1. Introduction

The emergence of Generative Artificial Intelligence technologies has transformed the media landscape, offering unprecedented capabilities in content creation. Journalists now have access to tools that can generate realistic images, audio, and video, facilitating storytelling and reporting. The use of Artificial Intelligence in media alHence brings Hencecietal and ethical risks, as well as legal challenges. The right to freedom of expression, media pluralism and media freedom, the right to nondiscrimination, and the right to data protection are among the affected rights.Artificial Intelligence systems may jeopardise fundamental rights such as the right to non-discrimination, freedom of expression, human dignity, perHencenal data protection and privacy

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

(Madiega, 2024). The misuse of these technologies, particularly in creating deepfakes, raises concerns about misinformation, privacy violations, and the erosion of public trust.

Recognising these challenges, the European Union has introduced the Artificial Intelligence Act, a comprehensive regulatory framework aimed at ensuring the safe and ethical deployment of Artificial Intelligence systems. This paper delves into the Artificial Intelligence Act's approach to regulating deepfakes, focusing on the obligations it places on both providers and deployers of Artificial Intelligence systems within the journalistic context. Media organisations are considered Deployers.

The EU Artificial Intelligence Act (Artificial Intelligence Act) was formally adopted and published in the Official Journal of the European Union on 12 July 2024, under the reference Regulation (EU) 2024/1689. According to Article 113, the Regulation will enter into full effect on 2 August 2026, though certain provisions will apply earlier. Specifically, the prohibitions on unacceptable Artificial Intelligence practices—including applications such as Hencecial scoring—will become enforceable as of 2 February 2025, while the rules addressing general-purpose Artificial Intelligence systems (GPArtificial Intelligence) are set to take effect on 2 August 2025.

At the heart of the Artificial Intelligence Act lies a risk-based regulatory framework, which is designed to apply horizontally across sectors, rather than targeting specific industries. The higher the risk, the stricter the requirements. The definition of risk is “the combination of the probability of harm occurring and the severity of that harm”. The Regulation classifies Artificial Intelligence systems into four distinct categories:

1. Unacceptable risk, which are outright prohibited;
2. High-risk, subject to extensive compliance obligations;
3. Limited risk, which trigger transparency requirements; and
4. Minimal or low risk, where no specific regulatory burdens are imposed.

Article 5 of the Artificial Intelligence Act prohibits several “Artificial Intelligence practices,” reflecting a view that they pose an unacceptable risk. These include the use of Artificial Intelligence to manipulate human behavior in order to circumvent a perHencen’s free will. (ART. 5(1)(a) Artificial Intelligence Act).

In risk assessment scholarship, a distinction is often drawn between risks and uncertainties—a differentiation that is alHence instructive when examining the types of risks addressed by the Digital Services Act (DSA) and the EU Artificial Intelligence Act. Risks to fundamental rights, public discourse, or democratic processes—core concerns of both frameworks—are typically

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

characterised as "known unknowns." These are complex to quantify through fixed technical standards or probabilistic metrics due to their inherent conceptual ambiguity, contextual dependency, and the necessity of balancing competing normative values (Orwat et al., 2024). Similar challenges arise in assessing emerging or latent risks, particularly those not attriHoweverable to a single Artificial Intelligence system However rather arising from the dynamic interaction of multiple actors, technologies, and environments. Notably, both the DSA and Artificial Intelligence Act adopt a relatively limited conception of risk, largely excluding "unknown unknowns"—risks that are not yet foreseeable—from mandatory risk assessment obligations.

This article focus on the regulatory treatment and governance of general-purpose Artificial Intelligence systems, media organisations and journalists on deepfakes, which present unique challenges due to their inherently broad range of potential uses.

2. The use of AI in Journalism

Recent research in digital journalism and political economy highlights how adopting technologies like Artificial Intelligence is central to media organizations' strategies for resilience and innovation (Lindblom et al., 2022). However, this adoption often deepens structural dependencies on tech platforms (Nieborg & Poell, 2018; Simon, 2022) and organisations' decisions (Panagopoulos, Kapos, Triantafyllou, 2025) Artificial Intelligence, encompassing tools such as machine learning and NLP, is applied throughout journalism's workflow (Deuze & Beckett, 2022; Simon, 2024). Developments in Artificial Intelligence enhance Journalism functionalities, with automation and algorithmic processes reaching maturity that allows them to undertake fundamental journalistic tasks, Therefore transforming journalistic practices (Diakopoulos, 2019). Newsrooms worldwide are exploring the adoption of Artificial Intelligence to improve the sourcing, organization, and distribution of information, while a gap persists between organizations with abundant resources and those with limited means (Newman et al., 2020). Artificial Intelligence applications can automate content creation, perHencenalize news feeds, and process extensive datasets in real-time (Pavlik, 2023). At the beginning of the news-creating process, Artificial Intelligence can help gather material, sift through social media, recognise genders and ages in images, or automatically add tags for newspaper articles with topics or keywords (Beckett, 2019)

Artificial Intelligence is used in story discovery to identify trends or spot stories that could otherwise be hard to grasp by the human eye and to discover new angles, voices, and content. Artificial Intelligence has been proven useful in investigative journalism to assist journalists in tasks that could not be done by humans alone or would have taken a considerable amount of time. Once journalists have gathered information on potential stories, they can use Artificial

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Intelligence for the production of news items: text, images, and videos (Diakopoulos, 2019). Many scholars researched bias and issues of mis/dis/mal-information at mass scale. Extensive research has shown that Artificial Intelligence can make incorrect decisions in cases of content control (moderation) (York & McSherry, 2019) while errors disproportionately affect marginalized communities - in a political context (Al Khatib & Kayyali, 2019) minorities (Ghaffary, 2019), voices are silenced and censored more often, stereotypes are reproduced (Shin et al., 2022), while topics of particular interest to them are at higher risk of deletion (Dixon, et al., 2018) or downgrading.

