LWP | USFQ

LAW WORKING PAPERS

LAW WORKING PAPERS

# What is real? Understanding Deepfakes and its implications

# Que es real? Entendiendo las Deepfakes y sus implicaciones

**Hannah Kirschbaum**
**Tania Micaela Pineda González**
**Charlotte Senecat**
**Julia Soszyńska**

2026 / 04

**USFQ Law Working Papers**
Colegio de Jurisprudencia
Universidad San Francisco de Quito USFQ
Quito, Ecuador

**En contestación a:** n/a

**Recibido:** 2026 / 02 / 18 **Difundido:** 2026 / 04 / 16

**Materias:** Tech Law / Protección de Datos

**DOI:** https://doi.org/10.18272/0qg9bv87

**Citación sugerida:** Kirschbaum, Hannah. Pineda Gonzales, Tania Micaela. Senecat, Charlotte. Soszyńska, Julia "What is real? Understanding Deepfakes and its implications". *USFQ Law Working Papers*, 2026/04, https://doi.org/10.18272/0qg9bv87

Acerca de

# USFQ Law Working Papers

USFQ Law Working Papers es una serie académico-jurídica de difusión continua, con apertura autoral para profesionales y de acceso abierto. Introduce en Ecuador un novedoso tipo de interacción académica que, por sus características particulares, tiene el potencial de ser pionero en rediseñar el discurso público del Derecho. Su objetivo es difundir documentos de trabajo (*working papers*) con impacto jurídico, que pueden abarcar cualquier asunto de las ramas de esta ciencia y sus relaciones con otras áreas del conocimiento, por lo que está dirigida a la comunidad jurídica y a otras disciplinas afines, con alcance nacional e internacional.

USFQ Law Working Papers difunde artículos académicos y científicos originales, entrevistas, revisiones o traducciones de otras publicaciones, entre otros, en español o inglés. Los contenidos son de exclusiva responsabilidad de sus autores, quienes conservan la titularidad de todos los derechos sobre sus trabajos. La difusión de los documentos es determinada, caso a caso, por el Comité Editorial. Se prescinde de la revisión por pares con el fin de dar a toda la comunidad académica la oportunidad de participar, mediante la presentación de nuevos trabajos, en la discusión de todos los contenidos difundidos.

USFQ Law Working Papers nace, se administra y se difunde como una iniciativa de la profesora Johanna Fröhlich (PhD) y un grupo de *alumni* del Colegio de Jurisprudencia de la Universidad San Francisco de Quito USFQ (Ecuador). Su difusión se realiza gracias al apoyo del Instituto de Investigaciones Jurídicas USFQ (Ecuador).

***Más información:*** http://lwp.usfq.edu.ec

# WHAT IS REAL? UNDERSTANDING DEEPFAKES AND ITS IMPLICATIONS

## *¿QUÉ ES REAL? ENTENDIENDO LAS* DEEPFAKES *Y SUS IMPLICACIONES*

Hannah Kirschbaum[1]
Tania Micaela Pineda González[2]
Charlotte Senecat[3]
Julia Soszyńska[4]

**ABSTRACT**

Deepfakes have been widely used for several purposes in the last years. The examination of its technical architecture is key to understanding the resulting societal and legal challenges. It involves computational mechanisms like Generative Adversarial Networks (GANs), Autoencoders, and Convolutional Neural Networks (CNNs), as well as hardware and networking infrastructure for high-fidelity synthesis. Deepfakes disrupt non-AI sectors, particularly cybersecurity biometrics, media broadcast credibility, and legal forensic analysis, where the phenomenon of the "Liar's Dividend" threatens the integrity of digital evidence. While acknowledging potential benefits in creative industries and assistive medicine, there are severe ethical risks, including the proliferation of non-consensual sexualized deepfakes and political disinformation. Therefore, it is necessary to evaluate the sufficiency of current law, in this case, the European Union's regulatory framework, specifically the AI Act, DSA, and GDPR.

**KEYWORDS**

Deepfakes, disinformation, reality, image generation, AI Act, Digital Service Act, General Data Protection Regulation.

***RESUMEN***

*Los* deepfakes *se han utilizado ampliamente para diversos fines en los últimos años. El estudio de su arquitectura técnica es fundamental para comprender los retos sociales y jurídicos que plantean. Implican mecanismos computacionales como las redes generativas adversarias (RGA), los autoencodificadores y las redes neuronales convolucionales (RNN), así como infraestructura de hardware y redes para la síntesis de alta fidelidad. Los* deepfakes *afectan sectores no relacionados a la IA, en particular la biometría de la ciberseguridad, la credibilidad de los medios de comunicación y el análisis forense jurídico, donde el fenómeno del «dividendo del mentiroso» amenaza la integridad de las pruebas digitales. Si bien se reconocen los posibles beneficios en las industrias creativas y la medicina asistencial, existen graves riesgos éticos, como la proliferación de* deepfakes *sexualizados no consentidos y la desinformación política. Por lo tanto, es necesario evaluar la suficiencia de regulaciones actuales, en este caso el marco regulatorio de la Unión Europea, concretamente la Ley de IA, la LSD y el RGPD.*

***PALABRAS CLAVE***

Deepfakes*, desinformación, realidad, generación de imágenes,* AI Act, *Ley de Servicios Digitales, Reglamento General de Protección de Datos.*

[1] Independent Researcher. Email: hannah.kirschbaum@gmx.net
[2] Independent Researcher. Email: taniamica2002@hotmail.com
[3] Independent Researcher. Email: senecat.charlottepro@gmail.com
[4] Independent Researcher. Email: juliasosik1323@gmail.com

**Table of Contents**

## Bibliography

### Primary sources

Regulation (EU) 2024/1689 (AI Act)
Regulation (EU) 2016/679 (GDPR)

### Secondary sources

Allen C and others, 'What You See Is Not What You Know: Studying Deception in Deepfake Video Manipulation' (2023) *Journal of Cybersecurity Education, Research and Practice*

Bailey M and Cunningham S, 'Introduction to Computer Graphics' (SIGGRAPH Asia Conference, September 2010)

Bass D. F., Shlaimon N., Penning N., J., 'Deepfakes and Intellectual Property Legal Issues' (2025).

Bellas G., 'Deepfakes in the Courtroom: Problems and Solutions' (March 2025).

Bello RW and others, 'Cloud-based Face Swapping Application' (Pretoria 2024)

Bernaciak C and Ross D, 'How Easy Is It to Make and Detect a Deepfake?' (Carnegie Mellon University Software Engineering Institute, 2022) accessible at: https://www.sei.cmu.edu/blog/how-easy-is-it-to-make-and-detect-a-deepfake

Birchwood University, 'Exploring the Bright Side of Deepfake Technology' (2024), accessible at https://www.birchwoodu.org/exploring-the-bright-side-of-deepfake-technology/

Broklyn P., Egon A. and Shad R., 'Deepfakes and Cybersecurity: Detection and Mitigation' (July 03, 2024)

Constantineanu, C., 'Deepfakes: The New Frontier in Political Disinformation' (2024), accessible at https://thesecuritydistillery.org/all-articles/deepfakes-the-new-frontier-in-political-disinformation?utm

Dan V., 'Deepfakes as a Democratic Threat: Experimental Evidence Shows Noxious Effects That Are Reducible Through Journalistic Fact Checks' (2025).

'Deepfake, n' *Oxford English Dictionary* (Oxford University Press, March 2023) accessible at https://www.oed.com/dictionary/deepfake_n

Eitren W., 'Deepfakes in the AI Act' (2024).

Eliot L, 'Explaining Deepfakes Versus Cheap Fakes and the Role of Generative AI' (Forbes, 25 June 2024) accessible at https://www.forbes.com/sites/lanceeliot/2024/06/25/cheap-fakes-and-rescuing-humankind-via-generative-ai/

European Commission, 'AI Act' (Last update: 5 December 2025)

European Commission, 'The Digital Services Act' (Last update: 15 December 2025)

Hameleers M, 'Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting' (2024) 36 *International Journal of Public Opinion Research*.

