Deepfake: A Multifaceted Dilemma in Ethics and Law

Article *in* Journal of Information Ethics · December 2023

DOI: 10.2307/JIE.32.2.109

CITATION
1

READS
1,194

2 authors:

Gergely Gosztonyi
Eötvös Loránd University
81 PUBLICATIONS **104** CITATIONS
SEE PROFILE

Gergely Ferenc Lendvai
University of Richmond
35 PUBLICATIONS **23** CITATIONS
SEE PROFILE

Deepfake

A Multifaceted Dilemma in Ethics and Law

Gergely Gosztonyi *and* Gergely Ferenc Lendvai

Abstract

The present paper explores the relationship between deepfake and fake news through the tools of the law. The paper first introduces the conceptual basis, then presents the relationship between disinformation and deepfake, the relevant U.S., European and alternative regulations in the context of unlawful deepfake content, and possible solutions. Particular attention is paid to the legal perception of disinformation in the context of deepfake technology, highlighting the harmful social and legal processes involved. The structure of the study is based on a review of national and international regulations and relevant literature and the authors' proposed solutions to the controversies caused by deepfake disinformation.

1. Introduction

Hungary's Prime Minister Viktor Orbán is notoriously averse to migration to Europe and wants to preserve the continent's "Christian character" (Rankin 2018). Despite all this, would it be possible to find a video where he says that he is a big fan of migration and would like to tear down all the borders and fences around Hungary to welcome anyone who asks for help because they have had to leave their homeland due to severe crises? This now seems impossible. But is it impossible? All it would take is a free program, as many photos as possible, a serious computer, and some time to make it happen. Will deepfakes change the way we live online? Will it change what we see and believe with our own eyes? The study examines how the law can fight against AI-generated reality and disinformation.

Journal of Information Ethics / Volume 32, Number 2 / Fall 2023 / pp. 109–121 /
ISSN 1061-9321 / eISBN / DOI:10.2307/JIE.32.2.109 /

As the press and media have evolved in recent years, and social media have become an integral part of our daily lives, we increasingly lose trust in edited reality.[1] People are less and less trusting of the press, as articles can be edited online in an instant, and the reality of fake news is with us—even if most politicians use the phrase (inappropriately and damagingly) as "something I disagree with" (CNBC 2017; Index 2018). Several studies have listed events that have eroded public trust in the media industry and journalism. According to Farnaz Fassihi of the *Wall Street Journal*, "the rise of social media has made it harder for the average person to distinguish between verified facts and misinformation" (Dickinson 2018).

And this is not only the case in politics. We've seen disturbing videos of well-known actresses in obvious sexual situations that almost certainly didn't happen. It turns out that we are facing something that could change our lives forever: using artificial intelligence, it is easy to make fake porn footage of anyone.[2] And sexual content is only the first step; the next could be politics, then the world of ordinary people. If we lose trust in written articles, will we also lose trust in pictures and videos? If we can no longer believe our eyes, the world will be different.

2. Conceptual Framework: What Is Deepfake?

The term deepfake is a combination of the words deep learning and fake (Mráz 2021: 249); the first element of the term refers to deep learning, while the second element refers to the fake nature of the content produced (G. Karácsony 2022: 306–307; Citron-Chesney 2019). These are videos where someone's head is mounted on another person's body, sometimes for the purpose of making memes or jokes but sometimes with malicious intent. The story began when facial expressions started to spread on social media in 2016. Although the technology dates back to 2013, it took three years for apps such as MSQRD, Face Swap Live, Snapchat, or Face Stealer to become available to a broader audience (Dredge 2016). For a while, everyone's wall on social media was full of such images. Your head is on a doll's body, a cat's, or your lover's. Sometimes funny, but mostly with disgusting results. It was nothing more than quickly fading curiosity. And as soon as it came, it disappeared. But in late 2017, a much more alarming thing happened: on Motherboard, a user called DeepFakeApp allowed someone's head to be drawn onto a moving image, for instance, a full video (Vincent 2017). From there, technology and human creativity began to advance at lightning speed.