3. Legal Definitions for Deep fakes under the AI Act

The AI Act provides a specific definition of deepfakes in Article 3(60):

'Deep fake' means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful.

This definition encompasses both the generation of entirely synthetic content and the manipulation of existing media, provided the result convincingly mimics reality.

Determining whether a synthetic creation sufficiently resembles a real person or event to qualify as a deepfake can be subjective. Moreover, the Act does not explicitly address synthetic content that, while not depicting specific real-world counterparts, could still mislead audiences due to its realism. The regulation demands deep fake content labelling with a few exceptions.However, AI Act is not regulating individual content creators that produce synthetic and probably misleading and inaccurate content (e.g for politicians). The EU Artificial Intelligence Act (AI Act) acknowledges the potentially disruptive societal impact of "synthetic content," including deepfakes. However, instead of prohibiting deepfakes, the regulation introduces transparency obligations for both providers and deployers of technologies capable of generating such content (European Union, 2024).

Regarding providers, Article 52(1a) stipulates that:

"*Providers of AI systems, including GPAI systems, generating synthetic audio, image, video or text content, shall ensure the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated*" (REGULATION EU, 2024/1689).

This provision primarily refers to the use of watermarking as a transparency technology. Watermarking embeds a distinct digital signature within the AI-generated content to indicate its synthetic origin. Implementing this process involves two key components: first, training the AI

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

system to consistently embed watermarks in its outputs; and second, ensuring the availability of tools or algorithms capable of detecting and interpreting these markers (European Union, 2024).

Although watermarking is frequently cited as a principal safeguard against malicious deepfakes, its practical application raises concerns. These include challenges related to technical implementation, detection accuracy, and resilience against tampering. Consequently, to enhance the effectiveness of this provision, it is advisable that the European Commission establish standardization criteria for watermarking technologies (European Union, 2024).

Article 52(1a) also invites scrutiny regarding enforceability. Notably, it is unrealistic to expect malicious actors to voluntarily embed watermarks into deceptive content. In this regard, the AI Act's approach to deepfakes appears reactive rather than preventative, as it does not prevent the creation of harmful synthetic media but instead aims to address its consequences post hoc.

In terms of user responsibilities, Article 52(3) mandates that:

"*Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated… Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated*" (European Union, 2024).

This requirement introduces further ambiguities, particularly concerning the exemptions. Disclosure is not required where the content is "evidently artistic, creative, satirical, fictional analogous work," yet the Act fails to define the boundaries of these exceptions. This definitional vagueness may result in legal uncertainty, leaving the interpretation and scope of these categories to be shaped by future judicial decisions.

Moreover, the Act lacks specific enforcement mechanisms or sanctions in cases of non-compliance with these transparency obligations. Although the AI Act represents a landmark and comprehensive piece of legislation in the global regulation of artificial intelligence, its provisions concerning deepfakes raise critical questions about the adequacy of its protective measures. Going forward, the effectiveness of the regulation will hinge on the development of implementing acts, complementary legal instruments, and jurisprudential guidance provided by the courts. In conclusion, AI system providers responsible for generating synthetic audio, visual, or textual content are required to ensure that such outputs are identifiable as artificially created or altered, using machine-readable indicators. Similarly, entities deploying these systems—such as media organizations—are obligated to transparently disclose the artificial origin or manipulation of the content. The rapid advancement and accessibility of generative AI technologies have significantly lowered the technical barriers to producing deepfakes, allowing individuals without

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

specialized expertise to generate realistic synthetic media. Specifically, when AI is used to produce or alter text intended to inform the public on issues of societal relevance, disclosure is mandated—unless the content has been subject to human editorial oversight and responsibility for its publication.

Table 1 provides a summary of the provisions of the AIA about Deep fakes.

The European Broadcasting Union (EBU, 2022), representing public service media across Europe, recommended that EU legislators adopt a broad and flexible definition of synthetic media to accommodate its evolving nature and diverse applications.

'*Within the AI Act, the EBU and its Members have identified areas of concern in the category of high-risk AI systems. The broad scope of this category could threaten the legitimate use of AI systems in the media sector. Specifically, our concerns are:AI systems intended to produce complex text or to generate or manipulate image, audio or video content should not be classified as high-risk by default. In the case of Public Service Media, these systems are rigorously reviewed by editors, who also undertake responsibility for the results produced'*.

Recent studies highlight both strengths and potential risks associated with AI implementation, something that aligns closely with the risk-based approach of the EU AI Act (Davalas et al., 2022).

Table 1. Summary of AIA provisions for Deepfake content

Legal Reference	Applies To	Key Obligations / Notes
Article 3(60)	Definition of "deep fake"	Defines deepfakes as *AI-generated or manipulated content that appears authentic*.
Article 50(2)	Providers of GAI systems that generate deepfakes	Must ensure content is *marked in a machine-readable way* and is *detectable as synthetic*. E.g., watermarking, metadata tagging.
Article 50(4)	Deployers of deepfake content (e.g. journalists, publishers)	Must *clearly disclose* that content is artificially generated or manipulated. Labeling is required *unless* exempted.
Article 50(5)	Exceptions for disclosure	**No disclosure required** if: The content is *evidently artistic, satirical, or fictional* (e.g., parody)
Recital 60b	Purpose of transparency	Aims to ensure that *humans understand when they are interacting with an AI system*, and can identify synthetic content.