Finnerty N., 'Tackling the regulation of sexually explicit deepfakes' (2024).

Flynn A. et al., Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging form of Image-Based Sexual Abuse, The British Journal of Criminology, Volume 62, Issue 6, (November 2022), Pages 1341–1358

Flynn A. et. al., Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of Non-Consensually Created and Shared Sexualized Deepfake Imagery (2025), Journal of Interpersonal Violence.

Fragale M., Grilli V., 'Deepfake, Deep Trouble: The European AI Act and the Fight Against AI-Generated Misinformation'(2024)

GAO, 'Science & Tech Spotlight: Deepfakes' (2020).

Goodfellow I and others, 'Generative Adversarial Networks' (2014) arXiv:1406.2661

GRC Solutions, 'What is Cybersecurity: Definition and Best Practices' (2025)

G. Vivek, ‘What is Cyber security: A Complete Beginner’s Guide’ (last updated 2025
Hoek, S. et al., 'Promising for patients or deeply disturbing? The ethical and legal aspects of deepfake therapy' (2024).
IBM, ‘What is PyTorch?’ accessible at https://www.ibm.com/think/topics/pytorch
IBM Developer, ‘Learn the basics of computer vision and object detection’ (2020) accessible a thttps://developer.ibm.com/articles/learn-the-basics-of-computer-vision-and-object-detection/
Insikt Group, 'Targets, Objectives, and Emerging Tactics of Political Deepfakes' (2024), accessible at https://assets.recordedfuture.com/insikt-report-pdfs/2024/ta-2024-0924.pdf
Jain A and Jain N, ‘A Peek into Computer Generated Imagery Pipeline’ (2017) 10(2) *Bionano Frontier* 220
Kalmykov, M., 'Positive Applications for Deepfake Technology' (2023), accessible at https://www.dataart.com/blog/positive-applications-for-deepfake-technology-by-max-kalmykov
Kira B., 'When non-consensual intimate deepfakes go viral: The insufficiency of the UK Online Safety Act' (2024), Computer Law & Security Review: The International Journal of Technology Law and Practice 54 (2024) 106024.
Kuźnicka-Błaszkowska D., Kostyuk N., ‘Emerging need to regulate deepfakes in international law: the Russo–Ukrainian war as an example‘ (2025), Journal of Cybersecurity, Volume 11, Issue 1
Łabuz M., ‘Regulating Deep Fakes in the Artificial Intelligence Act’, Volume 2, No. 1 (2023)
Lampe C, ‘“Cheap Fake” Video Making the Rounds Today Likely Won’t Be the Last’ (University of Michigan School of Information) accessible at https://www.si.umich.edu/about-umsi/news/cheap-fake-video-making-rounds-today-likely-wont-be-last
Lees D., 'Deepfakes in documentary film production: images of deception in the representation of the real' (Studies in Documentary Film, 108-129, 2023)
Maio A., 'What is VFX? Defining the Term and Creating Impossible Worlds' (2025), accessible at https://www.studiobinder.com/blog/what-is-vfx/
Manovich L, *The Language of New Media* (MIT Press 2020)
Martinez, O. N. et al., 'Possible Health Benefits and Risks of DeepFake Videos: A Qualitative Study in Nursing Students' (2024)
Mirsky Y and Lee W, ‘The Creation and Detection of Deepfakes: A Survey’ (2021) 54(1) *ACM Computing Surveys* 1
Mitchell A., ‘Deepfaked Evidence: What Case Law Tells Us About How the Rules of Authenticity Needs to Change’, Berkeley Tech. L.J.: Blog (June 23, 2025)
Naffi N., 'Deepfakes and the crisis of knowing', UNESCO (2025) accessible at https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing?utm
Negreiro M., 'Children and Deepfakes' (2025), European Parliamentary Research Service, accessible at https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775855/EPRS_BRI(2025)775855_EN.pdf
Livingston C., ‘Deepfakes in the Courtroom: Challenges in Authenticating Evidence and Jury Evaluation’, University of Baltimore Law Review (2025)
O’Brien S. R., 'Positive Use Cases of “Deepfakes”' (2025), accessible at https://www.wilsoncenter.org/article/positive-use-cases-deepfakes
Pandit AR and Kirdat TV, ‘Impact of AI in the Animation Industry’ (2024) 12(3) *International Journal for Research in Applied Science & Engineering Technology* 2828

Paravision, 'A Practical Guide to Deepfake Detection' (White Paper) https://www.paravision.ai/whitepaper-a-practical-guide-to-deepfake-detection/ accessed 19 December 2025

Paris B and Donovan J, Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence (Data & Society 2019) accessible at https://datasociety.net/library/deepfakes-and-cheap-fakes/

Paszke A and others, 'PyTorch: An Imperative Style, High-Performance Deep Learning Library' (2019) 32 *Advances in Neural Information Processing Systems* 8024

Pawelec M., Łabuz M. 'Non-Consensual Sexualising Deepfakes - Threats and Recommendations for Legal and Societal Action' (2025), CEE Digital Democracy Watch: Warsaw, accessible at https://ceeddw.org/wp-content/uploads/2025/05/NCII_DeepFakes_ThreatsRecommendations.pdf

Perov I and others, 'DeepFaceLab: Integrated, Flexible and Extensible Face-Swapping Framework' (2020) arXiv:2005.05535

Powell P and Smalley I, 'What is a Central Processing Unit (CPU)?' (IBM Think, 2025) accessible at https://www.ibm.com/think/topics/central-processing-unit

Raina R, Madhavan A and Ng AY, 'Large-scale Deep Unsupervised Learning using Graphics Processors' (International Conference on Machine Learning 2009)

Rana MS, 'CGI vs AI in 2026. Which Visual Workflow Should You Use?' (7CGI, 8 November 2025) accessible at https://7cgi.com/blog/cgi-vs-ai-key-differences

Rehak B, 'Computer-Generated Imagery' (2011) *Oxford Bibliographies: Cinema and Media Studies accessible at* https://doi.org/10.1093/obo/9780199791286-0068

Rini M, 'Deepfakes: The Evolution of Synthetic Media' (2020) 63(1) *Communications of the ACM* 86

Robins-Early N., 'How did Donald Trump end up posting Taylor Swift deepfakes?', the Guardian (2024), accessible at: https://www.theguardian.com/technology/article/2024/aug/24/trump-taylor-swift-deepfakes-ai

Stagg M., 'Why deepfakes are the next big threat to broadcast credibility' (2025)

'Text-to-image model' (Wikipedia, 2025) https://en.wikipedia.org/wiki/Text-to-image_model accessed 19 December 2025

Tolosana R and others, 'DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection' (2020) arXiv:2001.00179

University of Virginia Information Security, 'What the Heck Is a Deepfake?'accessible at https://security.virginia.edu/deepfakes

US Government Accountability Office, 'Deepfakes: AI-Generated Video and Audio Can Be Used for Exploitation and Disinformation' (*Technology Assessment* GAO-20-379SP, 2020)

Seng JKP and others, 'Artificial Intelligence (AI) and Machine Learning for Multimedia and Edge Information Processing' (2022) 11 *Electronics* 2239

V. Soni, Infosys, 'Deepfake and Its Impact on Cybersecurity - A New Frontier to Address'

Wack M., Parry D. A., 'Synthetic Diversity: Examining the Effects of Ethnic Targeting Using AI-Generated Political Ads', (2025), International Journal of Communication 19(2025), 3736–3760

Westerlund M, 'The Emergence of Deepfake Technology: A Review' (2019) 9(11) *Technology Innovation Management Review* 39

Whisperly, 'Deepfake Regulation: Chaos That Matters Now More Than Ever' (2025)

Wazid M., 'A secure Deepfake Mitigation Framework: Architecture, Issues, Challenges, and Societal Impact', Cyber Security and Applications Volume 2 (2024)

## I. I. Introduction

The quick development of Artificial Intelligence and machine learning in recent years have resulted in the creation of deepfakes. A deepfake is "the manipulation of artificial generation (synthesis) of audio, video or other forms of digital content to make it appear that a particular event occurred, or that someone behaved or looked differently than they actually did"[5]. This new technology has brought innovative prospects and serious ethical dilemmas. The impact that deepfakes have had on social media and the spread of disinformation has increased and even changed the democratic debate online[6]. The use of improved methods to create deepfakes has also improved its quality, impeding the recognition of modified materials, especially by less experienced users. For that reason, we will examine the technology underlying deepfakes, AI implications and societal, ethical and legal implications including the benefits and risk of this technology involving its contribution to the production of child sexual abuse material (CSAM).