The definition of deepfake is very problematic (Ambrus 2021: 221), primarily if we use a specific set of terminology of law. Typically, deepfake refers to a technology that uses machine learning (Cover 2022: 609), artificial intelligence or artificial learning (G. Christmas 2022: 301), or a technological solution that

uses algorithms (Mezei–Szentgáli-Tóth 2022: 248), which the user can intentionally (Fernandez 2022; Paris-Donovan 2019; Ferraro 2019: 1–3) fake (Mráz 2021: 249), in some way copy or transform (Lyu 2020: 1) image and sound content (Van der Sloot–Wagensveld 2020: 1; Veszelszki 2022: 33), by creating a fake new content that convincingly (Gibson 2020: 260) makes it appear that the new content and its actor(s) are natural, realistic or original (Judge-Korhani 2021).[3]

There are numerous deepfake creation possibilities in the technology field, the most well-known being the "GAN," or Generative Adversarial Networks technology (Van der Sloot–Wagensveld 2020: 2; Pantserev 2020: 40–43). This technology covers a neural network, a generator, that creates new fake content by downloading data sets and data content. This means that by inputting specific images and audio material (Reid 2021: 209), deepfake technology can map the material we provide onto another, different content, i.e., "based on available images or audio recordings of a person, it creates a manipulated recording that depicts the real person in a fictitious situation in which the specific person was not placed, or attributes to him a statement that he did not make" (G. Karácsony 2022: 306).

Deepfake content is most often created of people, especially people's faces and/or voices (Lyu 2020: 1). Digital copying and transformation of the face are salient and significant in the context of freedom of expression and law for two reasons. First, the face is one of the most important physical characteristics of the individual, a person, an evolutionary feature designed for identification purposes. Secondly, facial images and sound recordings as media are particularly significant in relation to public trust, public discourse, and content credibility (Mezei–Szentgáli-Tóth 2022: 248). The latter is also crucial for freedom of expression and democratic publicity; the credibility of the face and the communication, as well as trust in the veracity of the communication, is the basis for the formation of socio-political relations (see public political discourse), the democratic rule of law, and the exercise of freedom of expression and opinion (Papp 2022: 147).

The most relevant cornerstone of the definition of deepfake from the point of view of pseudo-propagation is that a deepfake is itself a communication (Mráz 2021: 252–261). The deepfake-maker thus communicates, exercising his right to freedom of expression when he publishes humorous deepfake videos of his friends (Van der Sloot–Wagensveld 2022: 3), but he also communicates when he manipulates false political messages with deepfakes (Mráz 2021: 252–253). It is therefore worth underlining that, in addition to criminal law (cf. Ambrus 2021; Miskolczi-Szathmáry 2018: 140–141; Pantserev 2020: 50; Sorbán 2020: 85–100), data protection and civil law (private law) (Ebermann 2021: 39; Hine-Floridi 2022: 608–609) approaches, the fundamental rights aspect of deepfake is inescapable for understanding the regulatory directions.

3. Deepfake and Disinformation

3.1. Transparent Hack or a Heavy Weapon?

As Rob Cover puts it (2022: 615), deepfake is primarily a "social concern." This terminus technicus is apt for the phenomenon, as the use of deepfake can have adverse effects in many areas, such as its impact on public discourse and opinion, personal rights, and privacy (Van der Sloot–Wagensveld 2022: 10), on political deception or even its use in armed conflicts (Allen 2022). In the spread of misinformation, deepfake is, as Veszelszki puts it, "a real heavy weapon" (Veszelszki 2022: 33). Political disinformation and deception are highly relevant to deepfake (Ürmösné Simon–Nyitrai, 2021: 88): it is worth recalling here the "speech" of Ukrainian President Volodymyr Oleksandrovich Zelensky in March 2022, where he asks Ukrainian soldiers to stop fighting and return home in a deepfake recording (Allyn 2022) or the speech of "stunned Nancy Pelosi," the Democratic Speaker of the U.S. House of Representatives (Buo 2020: 2–3).

One of the biggest challenges in the fight against fake news is that the fake news generated by deepfake is most often not spread by the users who spread it with the intention of disinformation but by those who are deceived by the manipulated content (Veszelszki 2021: 94). Stopping the transmission of political fake news via deepfake seems to be impossible for the time being, and the harmful use of the technology is a significant threat to the quality of public discourse and the trust in the media (Citron-Chesney 2019, Graber-Mitchell 2021), and some experts even argue that it could be dangerous for the judiciary (Hasen 2019). A fascinating train of thought on the relationship between deepfake and character assassination is outlined by Péter Bajomi-Lázár (2022: 8). In a survey by Vaccari and Chadwick (2020: 6), half of the viewers of a 4-second video were misled by deepfake technology, and 35 percent of all viewers were unsure whether they had seen deepfake content in the video. This ratio perfectly illustrates that if half of the viewers were deceived by deepfake technology in 2020, then as time goes by and technology evolves, this ratio may become increasingly worrying—especially in the context of the fight against fake news (Judge-Korhani 2021: 16).