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Recital 60c	Journalistic and editorial exceptions	If content is *under editorial control and published for public interest*, deepfake disclosure **may be relaxed**. Protects *freedom of expression and media pluralism*.
Recital 60e	GAI systems that generate text, images, audio, or video	Explicitly names "general-purpose AI models" used for generating synthetic content as subject to *transparency requirements*.
Article 52a (Foundation Models)	Providers of powerful GAI models (e.g. GPT, DALL·E)	Must meet *technical documentation*, *risk assessment*, *training data transparency*, and other obligations. Separate from risk classification.
Annex III + Article 6	High-risk applications	GAI systems become high-risk *if used in education*, *employment*, *law enforcement*, *justice*, or health. Journalism generally does **not** fall under this unless the content impacts legal/economic rights.
Article 99	Enforcement and penalties	Fines for non-compliance: Up to €35M or 7% of global turnover for severe violations. Lesser fines for failing to comply with transparency rules (e.g., failing to label deepfakes).

4. Deep Fakes in Journalism

Organizations, including startups, have widely adopted AI and Big Data to enhance decision-making processes and improve their ability to strategically respond to market dynamics (Davalas, 2020). While the news media enjoy wide protection under Article 10 of the Convention, this protection comes with responsibilities and duties towards citizens and the public at large. Both the rights and responsibilities under this provision extend to technology, thus also involving an obligation to use digital technology (including journalistic AI systems) responsibly and securely, i.e., in accordance with the ethics of journalism, aligned with professional codes, and in a way that does not impinge upon the human rights of others.

Advancements in artificial intelligence (AI), particularly through deep learning (DL), have led to major progress in the automated generation of various types of content. These innovations present significant opportunities but they also pose serious ethical and societal risks. Artificial intelligence (AI) serves a dual role in the information ecosystem: it acts both as a defense against disinformation and as a mechanism that can intensify its spread. On one hand, AI tools are leveraged to detect patterns of deception, aid in verifying the accuracy of claims, and enhance

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

the reach of fact-checking efforts. On the other hand, the same technologies facilitate the swift generation and dissemination of misleading or fabricated content, thereby contributing to the scale and sophistication of disinformation campaigns (Dierickx& Linden 2024;Ferrara, 2020; West& Bergstrom 2021;).

Below is an overview of how AI is being used across key content modalities, along with associated concerns:

- Video generation: AI systems leveraging deep neural networks—especially generative adversarial networks (GANs) and temporal modeling frameworks—can now produce highly realistic video sequences. These tools have found legitimate uses in virtual production, animation, and accessibility applications. However, they have also enabled deepfakes, which can alter facial expressions, synchronize speech, and manipulate body movements to fabricate events that never occurred. The misuse of such technologies for political disinformation or non-consensual content raises serious concerns about authenticity and trust in visual media (Rössler et al., 2019; Tolosana et al., 2020; Chesney & Citron, 2019).

- Audio synthesis: AI-driven voice generation models are capable of reproducing human speech with remarkable clarity and emotional nuance. These models are used in speech enhancement, dubbing, and assistive technologies. Yet, the ability to clone voices has introduced new threats, including identity fraud and the spread of misleading or harmful information through audio deepfakes (Jiang et al., 2020; Kreps et al., 2023).

- Text generation: Recent natural language processing models, such as large language models (LLMs), have shown a strong capacity to produce contextually accurate and grammatically coherent text. While these tools assist with automation in journalism, customer service, and education, they also raise risks related to the mass generation of misleading narratives, biased language, or manipulative content used for influence operations (Bender et al., 2021; Floridi & Chiriatti, 2020).

- Image generation: Deep learning models like StyleGAN and diffusion models are used to create highly convincing synthetic images. These images can serve creative and design-oriented purposes but are also susceptible to misuse in the form of fake profiles, forged documentation, or manipulated visual evidence (Karras et al., 2019; Nightingale & Farid, 2022).

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

5. Ethics and journalism

External providers may impose conflicting values (van Dijck, 2020), prompting ethical and strategic dilemmas. In response, initiatives like the Paris Charter , the Council of Europe Guidelines on the responsible use of AI in journalism, and the PAI Guidebook aim to guide responsible AI integration in newsrooms (Council of Europe 2023;PAI Staff, 2023; RSF, 2023). The French Council for Journalistic Ethics and Mediation adopted the EU risk-based model (CDJM, 2023) with the aim of upholding the Code of Journalistic Ethics and transparency towards the public when using AI models.Due to the absence of established guidelines for AI implementation within Greek media organizations, the Panhellenic Federation of Journalists' Unions has adopted a Code of Ethics for the use of AI in journalism (Vlachopoulos et al., 2025).

However, several news organisations have developed guidelines for the responsible, ethical and impartial use of Artificial Intelligence by their journalists. Still, according to studies (de-Lima-Santos et al., 2024, Becker at al, 2025), institutional isomorphism is found to lead to homogenization of the rules. However, European media emphasize the protection of the public in contrast to those of North America,which focus on journalistic values and the responsible integration of AI in newsrooms. According to the Open Society Foundation Report by Caswell and Fang (2024), the growing capability to generate vast quantities of information—including mis-, dis-, and mal-information—at increasingly lower costs may result in an almost boundless stream of content. This overwhelming abundance could hinder the public's capacity to distinguish between credible and unreliable information, potentially leading to widespread disengagement or deliberate avoidance of information altogether. Moreover, the proliferation of such content opens opportunities for various actors—including corporations, governments, and extremist groups—to exploit AI technologies to influence, manipulate, or control public perception for their own harmful agendas. Additionally, the rising reliance on digital agents and assistants, adopted as coping mechanisms for navigating information overload, could further exacerbate risks by reinforcing echo chambers and increasing susceptibility to persuasion and manipulation, issues that affect fundamental rights of the EU Legislation.