## II. II. Technical overview of deepfakes

Before addressing the complex technical architecture and main components of deepfakes, it is essential to establish a clear definition by distinguishing between the mainstream interpretation and the computational perspective adopted by the engineering standpoint.

Generally speaking, this concept refers to artificial media (such as images, videos, or audio) generated by AI to appear convincingly real.[7] The phenomenon is mostly known for face-swapping or voice-cloning, used to create scenes or conversations that never actually happened. Classic definitions describe the deepfake as "any of various media, especially a video, that has been digitally manipulated to replace one person's likeness convincingly with that of another, often used maliciously to show someone doing something that he or she did not do".[8]

A common misconception is to compare deepfakes to standard editing tools like Photoshop or social media filters.[9] This comparison is technically inaccurate. While traditional

---

[5] European Data Protection Supervisor, 'Deepfake detection' (2023); available at <https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/deepfake-detection_en> .

[6] Chesney R., Citron D.K., 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security', 107 California Law Review 1753, 1777 (2019).

[7] 'What the Heck Is a Deepfake?' (University of Virginia Information Security) <https://security.virginia.edu/deepfakes>

[8] 'Deepfake, n' Oxford English Dictionary (Oxford University Press, March 2023) <https://www.oed.com/dictionary/deepfake_n>

[9] 'What the Heck Is a Deepfake?' (University of Virginia Information Security) <https://security.virginia.edu/deepfakes>

software merely alters existing images, deepfake algorithms rely on generative models that produce entirely new pixels. Instead of simple retouching, they learn and predict visual patterns to achieve a far more realistic result.[10] Since deepfakes are a broad term, the main focus of this work will be on deepfakes involving face-swapping techniques, rather than the ones created from scratch such as text-to-video or text-to-images models. Once this distinction has been established, we will examine the computational architecture that underpins the technology.

As its name implies, "the word deepfake is a combination of the words 'deep learning' and 'fake' and primarily relates to content generated by an artificial neural network, a branch of machine learning".[11] From a technical perspective, although there are a wide variety of neural networks, most deepfakes are created using variations or combinations of generative networks and encoder decoder networks. Specifically, these architectures include Generative Adversarial Networks (GANs) and autoencoders.[12] The creation of deepfake videos relies on a broader ecosystem that goes far beyond the algorithm itself. It requires the coordinated use of five key elements: high-performance hardware, specialized software, networking components, advanced AI models, and human intervention.

1. Deepfakes generation process

The process to generate deepfakes encompasses three main steps. Firstly, collecting data; secondly, training the model; thirdly, generation and post-processing [13] . During the collection of data there are human active roles, mainly, data engineers, who will oversee the design of the infrastructure required to extract vast amounts of data including voice recordings, images and videos of the target person. AI researchers are involved in defining the data requirements and applying statistical methods to ensure the quality relevance and the diversity of the dataset. The quality and amount of this data will impact on the performance of the model. After the data has been collected, it must be organized to ensure consistency by aligning,

[10] US Government Accountability Office, Deepfakes: AI-Generated Video and Audio Can Be Used for Exploitation and Disinformation (Technology Assessment GAO-20-379SP, 2020)

[11] Enes Altuncu, Virginia NL Franqueira and Shujun Li, 'Deepfake: Definitions, Performance Metrics and Standards, Datasets, and a Meta-review' (2024) 7 Frontiers in Big Data 1400024,1-2

[12] Yisroel Mirsky and Wenke Lee, 'The Creation and Detection of Deepfakes: A Survey' (2020) arXiv:2004.11138,1

[13] Timonera Kathryn, 'What Is Deepfake Technology? Ultimate Guide To AI Manipulation' (EWeek, 11 October 2024) <https://www.eweek.com/artificial-intelligence/deepfake/>

normalizing and adjusting the features[14]. Then, the training of the model begins, and involves several subfields of AI, like deep learning and generative models[15].

To train the model, Generative Adversarial Networks (GANs), Autoencoders and Convolutional Neural Networks (CNNs) are employed. GANs are the most popular method due to the high level of realism that it provides. It contains pairs of neural networks, a generator that is learning by trying to trick the discriminator with the produced synthetic images or media; and a discriminator that is learning to distinguish between real data and synthetic data. Each component improves realism through adversarial training; hence, it's a zero-sum game against each other[16]. There are several applications of this technology, with models like StyleGAN, that can manipulate latent spaces, altering features such as age, gender and facial expressions with a result of a high-quality image. This model can be especially relevant and used for videos where one person's face is replaced with a completely different one. Another application is WAV2lip that creates deepfakes where the person appears to say things they haven't said by creating a speech and synchronizing lip movements with audio.[17]

For earlier image processing methods autoencoders were utilized to reduce dimensionality or compress images. In the process of deepfakes, this technology is used by training a neural network to encode and decode images or videos. It has three main components, encoder, latent space and decoder. The encoder is responsible for getting all the data from the images and compressing it with only the most important features such as skin tone, skin texture, facial expression, structure of the face, state of eyes, and any other necessary features to transmit it to the latent space. The latent space is in charge of finding the mapping of patterns and structural similarities among one or more data points in a latent code and sharing it with the decoder. Finally, the decoder is left with reconstructing the image, making it as realistic as the original image. In this process the autoencoder is provided with both authentic and altered media. Therefore, it learns to produce comparable data representation for the latent space of

[14] Regan Gabe, 'How Deepfakes Are Made' (Reality Defender, 16 June 2025) <https://www.realitydefender.com/insights/how-deepfakes-are-made>

[15] Bendiab Gueltoum and others, 'Deepfakes in digital media forensics: Generation, AI-based detection and challenges' [2025] 88. Journal of Information Security and Applications. <https://doi.org/10.1016/j.jisa.2024.103935.>

[16] Mirkute Dipti, 'Truth vs Fabrication: Exploring AI's Role in the Detection of Deepfakes' [2025] 12(10) International Journal of Informative & Futuristic Research <https://ijifr.org/pdfsave/24-06-2025278IJIFR-V12-E10-008.pdf>

[17] Ibid.

both real and deepfake materials. Thus, the production of highly convincing deepfake content is facilitated with the reconstruction performed by the decoder[18].

In the steps of facial and visual processing of deepfake generation, Convolutional Neural Networks (CNNs) are a fundamental aspect involved in Autoencoders and GANs. It processes input images by passing them through several layers which can be understood as filters. The first layers identify simple features, such as edges and lines, while the deeper layers recognize patterns, shapes, and eventually more complex whole objects. This method of extracting features in a hierarchical way is what makes CNNs so powerful for image recognition and processing[19]. The entire process to generate high quality deepfakes could require millions of labelled data points for training. Therefore, models must be trained with high-power processors, such as a GPU or an NPU, if they are to produce results quickly enough to be useful.

2. Hardware involved in deepfakes

While AI algorithms are the core of deepfakes generation, they can't function without the supporting infrastructure (hardware and software component) that allows them to work in practice.