3.2 Regulatory Solutions—Examples of Initiatives in the U.S. and Europe

3.2.1. American solutions

In the context of U.S. deepfake regulations, it is important to briefly discuss the different and often divergent practices of various states. In the state of California, action against deepfakes was primarily motivated by political considerations; under the state's election law, a candidate seeking political office may, within 60 days of an election, take action against a person who produced deepfake content and who sought to "discredit" the candidate with the content

during the campaign (Van der Sloot–Wagensveld 2022: 11; Californian Elections Code: §20010). The Californian concept seeks to address relevant problems, but its effectiveness is highly questionable; investigations are greatly hampered by the need to identify anonymous deepfake producers, which ultimately undermines the effectiveness of the provision (Cover 2022: 616). The Californian legislation was modeled on the Texas legislation, which was the first to prohibit deepfake videos that harm or disadvantage political candidates (Ferraro 2019: 13).

The states of New York and Nebraska have moved toward direct deepfake bans (Gibson 2020: 269, 282–283; Ferraro 2019: 12). The two states would protect the public and penalize the creation and publication of content "created with bad intent" (S. 3805, 115th Cong. 2018: §1041). The problem, however, is the same as with the California legislation; with mild skepticism, it is feared that these provisions will remain merely symbolic because their enforcement (1) is not transparent and predictable, and (2) there is no reliable, mature technological or legal means to track down deepfake producers.

In Virginia, we have the Unlawful Dissemination or Sale of Images of Another Person Act (Ferraro 2019: 14; Va. Code Ann. § 18.2-386.2, HB 2678, SB 1736 2019). The law makes the dissemination of non-consensually "falsely made" images and videos a misdemeanor (HB 2678 VA 2019), with possible penalties for this type of offense ranging from a fine to up to 1 year in prison.

Of particular note is the 2019 DEEPFAKES[4] bill in Congress, which, although it failed at the committee level, was a meaningful step against the illegal use of deepfakes. The DEEPFAKES Act is intended to remedy all possible illegal activities that deepfakes may have predicted (Ferraro 2019: 7). The bill also requires that when displaying images and videos manipulated by deepfakes, creators must post a visual notice indicating that the content is manipulated or generated by a deepfake (DEEPFAKES Act 2018: §1041). Failure to do so can result in a severe fine of up to $150,000 (DEEPFAKES Act 2018: §1041(f)(2)).

3.2.2. European and Chinese Proposals

Regarding regulation at the European level, the EU's draft Regulation on Artificial Intelligence (hereafter: AI Regulation) and the Digital Services Act (henceforth: DSA) are worth mentioning. According to the AI Regulation (AI Regulation 2021: Article 52(3); ERPS 2021), deepfake systems are low-risk AI systems. The MI Regulation proposes labeling deepfake content as primary protection against unlawful deepfake content (G. Karácsony 2022: 312). This would impose a duty of disclosure on users of artificial intelligence systems when generating content (images, audio, or video content) manipulated by deepfake technology, to the effect that the content should be labeled as artificially created or manipulated content (MI Regulation 2021: 52). The

MI Regulation has been heavily criticized as it has weakened the possibility of effective and productive enforcement by classifying deepfake technology as low risk (Fernandez 2022).

The DSA, although less closely related to deepfake regulation, is noteworthy for its prominence in the field of platform regulation in Europe. The regulation published in October 2022 (OJL EU L 277: 2022) sets obligations for hosting providers, including platforms, to ensure that the regulation of platforms guarantees complete protection of users and their fundamental rights in the online space (Török 2022: 195–200). The DSA seeks to achieve the above objectives through several innovations and solutions; it extends EU obligations to platforms on an extraterritorial basis and proposes a higher level of action against illegal content through notification and action mechanisms (DSA 2022: Article 16, Gosztonyi 2022: 147). Although there have been numerous criticisms of the broad terminology used in the DSA or the possible bureaucratization of the fight against illegal content (Gosztonyi 2022: 148–150; Peukert 2021), the DSA is undoubtedly a progressive and positive step in the field of platform regulation and thus in the fight against illegal deepfakes.