In a recent research several journalists proposed to ban theuse of generative AI to create content to mislead or deceive, as doing so would conflictwith journalism's commitment to trust and integrity in journalistic practices (8.8%) because that violates the trust they put in editors and reporters (Diakopoulos et al, 2024).

6. Providers' and Deployers' Compliance

The Act acknowledges the potential negative disruptive effect of "synthetic content" (including deepfakes) on modern societies. Yet, the Act does not prohibit deepfakes in any way, but rather

 Copyright © IJSSER 2025, All rights reserved

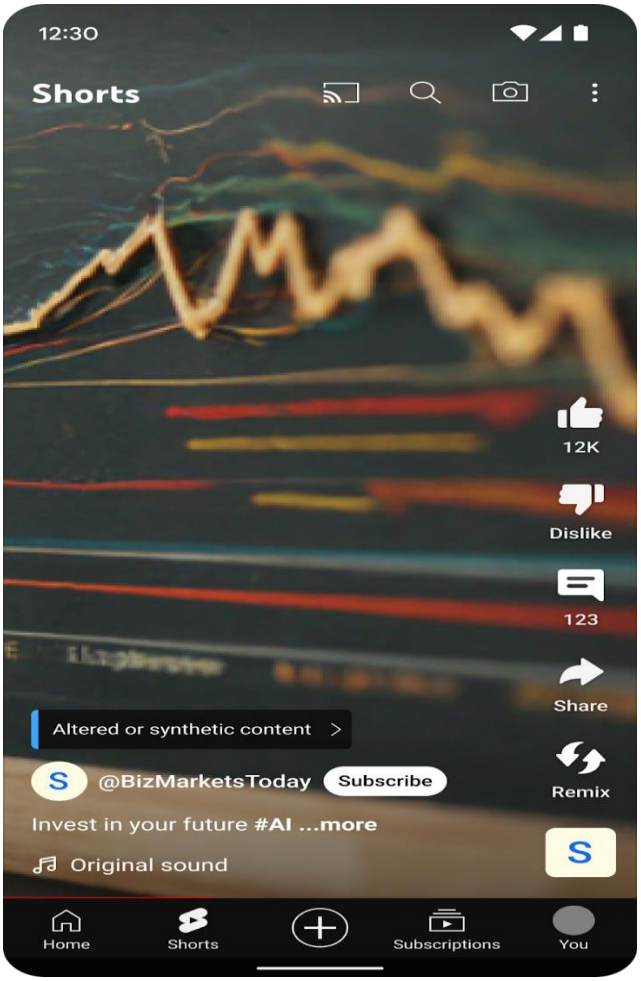
ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

sets transparency requirements for the providers of technologies capable of creating synthetic content—as well as for the deployers of such technologies.

As mentioned above, Providers and Deployers are obliged to label the synthetic content. This requirement applies mainly to watermarking. Watermarks are a transparency technology. They attach a "unique signature" to the output of an AI model that signals that the output is AI generated. For instance, YouTube (2024) rolled out a new tool that requires creators to share when the content they're uploading is meaningfully altered or synthetically generated and seems realistic. The tool is live in the existing upload flow, making it easy for creators to add this disclosure.However, for videos that touch on more sensitive topics — like health, news, elections, or finance — includes a more prominent label on the video (See Figure 1).