Deepfake generation requires specific hardware that provides the computing power needed to train and execute complex neural network models.[20] Unlike conventional video editing, deepfake creation involves complex mathematical calculations that demand parallel processing capabilities.[21] The Graphic Processing Units (GPU) serves as the key hardware component.[22] In fact, it accelerates the matrix computation (operations on large arrays of data) used by neural networks.[23] It makes the model training feasible within a reasonable time. However, the GPU does not work alone. The Central Processing Unit (CPU) handles data loading and pre-processing tasks.[24] It reads video files and prepares datasets before transferring

[18] Alanazi Sami and Asif Seemal, 'Exploring deepfake technology: creation, consequences and countermeasures' [2024] 6 Human-Intelligent Systems Integration <https://doi.org/10.1007/s42454-024-00054-8>

[19] Google cloud, 'What is a convolutional neural network?' (Subjects of Google Cloud) <https://cloud.google.com/discover/what-are-convolutional-neural-networks>

[20] Yisroel Mirsky and Wenke Lee, 'The Creation and Detection of Deepfakes: A Survey' (2020) arXiv:2004.11138, 5-6

[21] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio, 'Generative Adversarial Networks' (2014) arXiv:1406.2661, 1

[22] Caitlyn Bernaciak and Dena Ross, 'How Easy Is It to Make and Detect a Deepfake?' (Carnegie Mellon University Software Engineering Institute, 2022) <https://www.sei.cmu.edu/blog/how-easy-is-it-to-make-and-detect-a-deepfake>

[23] Yisroel Mirsky and Wenke Lee, 'The Creation and Detection of Deepfakes: A Survey' (2021) 54(1) ACM Computing Surveys 7–8.

[24] Adam Paszke and others, 'PyTorch: An Imperative Style, High-Performance Deep Learning Library' (2019) 32Advances in Neural Information Processing Systems 8026–8037

them to the GPU. The CPU also manages the workflow by distributing tasks and controlling data movement across the system.[25]

To support these processors, memory components are needed. Video Random Access Memory (VRAM) stores the visual data that the model processes at each training step. It provides the GPU with immediate access to images during computation.[26] Classic Random Access Memory (RAM) provides temporary storage for pre-processing operations.[27] It holds datasets and application data while the CPU prepares them for the GPU processing. Finally, high-performance storage is needed because deepfake training requires managing massive amounts of video data. Therefore, it allows fast reading of large files but also ensures that processors receive data instantly when they need.[28]

3. Software involved in deepfakes

Moreover, the deepfake process involves three main types of distinct software. At the beginning of the process, video pre-processing software prepares raw footage for model training. Neural networks require structured, aligned and clean data to function effectively.[29] For example, OpenCV converts raw video data into a model-acceptable format by extracting individual frames, resizing them and aligning faces to standard positions.[30]

Once the data is prepared, machine learning frameworks take over the training phase. Framework like PyTorch manage the model architecture, control the training procedures and learning process.[31] They determine how the neural network learns patterns from the prepared data.[32] At the final stage, deepfake applications apply the trained model to produce output. Applications like DeepFaceLab used the trained model to swap faces in videos or photos.[33]

---

[25] Phill Powell and Ian Smalley, 'What is a Central Processing Unit (CPU)?' (IBM Think, 2025) <https://www.ibm.com/think/topics/central-processing-unit>

[26] Stock Must Go, 'GPU VRAM (Graphics Processing Unit Video Random Access Memory)' (Stock Must Go) <https://www.stockmustgo.co.uk/pages/gpu-vram-graphics-processing-unit-video-random-access-memory>

[27] What is RAM (Random Access Memory)? (Lenovo Glossary, 2025) <https://www.lenovo.com/gb/en/glossary/what-is-ram/#:~:text=Random-access memory (RAM),use while it is running>

[28] What are the System Requirements for Creating Deepfakes? (Massed Compute FAQ Answers, 2025) <https://massedcompute.com/faq-answers/?question=What%20are%20the%20system%20requirements%20for%20creating%20deepfakes%3F>

[29] Yisroel Mirsky and Wenke Lee, 'The Creation and Detection of Deepfakes: A Survey' (2021) 54(1) ACM Computing Surveys 1.

[30] IBM Developer, 'Learn the basics of computer vision and object detection' (2020) <https://developer.ibm.com/articles/learn-the-basics-of-computer-vision-and-object-detection/>

[31] Adam Paszke and others, 'PyTorch: An Imperative Style, High-Performance Deep Learning Library' (2019) 32 Advances in Neural Information Processing Systems 8024

[32] IBM, 'What is PyTorch?'< https://www.ibm.com/think/topics/pytorch>

[33] Ivan Perov and others, 'DeepFaceLab: Integrated, Flexible and Extensible Face-Swapping Framework' (2020) arXiv:2005.05535, section 3.

Then, they take the learned patterns and apply them to the new footage. It allows the user to create the final deepfake content by using simplified interfaces.[34]

4. Networks in deepfakes processing

Finally, for all of these different components to function as a complete and working system, networking infrastructures enable their connection and coordination. First, deepfake training requires moving massive video datasets between different computers.[35] Data centers use high-speed networks to connect GPU/CPU servers with storage systems.[36] Many creators use cloud platforms instead of local computers. Cloud processing approaches collect data and send it to centralized servers for information processing.[37] In practice, users upload their video data through network connections to cloud servers. After security validation, serverless functions process the data and store results in cloud storage, before notification services return the processed output to users.[38]

Once a deepfake model is trained, networking infrastructure distributes it to end-users. Networks deliver these models through APIs (application programming interfaces) that connect applications to processing servers. [39] Therefore, when users interact with face-swapping applications, their requests pass through security firewalls before reaching API gateways that trigger processing functions, ultimately delivering results through notification systems.[40]

5. Related media manipulation technologies

a. Deepfakes and Computer-Generated Imagery

Deepfakes are digitally manipulated videos created using artificial intelligence that can deceive viewers into believing they reflect real events.[41] A closely related and historically connected technology is Computer-Generated Imagery (CGI). CGI emerged in the late 1960s and progressively became central to film production, particularly in animation and visual

[34] Mika Westerlund, 'The Emergence of Deepfake Technology: A Review' (2019) 9(11) Technology Innovation Management Review under 'Applications and Societal Implications'

[35] Jasmine Kah Phooi Seng and others, 'Artificial Intelligence (AI) and Machine Learning for Multimedia and Edge Information Processing' (2022) 11 Electronics 2239, 9

[36] Rajat Raina, Anand Madhavan and Andrew Y Ng, 'Large-scale Deep Unsupervised Learning using Graphics Processors' (International Conference on Machine Learning 2009) 934.

[37] ibid 1.

[38] Rotimi Williams Bello and others, 'Cloud-based Face Swapping Application' (Pretoria 2024) 4.

[39] ibid 1.

[40] ibid 4.

[41] Cathryn Allen and others, 'What You See Is Not What You Know: Studying Deception in Deepfake Video Manipulation' (2023) Journal of Cybersecurity Education, Research and Practice

effects.[42] Both CGI and deepfakes pursue the same objective. They aim to generate images that appear realistic and artificial images become nearly impossible to distinguish from reality.[43] However, despite this common goal, they differ technically in how images are created and controlled.