The Hungarian legislation has not yet developed any deepfake-specific regulations. Nevertheless, the relevant provisions of the Hungarian regulatory environment, particularly the 2012 Criminal Code, such as harassment, defamation, and sexual blackmail (Sorbán 2020: 99), may apply to deepfake images and sound recordings. Kitti Mezei and Boldizsár Szentgáli-Tóth (2022: 248) also raise the issue of deepfake regulation in the context of offenses not strictly linked to the individual; in their view, deepfake content may be effective even in the context of the Covid-19 world pandemic for disinformation and to inhibit defensive action. The new paragraph on the dissemination of rumors (Ambrus (2021) agrees with Sorbán on the need to regulate deepfake; in his opinion, it is not necessary to include deepfake in a different factual situation, as "the category of deepfake does not carry any additional danger to society," while Miskolczi and Szathmáry (2018: 140–141) argue that a higher level of legal protection should be created for deepfake, especially in relation to data protection.

In the context of China's regulation, it is worth briefly mentioning the existence of alternative, less cautious deepfake regulations. On January 28, 2022, the China Cyberspace Administration issued a set of rules on the administration of Internet information services using deepfake synthesis (Creemers-Webster 2022). Deep-synthesis technology is an umbrella term in the regulation (§ 2), which covers virtually all audiovisual content (Hine-Floridi 2022: 608). The regulation aims to take concrete, severe and strict action against deepfakes. It differs from U.S. or European regulations in that the Chinese deepfake regulation uniquely targets deepfake synthesis software and software providers, not platforms or users of deepfake software (Hine-Floridi 2022: 608–609). This new legal thinking has received positive feedback, given that it is expected

to provide more excellent protection for Internet users and stricter, legally ensured accountability for companies developing deep-synthesis (Yan 2022).

4. Better Regulation Than Law? Alternatives and Perspectives Outside the Law

The reality of the problem is best illustrated by the jubilee statements issued in July 2019 by the UN Special Rapporteur on Freedom of Expression, the Organization for Security and Co-operation in Europe (OSCE) Special Rapporteur on Freedom of the Press, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, in which they outlined the challenges facing freedom of expression in the next decade and called on world leaders and online platform owners to adopt human rights-sensitive solutions to the challenges posed by disinformation, including the growing emergence of deepfake (The United Nations 2019).

The first deepfake regulatory instrument was not a law or a case law decision; Reddit decided in 2018 to restrict deepfake content on its website (Van der Sloot–Wegensveld 2022: 3). The regulation was preceded by the proliferation of deepfake pornographic content on the deepfake "subreddit"[5] (Sorbán 2020: 90).

Facebook also announced in 2020 that it would ban sharing deepfake videos created with artificial intelligence algorithms (Bickert 2020). However, Anita Molnár and Árpád Rab (2021) note that "according to a November 2020 Twitter post summarising the election efforts, 300,000 tweets since October 27 have been tagged with a misleading content warning, which was 0.2 percent of all election-related posts in that period. There was no mention of deepfakes."

In addition to deepfake production, deepfake detection software is also actively developing (G. Karácsony 2022: 313–314). Such technical directions include implementing multifactor identification systems or voice and image-based identification. However, circumvention of these security solutions is still not a significant problem for more advanced deepfake software. Another drawback of a regulatory approach that relies solely on technological nuggets is that, although more and more research is being conducted, it is clear that no deepfake detection system with 100 percent security and absolute precision exists and is unlikely to be developed in the near future (Pantserev 2020: 48). The addition of human resources to the technological toolbox, while it may seem a more efficient proposal, would still entail high costs and, again, inadequate protection against deepfake content (Cover 2022: 617). One positive step towards incorporating technological advances into the legal environment would be to

regulate faceswap and similar software and programs in the context of liability and accountability for illegal deepfake content (Gerstner 2020: 12–13).