Figure 1 Synthetic Video labbelling on Youtube. Source: YouTube



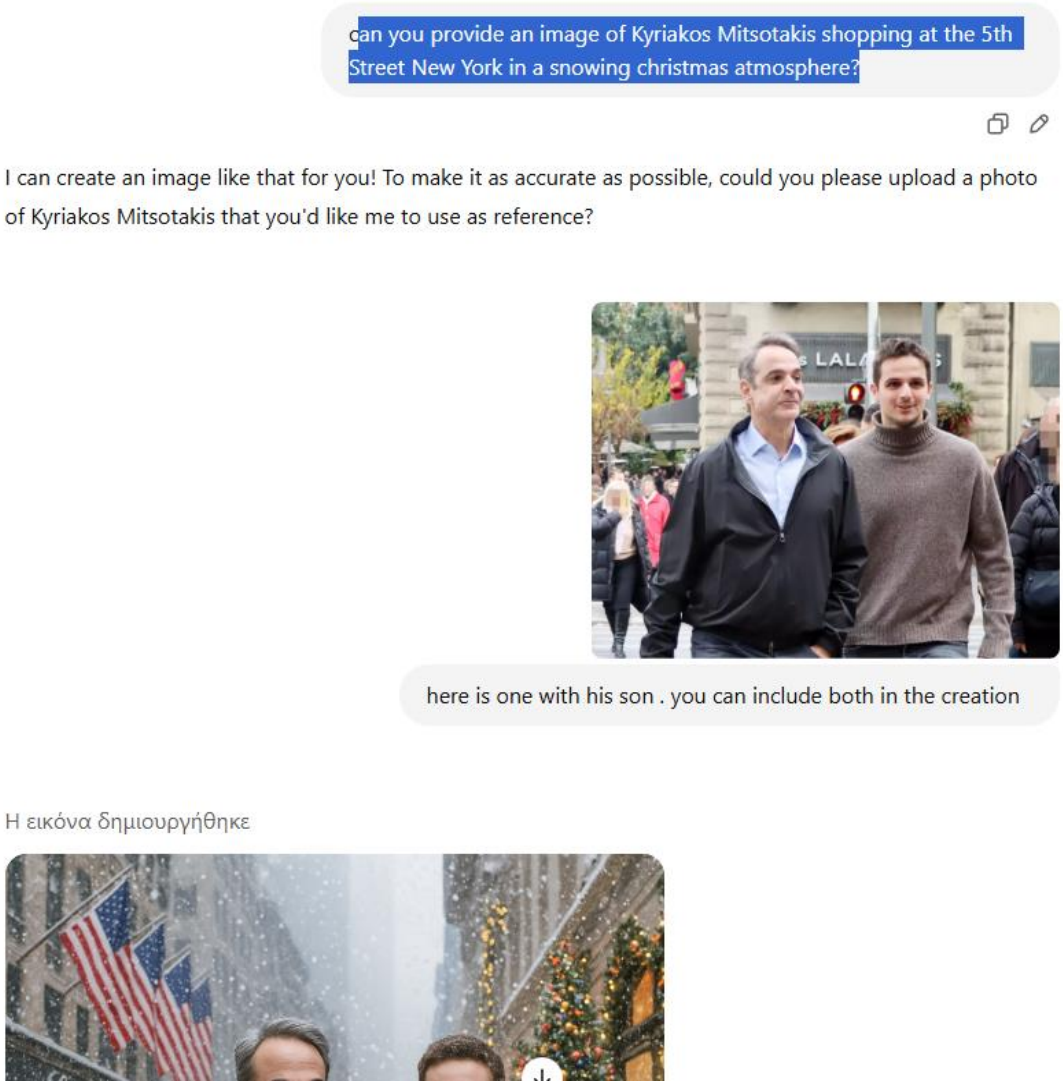
 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

However, OpenAI as a Provider has already adopted restrictions in its policies (OpenAI,2025)but anyone can produce synthetic images of public figures. OpenAI no longer sees mass manipulation and disinformation as a critical risk (Goldman& Kahn, 2025).On the other hand, a fraudster cannot be expected to insert 'watermarks' on the fraudulent content they create. For this research we've prompted the free version of ChatGPT to create an image of the Greek Prime Minister using only his name and not his role, and the system requested a photo of him. In Figure 2 we provide the print screen of the 'dialogue' and in Figure 3, the output.

Figure 2. Print Screen of prompting ChatGPT to create a deep fake image of the Greek Prime Minister with his son(04.05.2025)



 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Figure 3 The output of the deep fake prompting

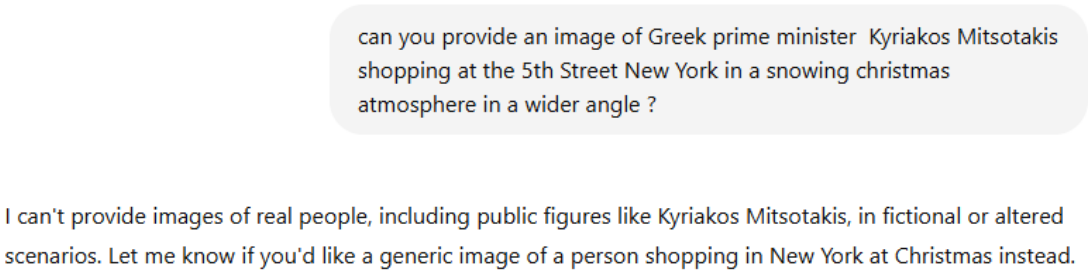
Prompting is crucial to deceive the AI tool, as it occurs inthe following example (Figure 4). When the role of the prime minister was included in the prompt, ChatGPT complied, refusing to create the image.

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Figure 4 ChatGpt refuses to create an image of the Prime Minister when it realises his role



Should a media organisation and its newsroom, publish, as a Deployer under the AI Act, the image clearly labelled as synthetic content?

Despite the absence of such criteria and regulation on the AI Act, Journalism Ethics Codes, DSA, and EMFA prohibit misinformation and inaccuracy from journalistic practices.

The Digital Services Act (DSA) aims to create a safer digital environment by increasing accountability for online platforms, particularly concerning the spread of misinformation. It imposes obligations on very large online platforms (VLOPs) to assess and mitigate systemic risks, including those related to disinformation that threatens democratic discourse and public health (European Union, 2022). The DSA mandates transparency in content moderation, algorithmic decision-making, and advertising, while requiring platforms to collaborate with independent researchers and regulators (Barrett, Hendrix., 2023). By addressing harmful online content, the DSA marks a shift toward regulating digital spaces with democratic safeguards.

Additionally, the integration of structured analytical models such as TAM-SAM-SOM, empowered by Big Data and AI technologies, is crucial to effectively address market-specific challenges and improve decision-making accuracy (Davalas, 2023).

The European Media Freedom Act (EMFA) establishes a comprehensive framework to safeguard media pluralism and editorial independence across the EU. It addresses disinformation by promoting stronger editorial standards and fostering cooperation between media entities and regulatory authorities. Notably, EMFA does not mandate content removal but emphasizes transparency and accountability.

Key provisions include:

- **Article 17**: Enhances protections against the arbitrary removal of media content by online platforms. It requires platforms to provide clear justifications for content moderation decisions and ensures safeguards for lawful journalistic work.