First, CGI relies on deterministic logic.[44] Every element of this image is manually designed and controlled by a human operator. Movements, facial expressions, lighting, and visual details are explicitly programmed using animation tools, for example (3D). [45] The computer merely executes predefined instructions within a structured pipeline. [46] On the contrary, deepfakes operate through inference. The system is trained on large datasets and predicts how a target face should appear based on a source video. Humans provide data rather than precise commands. The algorithm autonomously interprets the input and generates the final output.[47]

This difference is reinforced by production structure. CGI follows a sequential post-production pipeline, including modeling, texturing, animation and compositing. Each stage requires human expertise and validation.[48] On the other hand, deepfakes bypass this pipeline through end-to-end machine learning. Once trained, the model can synthesize realistic faces without manual intervention at each stage.

b. Deepfakes and cheap fakes

Another technology closely related to deepfakes is cheapfakes. The term “cheapfake” emerged around 2019 particularly in political context such as the 2019 incident involving a slowed-down video of the US House Speaker Nancy Pelosi.[49] Both deepfakes and cheapfakes have common features. In fact, they are considered sub-types of audiovisual manipulation, and

---

[42] Bob Rehak, ‘Computer-Generated Imagery’ (2011) Oxford Bibliographies: Cinema and Media Studies <https://doi.org/10.1093/obo/9780199791286-0068>

[43] Lev Manovich, The Language of New Media (MIT Press 2020); Mattias Rini, ‘Deepfakes: The Evolution of Synthetic Media’ (2020) Communications of the ACM 63(1) 86

[44]Md Shohel Rana, 'CGI vs AI in 2026. Which Visual Workflow Should You Use?' (7CGI, 8 November 2025) <https://7cgi.com/blog/cgi-vs-ai-key-differences>

[45] Abhishek R Pandit and TV Kirdat, 'Impact of AI in the Animation Industry' (2024) 12(3) International Journal for Research in Applied Science & Engineering Technology 2828, 2829.

[46] Mike Bailey and Steve Cunningham, 'Introduction to Computer Graphics' (SIGGRAPH Asia Conference, September 2010).

[47] Ruben Tolosana and others, 'DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection' (2020) arXiv:2001.00179.

[48]Ankit Jain and Nishita Jain, 'A Peek into Computer Generated Imagery Pipeline' (2017) 10(2) Bionano Frontier 220, 221.

[49] Clifford Lampe, '"Cheap fake" video making the rounds today likely won't be the last' (University of Michigan School of Information) <https://www.si.umich.edu/about-umsi/news/cheap-fake-video-making-rounds-today-likely-wont-be-last>

both can be used as tools to spread misinformation.[50] However, there is a clear distinction between the two. Deepfakes are synthetic videos created using AI and GANs to make people appear to say things they never said. Meanwhile cheapfakes are "lower-tech" manipulations where authentic, real footage is re-contextualized, cropped, or edited to deliberately change its meaning.[51]

In practice, the difference is straightforward. Deepfakes use AI to create fake videos. The person using AI needs little skill because the AI does all the hard work and handles the technical details. Cheapfakes create the same kind of fake content, but without AI. Instead, humans make them using simple editing tools. The user opens a photo, moves the mouse, clicks around, and changes how the person looks.[52]

These are not the only differences. Deepfakes are the most computationally reliant and least publicly accessible method, requiring significant expertise and high technical resources. On the contrary, cheapfakes use accessible software like Adobe Premiere, iMovie or even no software at all.[53] Moreover, deepfakes synthesize "virtual performances" from scratch, creating lip-synched speech and facial movements using data. Cheapfakes often rely on contextual manipulation, such as using lookalike stand-ins or relabeling footage from one event like another.[54]

A study conducted by Michael Hameleers from the University of Amsterdam examined the credibility of manipulated political content. The researcher carried out two separate experiments involving participants exposed to manipulated political messages. The scenario used a fictional speech about immigration. It was attributed to a conservative Dutch politician, and the message was made more extreme through manipulation. The main result of the study is that more advanced technology does not mean higher credibility. In fact, deepfakes were perceived as less believable than cheapfakes, even when they conveyed the same message (on a 7-point scale, cheapfakes received an average credibility score of 3.94, while deepfakes

[50] Britt Paris and Joan Donovan, 'Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence' (Data & Society Research Institute, September 2019) 3, 6.
[51] Michael Hameleers, 'Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting' (2024) 36 International Journal of Public Opinion Research 1.
[52] Lance Eliot, 'Explaining Deepfakes Versus Cheap Fakes and the Role of Generative AI' (Forbes, 25 June 2024) <https://www.forbes.com/sites/lanceeliot/2024/06/25/cheap-fakes-and-rescuing-humankind-via-generative-ai/>
[53] Britt Paris and Joan Donovan, 'Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence' (Data & Society Research Institute, September 2019) 4, 12, 14, 16.
[54] ibid

scored only 3.40) Therefore, in some cases, simple and low-tech manipulations can be just as credible and even more credible, than advanced AI-generated content.[55]

## III. AI implications

Since deepfake technology is an Artificial Intelligence innovation,[56] it severely impacts its preceding technologies that do not use any AI to function. There are many non-AI areas and tools which are affected by the prominent use of deepfakes in recent times. In this section, we will discuss some of these areas such as cybersecurity systems, legal and forensic analysis tools and media production and broadcast systems.

### 1. Cybersecurity Systems

The use of technologies, procedures and policies to defend systems, networks, software, devices, and data against cyberattacks is known as cyber security.[57] A cyber security system is therefore a combination of tools, guidelines, and procedures used to defend computers, networks and data from online threats. It seeks to ensure that information is accurate, private, and accessible only to those who need it.[58]

In relation to deepfakes, they can undermine biometric-based authentication systems, particularly facial recognition, voice identification as well as video-based identity verification. In automated verification systems or video calls, fraudsters might pose as actual persons, as well as phone-based security checks can be tricked by voice deepfakes. This technology has an influence on cybersecurity systems because it makes it possible for extremely realistic social engineering assaults, like voice and video impersonation in phishing and fraud, which circumvent conventional authentication and verification techniques and result in data breaches and multimillion-dollar losses.[59] The sophisticated modifications made possible by deepfakes frequently outperform conventional techniques like manual inspection or metadata analysis. Therefore, different strategies for detection and mitigation are needed to overcome the problems caused by deepfakes. AI and ML algorithms are used in detection tactics to find irregularities or contradictions in artificial media. Deepfake detection can be divided into two categories: deep learning approaches, which focus on learnt features and conventional methods,

---

[55] Michael Hameleers, 'Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting' (2024) 36 International Journal of Public Opinion Research 7.
[56] GAO, 'Science & Tech Spotlight: Deepfakes' (2020); available at <https://www.gao.gov/assets/gao-20-379sp.pdf.>
[57] GRC Solutions, 'What is Cybersecurity: Definition and Best Practices' (2025); available at <http://grcsolutions.io/what-is-cybersecurity/>.
[58] G. Vivek, 'What is Cyber security: A Complete Beginner's Guide' (last updated 2025); available at <https://www.simplilearn.com/tutorials/cyber-security-tutorial/what-is-cyber-security>.
[59] V. Soni, Infosys, 'Deepfake and Its Impact on Cybersecurity - A New Frontier to Address' (2025). https://www.infosys.com/services/cyber-security/documents/deepfake-impact-cybersecurity.pdf

which rely on constructed features,[60] which are not covered by this essay. In order to solve the issues around deepfakes and prevent their harmful usage, continued study and cooperation will be essential as the technology develops.[61]

2. Legal Forensic Analysis Tools

When it comes to handling video and image evidence, courts have serious challenges due to the simplicity with which deepfakes can be produced. We can no longer presume that a video or a recording is real when it could be a deepfake.[62] In the context of a courtroom, deepfakes can affect the veracity of the evidence, the credibility of witnesses as well as the integrity of the legal system. This is due to the fact that real evidence can now be accused of being false, necessitating its refutation.[63] Among challenges posed by AI and deepfakes that judges and lawyers must face are witness credibility, higher litigation costs or defamation and damage claims.