Van der Sloot and Wagensveld (2022: 12) argue that regulating deepfake with the tools of law is a hazardous area, not from the point of view of lawmaking but rather from the point of view of enforcement. As the authors put it, "deepfake is democratizing"; by this, they mean that it is no longer only the big studios that can afford to "resurrect" celebrities of the past on screen using the technology, but that virtually any user can produce deepfake content, mainly using free apps (Ibid.; Cover 2022: 613).

The authors of this paper propose a triangular regulatory model for deepfakes. The model's skeleton would blend the above regulatory methodologies and perspectives; that is, a harmonized linking of state and international legislation with the platforms' own monitoring activities, using and developing the constantly evolving technological capabilities to detect and identify deepfakes. This three-step model would, of course, face some challenges, such as harmonizing rules, and reconciling legislation with the private regulation of platforms (here, for example, a solution could be for platforms to seek the help of a monitoring committee to develop their rules [Hall 2018: 73]), standardizing conceptual frameworks (for which, for example, Mathilde Pavis has proposed a comprehensive and well-defined system [Pavis 2021: 981–982]), and ensuring that the platforms' rules are consistent with each other, not to mention the consideration of economic interests, while joint, shared action could also provide a comprehensive, holistic solution (Meskys-Kalpokiene-Jurcys 2020: 11–12) to reduce the vulnerability of users and individuals (Cover 2022: 617), to preserve the development of communication channels and to fight against pseudo-news. However, the triple regulation detailed above can only make a real difference if awareness and sensitization of users and individuals are prioritized in all relevant aspects of deepfake technology that affect individuals (Hall 2018: 74; Farish 2020: 48).

5. Conclusion: "It doesn't have to be perfect, just good enough" (Warzel 2018)

Deepfake content production as a tool and method of fake news dissemination is likely to become increasingly popular in the coming years (Allen 2022: 77–78). It will reach a broader and wider audience of both consumers and creators. To borrow from Charlie Warzel (2018), the deepfake problem will be one of the most exciting online regulatory issues of the near future, as technological advances will make it possible to deceive millions of people with imperfect, just good enough fake content.

From the above, it also seems clear that the law as a single, particularistic regulator is insufficient to combat illegal online content. The spectacular

and practical development of deepfake detection systems is welcome, as are international conventions that go beyond national—often divergent—regulations. However, the authors also argue that social awareness is essential. In this context, it is proposed to implement a media awareness process and to train to educate internet users on the dangers of deepfake and the steps they can take to identify deepfake content. This, therefore, calls for a holistic approach in which the actors in the "platform regulation triangle" (Gorwa 2019) jointly take steps to address the problems that arise, for which not only the tools of law but also the development of general media literacy (Nagy 2018) are considered desirable.

Notes

1. On average, only six in ten internet users (63 percent) in twenty countries said they trust the internet. This is an 11 percentage point drop since a similar survey in 2019 (Ipsos 2022).
2. For example, *Wonder Woman* star Gal Gadot's face has been superimposed by AI over a porn star's, so she "stars" in a whole movie. Scarlett Johansson, Maisie Williams, Taylor Swift and Aubrey Plaza have all been victims of the same distortion (Cole 2017).
3. There are already easily accessible, low-cost versions of deepfake videos, which the literature calls cheap fake (Kiss 2021: 27).
4. The name of the law itself is an acronym: "Defending Each and Every Person from False Appearance by Keeping Exploitation Subject to Accountability Act."
5. The subreddit is effectively a "subblog," where specific topics serve as the subgroup's talking and discussion points.

Literature

Allen, Major D. Nicholas (2022). Deepfake Fight: AI-Powered Disinformation and Perfidy Under the Geneva Conventions, *Journal of Emerging Technologies*, 3, 2.

Allyn, Bobby (2022). *Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn*. NPR, March 16 2022, https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

Ambrus, István (2021). *Digitalizáció és büntetőjog (Digitalisation and criminal law)*. Budapest, Wolters Kluwer.

Bajomi-Lázár, Péter (2022). Karaktergyilkosság vagy médiabotrány: mi a különbség? (Character assassination or media scandal: what is the difference?) *Médiakutató*, 23, 2: 7–13.

Bencze, Mátyás, and Győry, Csaba (2021). Hírek szárnyán: a rémhírterjesztés bűncselekménye és a jogbiztonság (On the wings of news: the crime of spreading rumours and legal certainty). *Magyar Tudomány* 182, 5: 614–624.