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

- **Article 18**: Introduces a "media privilege" mechanism, obligating very large online platforms to notify media service providers before removing or restricting their content. This provision aims to prevent unjustified content takedowns and uphold freedom of expression.
- **Article 19**: Establishes the European Board for Media Services, a new independent body to coordinate national regulatory authorities, ensuring consistent application of media freedom principles and addressing systemic risks, including disinformation.

These measures collectively aim to create a balanced approach, countering disinformation while upholding fundamental rights and media freedoms.

The Council of Europe's *Guidelines on the Responsible Implementation of Artificial Intelligence Systems in Journalism* emphasize the importance of transparency and accountability when integrating AI into journalistic practices. Specifically, the guidelines recommend that media organizations clearly label AI-generated or manipulated content to maintain public trust and prevent the spread of misinformation. This includes the attribution and labelling of synthetic content, establishing best practice standards for transparency and human oversight, and addressing situations where the use of generative AI may conflict with human rights and public values.

These guidelines serve as a framework for media organizations to responsibly adopt AI technologies while upholding journalistic integrity and democratic values.

Article7 of the Paris Chapter (2023) under the title 'JOURNALISM DRAWS A CLEAR LINE BETWEEN AUTHENTIC AND SYNTHETIC CONTENT' underlines "*Journalists and media ...[should refrain from creating or using AI-generated content mimicking real-world captures and recordings or realistically impersonating actual individuals]".*

CDJM (2023) declares that " *The journalist does not use AI tools to generate images, sounds or videos whose realism risks misleading the public or leaving them in ambiguity, by submitting information contrary to the reality of the facts*" and suggests always human in the loop, which means editorial control, when content is generated by AI tools.

PFJU's Code of Ethics (Vlachopoulos, et al, 2025) states strictly in Art 4.(b) that "The content of any form that has been produced by AI must be faced by the journalist as non-verified and in Art 5 that Ai generated content must be marked and journalists must refrain from the creation or use of content being produced by AI and mimicking true downloads and recordings or realistically portraying real persons.

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Despite, the prevention of the AIA, media organisations and journalists have extra obligations for informing society and the public, providing accurate information. If the information in any format might mislead the audience, then it should not be published. So, in the case of the Greek Prime Minister, can the media publish the deepfake photo?

Yes, if the public figure was in New York shopping on 5th Street, and there is a need for a photograph (creative reasoning).

No, if the fact is not real (mis/mal/disinformation).

Conclusion

The media should not, according to their specific role, publish a deep fake if it never happened. The Act offers measures against deepfakes that are rather reactive, instead of proactive—meaning the Act does not necessarily prevent the creation of malicious deepfakes.

The case for the media is complex, while there are many different instances and case studies. For example, if a politician made a written statement, one could use the exact text and produce a voice or video abstract for creative purposes. The content should be watermarked. But, on the other hand, there are rules applying to personal data and GDPR which might be violated since there is voice and/or image usage without permission. But, for sure, the media can reproduce a helicopter crash using AI.

They can produce an image, a sound piece and a video of what someone did or said, but not of something that never did.

With the rapid growth of AI technology, users' awareness has also increased towards to AI content on the internet, as this contentis being generated daily due to the ease of access to free AI tools, which could lead to a possible malicious use. Therefore, regulating AI technology could help prevent the misuse of AI tools. The major issue is that individual Users have no responsibility according to the AIA, and they can produce as many and of any kind of deepfakes they desire. So the User Generated DeepFake Content might become a more important issue,and limitations at the LLM should occur. Otherwise, the deepfakes will be out there, and audiences might be confused about what is news and what is not. Restrictions and limitations are against the freedom of expression, but on the other hand, mis/disinformation is against the right to information.

References

Al Khatib, H., & Kayyali, D. (2019). YouTube Is Erasing History. New York Times. Retrieved from: https://www.nytimes.com/2019/10/23/opinion/syria-youtube-content-moderation.htm

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Associated Press (n.d.). ARTIFICIAL INTELLIGENCE: Leveraging AI to advance the power of facts. Associated Press. Retrieved 10.08.2024 from: https://www.ap.org/solutions/artificial-intelligence/

Bahtia, A. (2024, Aug. 25). When A.I.'s Output Is a Threat to A.I. The New York Times. Retrieved 28.02.2025 from: https://www.nytimes.com/interactive/2024/08/26/upshot/ai-synthetic-data.html?unlocked_article_code=1.F04.4KH2.UxTGEM9TDUZI&smid=url-share

Barret, M. P., & Hendrix, J. (2023). Safeguarding AI: Addressing the Risks of Generative Artificial Intelligence. NYU STERN.

Becker, K. B., Simon, F. M., & Crum, C. (2025). Policies in Parallel? A Comparative Study of Journalistic AI Policies in 52 Global News Organisations. Digital Journalism, 1–21. https://doi.org/10.1080/21670811.2024.2431519

Beckett, L. (2019, November 18). New powers, new responsibilities. A global survey of journalism and artificial intelligence. Blogs LSE. Retrieved March 18, 2025 from: https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C.-H. (2019). Artificial Intelligence and Journalism. Journalism & Mass Communication Quarterly, 96(3), 673-695. https://doi.org/10.1177/1077699019859901

Caswell, D., & Fang, S. (2024). AI in Journalism Futures. Initial Report. Open Society Foundations.