Deepfakes can be used to create audio or video footage of people who seem to be making false or damning claims, compromising their credibility. By threatening to expose fictitious but compromising materials, parties may utilize deepfakes to scare witnesses and deter them from testifying.[64] Sociological phenomenon known as the "liar's dividend" make the fear that deepfakes would sway juries'/judge's opinions more plausible. This strategy was already used in the Elon Musk and Tesla lawsuit, where the plaintiff submitted a request for Tesla to admit the "authenticity of a video in which Elon Musk made statements about the safety of its Autopilot feature".[65] In response, Tesla claimed that because Musk is a well-known figure, he is frequently the target of deepfake video attempts, and as a result, it is unable to confirm or refute the video's veracity, nevertheless the court rejected its argument since otherwise that could mean Musk, and others in his position could make real, public statements and then claim that their remarks were possibly deepfakes to avoid accountability. Nevertheless,

[60] Broklyn P., Egon A. and Shad R., 'Deepfakes and Cybersecurity: Detection and Mitigation' (July 03, 2024). Available at <https://ssrn.com/abstract=4904874>.
[61] Ibid.
[62] Bellas G., 'Deepfakes in the Courtroom: Problems and Solutions' (March 2025); available at <https://www.isba.org/sections/bench/newsletter/2025/03/deepfakesinthecourtroomproblemsandsolutions>.
[63] Ibid.
[64] Ibid.
[65] Mitchell A., 'Deepfaked Evidence: What Case Law Tells Us About How the Rules of Authenticity Needs to Change', Berkeley Tech. L.J.: Blog (June 23, 2025), <https://btlj.org/2025/06/deepfaked-evidence-what-case-law-tells-us-about-how-the-rules-of-authenticity-needs-to-change/>.

as the public grows more conscious of the dangers deepfakes represent, this strategy gains credibility due to liar's dividend phenomena.[66]

Deepfakes may also cause additional litigation costs, as to detect and disprove deepfakes, litigants could need to employ digital forensics specialists, which would greatly raise the expense of litigation.

3. Media Production and Broadcast Systems

Deepfakes also affect broadcast credibility and media production systems. This example is more ethical rather than technological, however, essential since this is one of the mainly affected environments.

By undermining audience confidence in the authenticity of content, making verification processes more difficult, and raising the possibility of false information spreading through altered footage that imitates actual broadcasts or productions, deepfakes have an impact on media production and broadcast systems.[67] The stakes are very high for broadcasters. The incentive to publish quickly, without verification, is always present in a world where speed is more important than ever. However, a single mistake, such as broadcasting a convincing deepfake, might drastically undermine public confidence and credibility. Additionally, if broadcasters unintentionally distribute modified content, they may be subject to legal action, negative business reactions, and a decline in trust from audiences, right holders, and regulatory agencies.[68]

Therefore, as presented, deepfakes have various AI implications on non-AI technologies, and thus deepfake detection and different mechanisms are necessary in order to manage the new reality with the prevalent presence of deepfakes in more areas of our lives.

## IV. Societal, ethical and legal implications

In the following section, we will explore the advantages and risks of deepfake technology as well as its ethical, societal and legal implications.

1. Real, proposed or potential benefits of deepfake technology

For a comprehensive examination of deepfake technology, it is important to consider the benefits of it. First, it should be clarified that this doesn't refer to the benefits of deepfakes

[66] Livingston C., 'Deepfakes in the Courtroom: Challenges in Authenticating Evidence and Jury Evaluation', University of Baltimore Law Review (2025); available at <https://ubaltlawreview.com/2025/12/01/deepfakes-in-the-courtroom-challenges-in-authenticating-evidence-and-jury-evaluation/#_ftn30>.

[67] Stagg M., 'Why deepfakes are the next big threat to broadcast credibility' (2025); available at <https://www.broadcastnow.co.uk/broadcasting/why-deepfakes-are-the-next-big-threat-to-broadcast-credibility/5204196.article#:~:text=Examples%20could%20include%20a%20fake,still%20plays%20a%20critical%20role>.

[68] Ibid.

themselves, but rather the benefits of using the underlying technology. So, we are talking more about synthetic content in general, i.e. media that has been generated at least in part by AI.[69] The word "deepfake" is regularly associated with the underlying intention of spreading false information.[70]

a) Creative industries, entertainment

One of the greatest actual benefits of this technology is evident in the creative and entertainment industries. Synthetic content has been used in film production and advertising for years. Visual effects (VFX), which is imagery created, manipulated, or enhanced for any movie that doesn't take place during live-action shooting[71], are becoming cheaper and more realistic thanks to deepfake technologies[72]. This is because faces and movements can be generated or adjusted automatically[73].

In this context, deepfake technology also enables historical reconstructions, as demonstrated in the project "In Event of Moon Disaster", in which synthetic face and voice manipulation was used to create an alternative speech by Richard Nixon[74]. The technology also allows the reanimation of actors who have already passed away or the digital rejuvenation of characters, such as the reconstruction of the young Luke Skywalker in "The Mandalorian"[75]. Another advantage is the possibility of linguistic adaptation of scenes, where deepfake methods make it possible to adapt lip movements to new synchronized versions[76] . A well-known example is the "Malaria Must Die" campaign in which David Beckham appears to speak about the disease in nine different languages with the help of AI to raise awareness[77].

b) Assistive technologies, communication and medicine

Another potential use case for deepfake technology is in therapy and care. In Alzheimer's and dementia therapy, visual representations of relatives from earlier stages of life could be generated to facilitate communication with patients and inspire trust.[78] There are also studies on deepfakes in psychological counseling, which could be used here, among other

[69] O'Brien S., 'Positive Use Cases of "Deepfakes"' (2025).
[70] Ibid.
[71] Maio A., 'What is VFX? Defining the Term and Creating Impossible Worlds' (2025).
[72] Birchwood University, 'Exploring the Bright Side of Deepfake Technology' (2024).
[73] Kalmykov M., 'Positive Applications for Deepfake Technology' (2023).
[74] Lees D., 'Deepfakes in documentary film production: images of deception in the representation of the real' (Studies in Documentary Film, 108-129, 2023).
[75] Kalmykov M., 'Positive Applications for Deepfake Technology' (2023).
[76] Ibid.
[77] Ibid.
[78] Ibid.

things, for grief counseling to reconstruct loved ones.[79] However, these ideas are not viewed exclusively as positive, but can also be perceived as disturbing.[80] Potential could also be seen in areas of doctor-patient communication, for example if deepfake technology was used to create avatars that explain complex diagnoses to patients on an individual basis.[81]

2. Proven or potential issues and risks of deepfake technology

Despite the (mostly potential) benefits listed above, the real and potential risks and problems associated with deepfake technology are a regular source of debate. Below are three selected applications that highlight the risks of deepfakes.

a) Non-consensual sexual deepfakes and CSAM

Non-consensual sexualized deepfakes and synthetic child sexual abuse material (CSAM) are particularly sensitive areas of risk. Sexualized deepfake abuse is "an expanded, but distinct form of image-based sexual abuse involving the non-consensual creation, distribution, or threat of creation/distribution of an image or video that has been altered in a nude or sexualized way using AI technologies."[82]

The spread of image and video-based sexual abuse is not a new phenomenon, but the development of AI tools and deepfake technology has opened up a new level[83]. Freely available tools have made it easier to produce such content, enabling even people without technical knowledge to create sexual deepfakes[84]. The rapid spread of these kinds of deepfakes is highly problematic, making it difficult and sometimes even impossible to delete completely [85] . Recently, there have been countless sexualized deepfakes depicting not only celebrities but also children[86]. The content is shared with friends "for fun" or created out of revenge, for example to humiliate ex-partners.[87] Sharing and distributing content on the internet is quick and often perceived as "funny" and "cool" by people in the user's circle.[88] Among other things,

[79] Hoek S. et al., 'Promising for patients or deeply disturbing? The ethical and legal aspects of deepfake therapy' (2024).
[80] Ibid.
[81] Martinez O. N. et al., 'Possible Health Benefits and Risks of DeepFake Videos: A Qualitative Study in Nursing Students' (2024).
[82] Flynn A. et al., Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging form of Image-Based Sexual Abuse (2022).
[83] Kira B., When non-consensual intimate deepfakes go viral: The insufficiency of the UK Online Safety Act (2024).
[84] Pawelec M., Łabuz M. 'Non-Consensual Sexualising Deepfakes - Threats and Recommendations for Legal and Societal Action' (2025).
[85] Ibid.
[86] Negreiro M., 'Children and Deepfakes' (2025); Pawelec M., Łabuz M. 'Non-Consensual Sexualising Deepfakes - Threats and Recommendations for Legal and Societal Action' (2025).
[87] Flynn A. et. al., 'Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of Non-Consensually Created and Shared Sexualized Deepfake Imagery' (2025).
[88] Ibid.