Bickert, Monika (2020). *Enforcing Against Manipulated Media*. Facebook, January 6, 2020, https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/

Buo, Shadrack Awah (2020). The Emerging Threats of Deepfake Attacks and Countermeasures. *arXiv*, *Cornell University*, December 14, 2020.

Citron, Danielle K., and Chesney, Robert (2019). *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security* Boston: Boston University School of Law.
Dickinson, Daniel (2018). *Interview with Farnaz Fassihi*. May, 1 2018, https://news.un.org/en/audio/2018/05/1008682
CNBC (2017). *Trump to CNN reporter: You are fake news.* January 11, 2017, https://www.cnbc.com/video/2017/01/11/trump-to-cnn-reporter-you-are-fake-news.html
Cole, Samantha (2017). *AI-Assisted Fake Porn Is Here and We're All Fucked*. *Vice*, 11 Dec 2017, https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn
Cover, Rob (2022). Deepfake culture: the emergence of audio-video deception as an object of social anxiety and regulation *Continuum Journal of Media & Cultural Studies* 36, 4: 609–621.
Creemers, Rogier, and Webster, Graham (2022). Translation: Internet Information Service Deep Synthesis Management Provisions (Draft for Comment). *Stanford University–Digichina*, February 4, 2022, https://digichina.stanford.edu/work/translation-internet-information-service-deep-synthesis-management-provisions-draft-for-comment-jan-2022/
Domokos, Andrea (2021). Egyes, a járványhoz kapcsolódó büntetőjogi szabályok különleges jogrend idején (Certain rules of criminal law related to epidemics in a special legal regime). *Glossa Iuridica,* 7, Special Issue: Law and Virus, 73–83.
Dredge, Stuart (2016). *Five of the best face swap apps*. *The Guardian*, March 17, 2016, https://www.theguardian.com/technology/2016/mar/17/five-of-the-best-face-swap-apps.
Ebermann, Axel (2021). The Effects of Deepfakes and Synthetic Media on Communication Professionals, *CUNY Academic Works,* Fall 2021: 1–155.
European Parliamentary Research Service (EPRS) (2021). *Tackling deepfakes in European policy*.
Farish, Kelsey (2020). Do deepfakes pose a golden opportunity? Considering whether English law should adopt California's publicity right in the age of the deepfake. *Journal of Intellectual Property Law & Practice,* 15, 1: 40–48.
Fernandez, Angelica (2022). *Regulating Deep Fakes in the Proposed AI Act*, Medialaws, March 23 2022, https://www.medialaws.eu/regulating-deep-fakes-in-the-proposed-ai-act/
Ferraro, Matthew F. (2019). Deepfake Legislation: A Nationwide Survey-State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media. WilmerHale report: Deepfake Legislation: A Nationwide Survey—State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media, September 25, 2019.
Gergely, G. Karácsony (2022). „Ideje megváltoztatni"—Az arcfelismerő rendszerek alkalmazásának alapjogi kockázatai a közösségi média és a deepfake korában („Time for a change"—The fundamental rights risks of using facial recognition systems in the age of social media and deepfake. In: Török, Bernát–Ződi, Zsolt (eds.) (2022) *Az internetes platformok kora (The Age of Internet Platforms)*. Budapest: Ludovika Egyetemi Kiadó. 299–318.
Gerstner, Erik (2020). Face/Off: "DeepFake" Face Swaps and Privacy Laws. *Defense Counsel Journal,* January 2020. 1–14.
Gibson, Kareem (2020). Deepfakes and Involuntary Pornography: Can Our Current Legal Framework Address This *Technology*? 66. 2020. 259–289.
Gorwa, Robert (2019). The platform governance triangle: conceptualising the informal regulation of online content. *Internet Policy Review*, 8, 2.
Gosztonyi, Gergely (2022). *Cenzúra Arisztotelésztől a Facebookig (Censorship from Aristotle to Facebook)*. Budapest: Gondolat Kiadó.
Graber-Mitchell, Nicolas (2021). Artificial illusions: deepfakes as speech, *Intersect*, 14, 3: 1–19.