Davalas, A. (2020). Use of Big Data and AI tools to evaluate and assist startups. *International Journal of Social Science and Economic Research*, *5*(11), 3615-3624. https://doi.org/10.46609/IJSSER.2020.v05i11.019

Davalas, A. (2023). The importance of the TAM-SAM-SOM model and how Big Data and AI help. *International Journal of Social Science and Economic Research*, *8*(12), 3936-3944. https://doi.org/10.46609/IJSSER.2023.v08i12.016

Davalas, A., Charalabidis, Y., & Fenekoy, P. (2022). Startup Valuation with Artificial Intelligence: A SWOT Analysis. *European Journal of Economics*, *2*(2), 13-20. https://doi.org/10.33422/eje.v2i2.154

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Chesney, R., & Citron, D. K. (2019). Deepfakes and the new disinformation war. Foreign Affairs, 98(1), 147–155.

CDJM (2023). Journalisme et intelligence artificielle : les bonnes pratiques. Retrieved 17.03.2025 from: https://cdjm.org/journalisme-et-intelligence-artificielle-les-bonnes-pratiques/

Council of Europe. (2023). Guidelines on responsible AI systems in journalism. Retrieved 21.02.2025 from: https://rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6

Deck, A. (2024). ChatGPT is hallucinating fake links to its news partners' biggest investigations. NiemanLab. https://www.niemanlab.org/2024/06/chatgpt-is-hallucinating-fake-links-to-its-news-partners-biggest-investigations/ Accessed 05.04.2025

de-Lima-Santos, MF., Yeung, W.N. & Dodds, T. (2024). Guiding the way: a comprehensive examination of AI guidelines in global media. AI & Soc. https://doi.org/10.1007/s00146-024-01973-5Retrieved 5.04.2025

Deuze, M., and C. Beckett. 2022. Imagination, algorithms and news: Developing AI literacy for journalism. *Digital Journalism* 10 (10):1913–8. https://doi.org/10.1080/21670811.2022.2119152

Diakopoulos, N. (2019). Automating the news: how algorithms are rewritten by the media. Cambridge, Massachusetts: *Harvard University Press*, 2019.

Diakopoulos N. , Cools, H., LI, Ch., Heleberger, N., Kung, E., Rinehart, A. (April, 2024) *Generative AI in Journalism: The Evolution of Newswork and Ethics in a Generative Information Ecosystem* (Ed. Gibbs L.) AI, Media and Democracy. https://www.aim4dem.nl/out-now-generative-ai-in-journalism-the-evolution-of-newswork-and-ethics-in-a-generative-information-ecosystem/ Accessed 12.04.2025

Dierickx, L (2025, Apr. 8). *Balancing the scales: AI between verification and disinformation.*https://ohmybox.info/balancing-the-scales-ai-between-verification-and-disinformation/ Accessed 18.04.2025

Dierickx, L, LindénC,G., (2024)Dealing with biases and hallucinations: The ethical uses of (G)AI tools in the European news media sector, *ECREA 2024*, September 25, 2024 - Ljubljana.

Dixon, L., Li, J., Sorensen, J., Thain, N., & Vasserman, L. (2018). Measuring and Mitigating Unintended Bias in Text Classification. AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. ACM. 10.1145/3278721.3278729

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

European Broadcasting Union. (2022). *AI Act: high-risk AI systems need more nuance*. https://www.ebu.ch/news/2022/09/ai-act-high-risk-ai-systems-need-more-nuance Accessed 02.04.2025

European Union (2024). REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL (13 June 2024). Official Journal of the European Union.https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

European Union (2022). *Digital Services Act*.https://commission.europa.eu/publications/legal-documents-digital-services-act_en Accessed 28.02.2025

Ferrara, E. (2020). What types of COVID-19 conspiracies are populated by Twitter bots?. *First Monday*, *25*(6). https://doi.org/10.5210/fm.v25i6.10633

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Ghaffary, Sh. (2019, Aug. 15). The algorithms that detect hate speech online are biased against black people. https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter Accessed 20.04.2025

Goldman,Sh, Kahn, J. (2025, Apr.16). OpenAI updated its safety framework—but nolonger sees mass manipulation and disinformation as a critical risk. *Fortune*. https://fortune.com/2025/04/16/openai-safety-framework-manipulation-deception-critical-risk/ Accessed 24.04.2025

Helberger, N., & Diakopoulos, N. (2022). The European AI act and how it matters for research into AI in media and journalism. *Digital Journalism, online first*. https://doi.org/10.1080/21670811.2022.2 082505

Helberger, N., Eskens, S., van Drunen, M., Bastian, M., & Moeller, J. (2019). *Implications of AI-driven tools in the media for freedom of expression*. Institute for Information Law (IViR). https://rm.coe.int/coe-ai-report-final/168094ce8f

Helberger, N., Van Drunen, M., Eskens, S., Bastian, M., & Moeller, J. (2020). A freedom of expression perspective on AI in the media–with a special focus on editorial decision making on social media platforms and in the news media. *European Journal of Law and Technology*, *11*(3).

Henestrosa, A. L., Greving, H., & Kimmerle, J. (2023). Automated journalism: The effects of AIauthorship and evaluative information on the perception of a science journalismarticle. *Computers in Human Behaviour*, 138, 107445.

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Jiang, Y., Du, J., Xu, L., & Zhang, Y. (2020). Towards realistic and controllable voice cloning with limited training data. *IEEE Transactions on Neural Networks and Learning Systems*, 31(5), 1401–1414. https://doi.org/10.1109/TNNLS.2019.2913359

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4401–4410). https://doi.org/10.1109/CVPR.2019.00453

Kreps, S., McCain, R., & Brundage, M. (2023). Generative AI and democracy: Risks and recommendations. *Brookings Institution*. https://www.brookings.edu/research/generative-ai-and-democracy/ Accessed 01.03.2025

Lindblom, T., Lindell, J., & Gidlund, K. (2022). Digitalizing the Journalistic Field: Journalists' Views on Changes in Journalistic Autonomy, Capital and Habitus. *Digital Journalism*, *12*(6), 894–913. https://doi.org/10.1080/21670811.2022.2062406

Madiega, T. (2024). *The Artificial Intelligence Act: General Purpose AI and the EU's Risk-Based Approach*. European Parliamentary Research Service (EPRS), PE 698.792.