these processes contribute to the further normalization of sexual violence against women and children.[89]

The development of synthetic CSAM is particularly serious in this context. Even though the content is altered or created using AI, it still endangers real children.[90] Such fantasies are normalized and reinforced by the simplicity of creating and disturbing the material.[91] Overall, deepfakes have a particularly negative impact in the area of image and video-based sexual abuse, endangering not only real individuals but also society's perception of sexual abuse.

b) Disinformation in politics and elections

In a political context, deepfakes are mostly used to deceive voters, discredit political opponents or increase social tensions. Deepfakes introduce a new dimension to the spread of disinformation. People tend to trust videos that they have seen themselves more than pure text.[92] This makes deepfakes more effective at spreading false information than pure text-based fakes.[93] In politics, and especially in the context of elections, the risks posed by deepfakes are particularly evident. A key problem is that deepfakes are primarily used during election campaigns, for example through fake speeches, interviews or made-up scandals.[94] Last-minute deepfakes are particularly risky in this regard. They are published immediately before the election and often cannot be proven false in time.[95]

The increasing technical quality of deepfakes also makes targeted political influence possible. One example is microtargeting, which allows content to be tailored to the language, identity or ethnicity of specific population groups.[96] Such targeted, personalized deepfake messages can be used to gain the support of specific target groups.[97] Deepfakes are now almost impossible for users to reliably detect. This increases the risk that manipulated content will influence political discourse and change public opinion without being noticed.[98] Deepfakes are

[89] Ibid.
[90] Negreiro M., 'Children and Deepfakes' (2025).
[91] Pawelec M., Łabuz M. 'Non-Consensual Sexualising Deepfakes - Threats and Recommendations for Legal and Societal Action' (2025).
[92] Dan V., 'Deepfakes as a Democratic Threat' (2025).
[93] Dan V., 'Deepfakes as a Democratic Threat' (2025).
[94] Insikt Group, 'Targets, Objectives, and Emerging Tactics of Political Deepfakes' (2024).
[95] Carmen Constantineanu, 'Deepfakes: The New Frontier in Political Disinformation' (2024).
[96] Wack M., Parry D. A., 'Synthetic Diversity: Examining the Effects of Ethnic Targeting Using AI-Generated Political Ads', (2025).
[97] Ibid.
[98] Carmen Constantineanu, 'Deepfakes: The New Frontier in Political Disinformation' (2024).

also used to generate artificial political approval in order to make political movements appear larger or to fake the support of prominent figures.[99]

As manipulated videos become increasingly convincing, there is a growing danger that politicians will dismiss genuine videos as fake ("liar's dividend") and that there will be widespread uncertainty about which information is still credible.[100] In the long term, this can weaken trust in the media, democratic institutions and political processes in general.[101] This effect is again intensified by the fact that advances in technology make it increasingly easy and inexpensive to create fakes, making political manipulation more accessible. Deepfakes are also used in military conflicts, as a tool for spreading misinformation in order to cause confusion, provoke panic or weaken the morale of soldiers and civilians, such as the deepfake video of the Ukrainian president allegedly ordering his troops to surrender.[102]

Overall, these developments illustrate that deepfakes go beyond isolated manipulations and come with the risk of fundamentally changing political decision-making processes.

3. Analysis of benefits and risks

The positive use cases mentioned above reflect the real problem with deepfake technology. The same technology is employed, but with malicious and dangerous underlying motives. What can be beneficial in movies, can quickly become a weapon when used with the wrong intentions. Not only can it be used against individuals in need of protection but recently it has increasingly been used to destabilize democracy and peace. These real-world implications make deepfakes a very dangerous technology, and the benefits simply do not outweigh these enormous risks. Moreover, consider that many of the advantages are not being exploited yet but mainly describe potential applications.

4. Societal implications

a) Distrust in media and digital evidence

Deepfakes contribute to a decline in trust in political institutions, public media and democratic processes. One of the main reasons for this is that it is becoming increasingly difficult for the public to distinguish between what is real and what has been manipulated.[103]

---

[99] Insikt Group, 'Targets, Objectives, and Emerging Tactics of Political Deepfakes' (2024); Robins-Early N., 'How did Donald Trump end up posting Taylor Swift deepfakes?' (2024).
[100] Naffi N., 'Deepfakes and the crisis of knowing' (2025).
[101] Wazid M., 'A secure Deepfake Mitigation Framework: Architecture, Issues, Challenges, and Societal Impact' (2024).
[102] Carmen Constantineanu, 'Deepfakes: The New Frontier in Political Disinformation' (2024); Kuźnicka-Błaszkowska D., Kostyuk N., Emerging need to regulate deepfakes in international law: the Russo–Ukrainian war as an example (2025).
[103] Carmen Constantineanu, 'Deepfakes: The New Frontier in Political Disinformation' (2024).

This promotes a fundamental mistrust of digital evidence, particularly in journalism and election campaigns.[104] This situation is worsened by the phenomenon of the "liar's dividend" explained above, which means that even authentic content is sometimes dismissed as a deepfake.

b) Gendered impacts and inequality

Sexualized deepfakes predominantly affect women, especially those in the public eye.[105] Such deepfakes reinforce and perpetuate structural misogyny, while the damage caused to those affected is systematically downplayed.[106] Deepfakes are often used against women in politics and journalism to undermine their credibility and intimidate them[107]. Above all, this makes it more difficult for them to participate in public life or, in some cases, prevent them from engaging in public activities altogether.[108]

Marginalized groups, such as LGBTQIA+ individuals and ethnic minorities, are disproportionately at risk. [109] They are already frequently targeted by digital harassment, regardless of deepfakes, and this could be further escalated by deepfakes.[110] One reason for this is that social prejudices are reproduced in technology and AI only reinforces stereotypical representations.[111] For the same reason, people with disabilities are also more affected by digital sexual assault.[112] In summary, it can be seen that deepfakes help intensifying social inequalities and thus contribute to reinforcing existing power structures such as misogyny.

5. Legal aspects

a) Relevant regulations

In this paper we have decided to focus on the European Union legal framework and how it addresses the use of deepfakes.

[104] Negreiro M., 'Children and Deepfakes' (2025); Insikt Group, 'Targets, Objectives, and Emerging Tactics of Political Deepfakes' (2024).
[105] Insikt Group, 'Targets, Objectives, and Emerging Tactics of Political Deepfakes' (2024).
[106] Asher Flynn et. al., 'Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of Non-Consensually Created and Shared Sexualized Deepfake Imagery' (2025); Pawelec M., Łabuz M. 'Non-Consensual Sexualising Deepfakes - Threats and Recommendations for Legal and Societal Action' (2025).
[107] Insikt Group, 'Targets, Objectives, and Emerging Tactics of Political Deepfakes' (2024).
[108] Ibid.
[109] Asher Flynn et. al., 'Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of Non-Consensually Created and Shared Sexualized Deepfake Imagery' (2025).
[110] Ibid.
[111] Ibid.
[112] Ibid.