Hall, Holly Kathleen (2018). Deepfake Videos: When Seeing Isn't Believing. *Catholic University Journal of Law and Technology* 27, 1: 51–76.
Hasen, R.L. (2019). Deep Fakes, Bots, and Siloed Justices: American Election Law in a "Post-Truth" World. *Saint Louis University Law Journal*, 64, 4: 535–568.
Hine, Emmie, and Floridi, Luciano (2022). New deepfake regulations in China are a tool for social stability, but at what cost? *Nature Machine Intelligence* 4, 2022. 608–610.
Index.hu (2018). *Hungarian PM Calls Country's Leading News Website Fake*. May 30, 2018, https://index.hu/video/2018/05/30/viktor_orban_index_fake_news/
Ipsos (2022). *Internet users' trust in the Internet has dropped significantly since 2019*. November 14, 2022, https://www.ipsos.com/en/trust-in-the-internet-2022
Judge, Elizabeth F., and Korhani, Amir M. (2021). Deepfakes, Counterfeits, and Personality, *Ottawa Faculty of Law Working Paper No. 2021–21.*
Kiss, Máté Attila (2019). Álhírek bűvöletében—A dezinformációk elleni fellépés kilátásai (Under the spell of illusions—The prospects of action against disinformation). *Nemzetbiztonsági Szemle*, 9, 2: 20–28.
Lendvai, Gergely Ferenc (2022). "Of Covid, [say] nothing but the truth": reflections on the consequences on media platforms and freedom of expression principles of the new scaremongering rules implemented in the Hungarian Penal Code during the pandemics. In: Gosztonyi, Gergely–Lazar, Elena (eds.) (2023) *Media Regulation during the Covid-19 Pandemic: A Study from Central and Eastern Europe*.
Lyu, Siwei (2020). Deepfake Detection: current challenges and next steps. *IEEE International Conference on Multimedia & Expo Workshops (ICMEW),* 2020. 1–6.
Meskys, Edvinas, and Kalpokiene, Julija, and Jurcys, Paulius (2020). Regulating deep fakes: legal and ethical considerations *Journal of Intellectual Property Law & Practice*— January 2020. 1–13.
Mezei, Kitti, and Szentgáli-Tóth, Boldizsár (2022). Az online platformok használatában rejlő veszélyek: a dezinformáció és a kibertámadások jogi kockázatai (The dangers of using online platforms: the legal risks of disinformation and cyber-attacks). In Chronowski, Nóra, and Szentgáli-Tóth, and Boldizsár–Szilágyi, Emese (eds.) (2022) *Demokrácia-dilemmák. Alkotmányjogi elemzések a demokráciaelvértelmezéséről az Európai Unióban és Magyarországon (Democracy dilemmas. Constitutional law analyses on the interpretation of democracy in the European Union and Hungary)*. Budapest: ELTE Eötvös Kiadó. 241–262.
Molnár, Anita, and Rab, Árpád (2021). *A deepfake technológia és az amerikai elnökválasztás (Deepfake technology and the U.S. presidential election)*. ITKI Blog, 18 Jan 2021, https://www.ludovika.hu/blogok/itkiblog/2021/01/18/a-deepfake-technologia-es-az-amerikai-elnokvalasztas/
Mráz, Attila (2021). Deepfake, demokrácia, kampány, szólásszabadság (Deepfake, democracy, campaign, freedom of speech). In: Török, Bernát–Ződi, Zsolt (eds.) (2021) *A mesterséges intelligencia szabályozási kihívásai (Regulatory challenges of artificial intelligence)*. Budapest: Ludovika Egyetemi Kiadó. 249–277.
Nagy, Krisztina (2018). *Műveltség—Média—Szabályozás. A médiaműveltség médiapolitikai jelentősége és szabályozási keretei (Literacy—Media—Regulation. The significance of media literacy in media policy and its regulatory framework).* Budapest: Gondolat Kiadó.
Pantserev, Konstantin A. (2020). The Malicious Use of AI-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability. In: Jahankhani, Hamid -Kendzierskyj, Stefan–Chelvachandran, Nishan–Ibarra, Jaime (eds.) (2020) *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*. Cham: Springer Switzerland. 37–56.
Papp, János Tamás (2022). *A közösségi média szabályozása a demokratikus nyilvánosság*