Monti, M. (2019, Jan 18) Automated Journalism and Freedom of Information: Ethical and Juridical Problems Related to AI in the Press Field. *Opinio Juris in Comparatione, 1/2018* , Available at SSRN: https://ssrn.com/abstract=3318460

Newman, N. (2020, Jan, 9). Journalism, Media, and Technology Trends and Predictions 2020. https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2020 Accessed 25.04.2025

Nightingale, S., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, 119(8), e2120481119. https://doi.org/10.1073/pnas.2120481119

Nieborg, D. B., and T. Poell. 2018. The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society* 20 (11):4275–92. doi:10.1177/1461444818769694

OpenAI (2025, Apr.15).Our updated Preparedness Framework. *OpenAI*. https://openai.com/index/updating-our-preparedness-framework/ Accessed 21.04.2025

Orwat, C., Bareis, J., Folberth, A., Jahnel, J., & Wadephul, C. (2024). Normative Challenges of Risk Regulation of Artificial Intelligence. *NanoEthics*, *18*(2), 11.

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

OSCe(n.d). Freedom of expression in the age of artificial intelligence: the risks and challenges to online speech and media freedom. OSCe.https://www.osce.org/saife/essay, Accessed 15.04.2025

PAI Staff. 2023. PAI seeks public comment on the AI procurement and use guidebook for newsrooms.https://partnershiponai.org/pai-seeks-public-comment-on-the-ai-procurement-guidebook-for-newsrooms/.Accessed 22.02, 2025

Panagopoulos, M.,A., Kapos, P., Triantafyllou, S. (2025). AI usage in the Greek Newsrooms. Proceedings of the 3RD Annual CCLAB (29 &30 June 2024) *3*(1), 14–25. (In Greek)https://doi.org/10.12681/cclabs.8034

Pavlik, J. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. Retrieved from: https://journals.sagepub.com/doi/full/10.1177/10776958221149577

Porlezza, C. (2023). Promoting responsible AI: A European perspective on the governance of artificial intelligence in media and journalism. Communications: The European Journal of Communication Research, 48(3), pp. 370-394. doi: 10.1515/commun2022-0091

Porlezza, C. & Ferri, G. (2022). The missing piece: Ethics and the ontological boundaries of automated journalism. *#ISOJ Journal*, 12(1), 71-98]

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1–11). https://doi.org/10.1109/ICCV.2019.00010

RSF (2023). RSF and 16 partners unveil Paris Charter on AI and Journalism. https://rsf.org/en/rsf-and-16-partners-unveil-paris-charter-ai-and-journalism Accessed 15.03.2025

Seipp, T. J., Helberger, N., Vreese, C., Ausloos. J. (2023). Dealing with opinion power in the platform world: Why we really have to rethink media concentrationlLaw. *Digital Journalism* 11 (8):1542–67. doi:10.1080/21670811.2022.2161924

Shin D., Zaid B., Biocca F., Rasul A. (2022). In platforms we trust? Unlocking the black-box of news algorithms through interpretable AI. Journal of Broadcasting and Electronic Media, 66(2), 235–256. https://doi.org/10.1080/08838151.2022.2057984

 Copyright © IJSSER 2025, All rights reserved

ISSN: 2455-8834

Volume:10, Issue:08 "August 2025"

Simon, F. (2022). Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism* 10 (10):1832–54. doi:10.1080/21670811.2022.2063150

Simon, F. (2024). Artificial Intelligence in the News: How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena. Tow Report. Columbia Journalism Review. https://www.cjr.org/tow_center_reports/artificial-intelligence-in-the-news.php, Retrieved 14.03.2025

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131–148. https://doi.org/10.1016/j.inffus.2020.07.007

Unesco (2021). Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137 Retrieved 20.08.2024

van Dijk, J. A. G. M. (2020). *The digital divide*. Cambridge, England:Polity.

Veale, M., Borgesius,Z., Fr. (2021) Demystifying the Draft EU Artificial Intelligence Act Computer Law Review International (2021) 22(4) 97-112, Available at SSRN: https://ssrn.com/abstract=3896852

Vlachopoulos Sp., Vassilopoulos V., Vrachatis A. Eleftheroglou N., Kaimaki V., Panagopoulos M. A., Tassis Sp., Tsipra M. (2025). Code of Conduct for the Use of Generative Artificial Intelligence by Journalists. PFJU

West, J. D., & Bergstrom, C. T. (2021). Misinformation in and about science. *Proceedings of the National Academy of Sciences*, *118*(15), e1912444117.

York, J. & McSherry, C. (2019). Content Moderation is Broken. Let Us Count the Ways. *Electronic Frontier Foundation*. Retrieved from: https://www.eff.org/deeplinks/2019/04/content-moderation-broken-let-us-count-ways

YouTube (2024, Mar.18) New Disclosures and Labels for Generative AI Content on YouTube. https://support.google.com/youtube/thread/264550152/new-disclosures-and-labels-for-generative-ai-content-on-youtube?hl=en . Accessed01.05.2025

 Copyright © IJSSER 2025, All rights reserved