*i) AI Act*

One of the most extensive initiatives to regulate AI technology, including deepfakes, is the EU AI Act. Its main purpose is to strike a balance between the need to uphold fundamental rights and social values and the development of innovation.[113] The Act includes a definition of a deepfake and the particular rights and obligations associated with them, such as fundamental transparency and disclosure guidelines. It also places deepfakes in the "specific" or "limited risk" quasi-category of AI systems.[114]

The AI Act defines them as an "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".[115] Furthermore, Article 50(4) and Recital 134 talk about the transparency obligation, according to which AI system deployers are required to reveal whether or not the content is artificially created or altered. Given the potential harmful nature of deepfakes, this notification requirement is somewhat lenient.[116] Article 50 recognizes that even though some AI systems are not deemed high-risk, they still carry the risk of impersonation or deception because they are made to interact with people or produce content. In order to address this, the Act establishes transparency requirements without changing current high-risk AI regulations. Primarily, people must be informed when interacting with AI unless it is evident from the context. Furthermore, people must be informed when AI systems use their biometric data to identify emotions and intentions or classify them.[117]

The AI Act will go into full effect in 2027,[118] even though some of its clauses have been completely applicable since 2024. Given the speed at which AI is developing and the processing needed to address problems like deepfakes and disinformation, there are concerns about the delay in mandating complete compliance. Nevertheless, enforcing transparency across international platforms is fraught with difficulties.[119]

---

[113] Eitren W., 'Deepfakes in the AI Act' (2024), available at <https://schjodt.com/news/deep-fakes-in-the-ai-act>.
[114] Łabuz M., 'Regulating Deep Fakes in the Artificial Intelligence Act', Volume 2, No. 1 (2023), available at <https://www.acigjournal.com/Regulating-Deep-Fakes-in-the-Artificial-Intelligence-Act,184302,0,2.html>.
[115] Regulation (EU) 2024/1689 (AI Act), Article 3(60).
[116] Eitren W., 'Deepfakes in the AI Act' (2024), available at <https://schjodt.com/news/deep-fakes-in-the-ai-act>.
[117] Fragale M., Grilli V., 'Deepfake, Deep Trouble: The European AI Act and the Fight Against AI-Generated Misinformation'(2024), available at <https://cjel.law.columbia.edu/preliminary-reference/2024/deepfake-deep-trouble-the-european-ai-act-and-the-fight-against-ai-generated-misinformation/>.
[118] European Commission, 'AI Act' (Last update: 5 December 2025); available at <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai#:~:text=Application%20timeline,the%20AI%20Act%20that%20will:>.
[119] Ibid.

*ii) DSA and GDPR*

The DSA establishes rules governing internet services that European citizens utilize on a daily basis.[120] These services include social media platforms, app stores and marketplaces.[121] The Digital Services Act (DSA) establishes two important requirements that relate to deepfake content even though it does not define the term specifically, i.e. a transparency requirement, similarly to the AI Act, according to which artificially created or altered content must be properly identified as such. This includes making sure consumers are aware of the content's false character by using watermarks or other identifiers. Secondly, it establishes a notice-and-action obligation, according to which when platforms receive information regarding harmful or illegal content, including deepfakes, they must take appropriate action. Recital 50 of the DSA states that this includes removing or limiting access to such content following appropriate evaluation.[122]

Personal data processing is nearly always a part of deepfakes. The GDPR nevertheless applies to such processing even if the data is inaccurate.[123] The GDPR's Article 9 limitation on processing certain types of personal data may apply to a deepfake. This occurs when a deepfake employs biometric information, such a person's voice or face, to identify the target. The deepfake may also be considered special category personal data if the altered material refers to a person's race, health, sexual orientation, or political views. Processing of a special category's personal data is, in theory, prohibited.[124] The deepfake creator then needs the data subject's express consent in order to produce and distribute the deepfake.

b) Legal implications

Deepfakes raise legal issues in several overlapping areas such as intellectual property and publicity rights, criminal law, and many more.

A deepfake may violate copyright if it makes unauthorized use of copyrighted content such as music, movie clips, or photographs. Trademark concerns may arise if brands, logos, or distinguishing signs are used. [125] Furthermore, when deepfakes are used for fraud, impersonation, blackmail, harassment, or sexually explicit content, like the aforementioned

[120] European Commission, 'The Digital Services Act' (Last update: 15 December 2025); available at <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act>.
[121] Ibid.
[122] Whisperly, 'Deepfake Regulation: Chaos That Matters Now More Than Ever' (2025); available at <https://whisperly.ai/deepfake-regulation/>.
[123] Ibid.
[124] Regulation (EU) 2016/679 (GDPR), Article 9(1).
[125] Bass D. F., Shlaimon N., Penning N., J., 'Deepfakes and Intellectual Property Legal Issues' (2025); available at https://www.lexisnexis.com/community/insights/legal/practical-guidance-journal/b/pa/posts/deepfakes-and-intellectual-property-legal-issues.

CSAM cases, criminal charges may result.[126] Deepfakes, however, have many more legal implications since they can affect many areas of life and law.

## V. Further discussion

There is even more to discover about deepfakes. The main focus of this work is about generation of deepfake by modifying photographs, videos or voices, but deepfakes can also be created from text. Now, with different generative AI models, a person can just type a sentence and the video, image or audio will be created, offering a very realistic result. To obtain these deepfakes there is another type of technology involved, called diffusion models. Text-to-image or text-to-video with the diffusion models works by turning a text description with optional identity cues into an image or video, through denoising random noise into realistic media[127]. After that, the model works by encoding the input text into an embedding that captures attributes like age, emotion, clothing[128], or a situation as unreal as the king of the Netherlands sitting on an orange moon, or something that could look more realistic and as a result deceive viewers.

Considering these new and easier alternatives to generate fake content, there is a higher risk of fabrication of hyper-realistic audio, video or images that mimic real people. Therefore, it is necessary to develop robust detection methods so trust can be restored. These deepfake detection methods use a mix of AI, digital forensics and infrastructure measures. The main detectors based on AI involve CNNs, vision transformers and others to classify media as real or fake based on different features of the media that humans would miss. Moreover, every person can also try to spot certain inaccuracies by analyzing faces or frames in the pictures, focusing, for example, on unnatural eye-blinking, lip-sync error, skin texture anomalies, etc. Diving deeper there are more advanced techniques regarding metadata forensics that will examine the compression history, editing traces and a file's technical history to verify if it aligns with the claimed origin[129].

## VI. Conclusion

The emergence of deepfakes and the technology involved in its creation represent an interesting shift in the digital landscape. It has moved from media manipulation passing

[126] Finnerty N., 'Tackling the regulation of sexually explicit deepfakes' (2024) <https://www.kingsleynapley.co.uk/insights/blogs/criminal-law-blog/tackling-the-regulation-of-sexually-explicit-deepfakes>.
[127] Yunzhuo Chen and others, 'Text-image guided Diffusion Model for generating Deepfake celebrity interactions' [2023] Cornell University <https://doi.org/10.48550/arXiv.2309.14751>
[128] 'Text-to-image model', Wikipedia (2025) <https://en.wikipedia.org/wiki/Text-to-image_model>
[129] Paravision, 'A Practical Guide to Deepfake Detection' (White Paper) <https://www.paravision.ai/whitepaper-a-practical-guide-to-deepfake-detection/>

through manual deterministic artistry (CGIs) to autonomous, AI driven processes. In that sense, the technical sophistication of Generative Adversarial Networks (GANs) and Autoencoders have simplified the ability to create hyper-realistic synthetic media. This analysis discussed the benefits and disadvantages of deepfakes. While it offers transformative potential for creative industries and accessibility, these benefits can be overshadowed by high-stakes risks. The ease of generating non-consensual sexual content and political disinformation poses a direct threat to individual dignity and democratic stability. Moreover, one of the most insidious impacts identified is that the mere existence of the technology provides a cover for bad actors to dismiss authentic evidence as "fake" thereby destroying public trust in the judiciary, the media and digital reality itself.

For that reason, the European Union's approach with the AI Act, DSA and GDPR is one of the most innovative attempts to enforce transparency and accountability. However, the quick evolution of diffusion models and text-to-video synthesis suggests that legislation may struggle to keep pace with the decreasing cost and increasing accessibility of these tools. Therefore, a groundbreaking regulation is needed, one that incorporates a comprehension of technology and its implications. Ultimately, the challenge of deepfakes cannot be solved by law or technology in isolation. It requires a multidisciplinary strategy including a societal push towards digital literacy. As we move forward into a world overflowed by synthetic media, the ability to verify "what is real" will become as valuable as the ability to create it.