védelmében (Regulation of social media to protect the democratic public). Budapest: Wolters Kluwer.
Paris, Britt, and Donovan, Joan (2019). *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence. Datasociety*, September 18, 2019, https://datasociety.net/library/deepfakes-and-cheap-fakes/
Pavis, Mathilde (2021). Rebalancing our regulatory response to deepfakes with performers' rights, The *International Journal of Research into New Media Technologies* 2021, 27, 4: 974–998.
Peukert, Alexander (2021). *Five Reasons to be Skeptical About the DSA*. *Verfassungsblog*, 31 Aug 2021, https://verfassungsblog.de/power-dsa-dma-04/
Rankin, Jennifer (2018). *Viktor Orbán: re-election of Hungary's anti-immigrant leader is major challenge for EU*. *The Guardian*, April 9, 2018, https://www.theguardian.com/world/2018/apr/09/viktor-orban-re-election-hungarys-anti-immigrant-leader-major-challenge-for-eu
Reid, Shannon (2021). The Deepfake Dilemma: Reconciling Privacy and First Amendment Protections *Journal of Constitutional Law* 23: 209–238.
Sorbán, Kinga (2020). A bosszúpornó és deepfake pornográfia büntetőjogi fenyegetettségének szükségességéről (On the need to criminalise revenge pornography and deepfake pornography). *Belügyi Szemle,* 10: 81–104.
Török, Bernát (2022). A szólásszabadság a közösségi platformokon és a Digital Services Act (Freedom of expression on social platforms and the Digital Services Act). In: Török, Bernát–Ződi, Zsolt (eds.) (2022) *Az internetes platformok kora (The Age of Internet Platforms)*. Budapest: Ludovika Egyetemi Kiadó. 195–208.
The United Nations (UN) (2019). Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information: twentieth anniversary joint declaration: challenges to freedom of expression in the next decade. 2019, https://www.ohchr.org/Documents/Issues/Opinion/JointDeclaration10July2019_English.pdf
Ürmösné Simon, Gabriella, and Nyitrai, Endre (2021). The phenomena of epidemic crime, deepfakes, fake news, and the role of forensic linguistics. *Információs Társadalom* 21, 4: 86–101.
Vaccari, Cristian, and Chadwick, Andrew (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, 6, 1.
Van der Sloot, Bart, and Wagensveld, Yvette (2022). Deepfakes: regulatory challenges for the synthetic society *Computer Law & Security Review* 46: 1–15.
Veszelszki, Ágnes (2021). deepFAKEnews: Az információmanipuláció új módszerei (deepFAKEnews: New methods of information manipulation). In: Balázs, László (ed.) (2021) *Digitális kommunikáció és tudatosság (Digital Communication and Awareness)*. Budapest: Hungarovox Kiadó. 93–105.
Veszelszki, Ágnes (2022). A tudományos influencerektől a deepfake-ig. A legújabb tudománykommunikációs lehetőségek (From scientific influencers to deepfake. The latest science communication opportunities). *Filológia* 13, 1–4: 27–39.
Vincent, James. *AI tools will make it easy to create fake porn of just about anyone The Verge*, December 12, 2017, https://www.theverge.com/2017/12/12/16766596/ai-fake-porn-celebrities-machine-learning
Warzel, Charlie (2018). *Believable: The Terrifying Future of Fake News*. *BuzzFeedNews*,

February 12 2018, https://www.buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news#.ncyONjQOw
Yan, Li (2022). *Using 'deepfake' technology in China may require identity verification in future: cyberspace regulator*. *ECNS*, January 29, 2022, http://www.ecns.cn/news/sci-tech/2022-01-29/detail-ihavfwhh0343801.shtml

---

Dr. habil Gergely Gosztonyi PhD, lawyer, lecturer and senior media researcher, Eötvös Loránd University of Budapest Faculty of Law. Main fields of expertise: censorship, alternative media and the liability of intermediaries. Member of the European Communication Research and Education Association and Community Media Forum Europe, coach of the Hungarian Team for the Monroe E. Price Media Law Moot Court Competition. Hungary-1122 Budapest, Abos utca 13/A. gosztonyi@ajk.elte.hu.

Dr. Gergely Ferenc Lendvai, PhD candidate, lawyer and researcher, Pázmány Péter Catholic University Faculty of Law. Research specializations: platform governance, platform regulation, AI and law, deepfake, digital constitutionalism. Hungary-1015 Budapest, Hattyú utca 17A. gergelyflendvai@gmail.com.