Defending Informational Sovereignty by Detecting Deepfakes: Risks and Opportunities of an AI-Based Detector for Deepfake-Based Disinformation and Illegal Activities

Tahraoui, Milan; Krätzer, Christian; Dittmann, Jana

Erstveröffentlichung / Primary Publication
Konferenzbeitrag / conference paper

Empfohlene Zitierung / Suggested Citation:
Tahraoui, M., Krätzer, C., & Dittmann, J. (2023). Defending Informational Sovereignty by Detecting Deepfakes: Risks and Opportunities of an AI-Based Detector for Deepfake-Based Disinformation and Illegal Activities. In B. Herlo, & D. Irrgang (Eds.), *Proceedings of the Weizenbaum Conference 2022: Practicing Sovereignty - Interventions for Open Digital Futures* (pp. 142-161). Berlin: Weizenbaum Institute for the Networked Society - The German Internet Institute. https://doi.org/10.34669/wi.cp/4.14

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

DEFENDING INFORMATIONAL SOVEREIGNTY BY DETECTING DEEPFAKES?

RISKS AND OPPORTUNITIES OF AN AI-BASED DETECTOR FOR DEEPFAKE-BASED DISINFORMATION AND ILLEGAL ACTIVITIES

Tahraoui, Milan
Berlin School of Economics and Law
Berlin, Germany
milan.tahraoui@hwr-berlin.de

Krätzer, Christian
Magdeburg University
Magdeburg, Germany
kraetzer@iti.cs.uni-magdeburg.de

Dittmann, Jana
Magdeburg University
Magdeburg, Germany
jana.dittmann@iti.cs.uni-magdeburg.de

KEYWORDS

deepfake detection; digital sovereignty; remote ID proofing.

DOI: 10.34669/wi.cp/4.14

ABSTRACT

This paper will first investigate possible contributions that an AI-based detector for deepfakes could make to the challenge of responding to disinformation as a threat to democracy. Second, this paper will also investigate the implications of such a tool—which was developed, among other reasons, for security purposes—for the emerging European discourse on digital sovereignty in a global environment. While disinformation is surely not a new topic, recent technological developments relating to AI-generated deepfakes have increased the manipulative potential of video and audio-based contents spread online, making it a specific but important current challenge in the global and interconnected information context.

1 Introduction

Google has recently forbidden the use of its Colaboratory (Colab) service—one of the most popular platforms online to train machine-learning and AI systems with free computational resources—to generate deepfakes.[32] This is one example among others of the increasing risks perceived to be associated with deepfakes. These risks have motivated public authorities, such as the European Commission,[33] Cyber Administration of China,[34] as well as global leading private firms, such as Google and Meta,[35] to regulate their generation and circulation. One of the most commonly perceived risks with deepfakes, beyond so-called "revenge porn" or harmful application cases, is the anticipated

[32] TechRadar.com, "Google is cracking down hard on deepfakes", 31 May 2022, at https://www.techradar.com/news/google-is-cracking-down-hard-on-deepfakes; reseach.google.com, "Colaboratory: Frequently Asked Questions", at https://research.google.com/colaboratory/faq.html (page visited on 28 August 2022): "We prohibit actions associated with bulk compute, actions that negatively impact others, as well as actions bypassing our policies. The following are disallowed from Colab runtimes: […] creating deepfakes."

[33] Reuters.com, "Exclusive: Google, Facebook, Twitter to tackle deepfakes or risk EU fines", 14 June 2022, at https://www.reuters.com/technolog[33] TechRadar.com, "Google is cracking down hard on deepfakes", 31 May 2022, at https://www.techradar.com/news/google-is-cracking-down-hard-on-deepfakes; reseach.google.com, "Colaboratory: Frequently Asked Questions", at https://research.google.com/colaboratory/faq.html (page visited on 28 August 2022): "We prohibit actions associated with bulk compute, actions that negatively impact others, as well as actions bypassing our policies. The following are disallowed from Colab runtimes: […] creating deepfakes."

[33] Reuters.com, "Exclusive: Google, Facebook, Twitter to tackle deepfakes or risk EU fines", 14 June 2022, at https://www.reuters.com/technology/google-facebook-twitter-will-have-tackle-deepfakes-or-risk-eu-fines-sources-2022-06-13/.

[33] Reuters.com, "China issues draft rules for fake in cyberspace", 28 January 2022, at https://www.reuters.com/world/china/china-regulator-issues-draft-rules-cyberspace-content-providers-2022-01-28/.

[33] Meta, "Enforcing Against Manipulated Media", 6 January 2020, at https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/.

[33] See for instance, Markus Appel, Fabian Prietzel, "The detection of political deepfakes", *Journal of Computer-Mediated Communication*, 2022, Vol. 27, No. 4, at https://academic.oup.com/jcmc/article/27/4/zmac008/6650406; Matthew Bodi, "The First Amendment Implications of Regulating Political Deepfakes", *Rutgers Computer and Technology Law Journal*, 2021, Vol. 47, No. 1, pp. 143-172; Marc Jonathan Blitz, "Deepfakes and Other Non-Testimonial Falsehoods: When is Belief Manipulation (Not) First Amendment Speech?", *Yale Journal of Law& Technology*, 2020, Vol. 23, No. 3, pp. 160-300; Bobby Chesney and Danielle Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, 2019, Vol. 107, No. 6, pp. 1753-1820.

[33] See for instance, Ido Kivovaty, "The international law of cyber intervention" *in* Nicolas Tsagourias and Russel Buchan (eds.), *Research Handbook on International Law and Cyberspace*, 2021, Cheltenham (UK)/Northampton (US)Edgar Elgar Publishing, 2nd ed., 2021, xxviii-634p., pp. 97-112, p. 104; Nicholas Tsagourias, "Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace" *in* Dennis Broeders and Bibi van den Berg (eds.), *Governing Cyberspace: Behavior, Power, and Diplomacy*, Lanham/Boulder/New York/London, 2020, Rowman & Littlefield, vii-327p., pp. 45-63.

[33] This contribution is based on the research project *FAKE-ID: Videoanalyse mit Hilfe künstlicher Intelligenz zur Detektion von falschen und manipulierten Identitäten (meaning "AI-based video analysis to detect false and manipulated identities")*, financed by the German Federal Ministry of Education and Research (BMBF) within the framework of the research programme *Künstliche Intelligenz in der zivilen Sicherheitsforschung ("AI in civil security research")* (FKZ: HWR/FÖPS 13N15737, OVGU 13N15736).
https://www.reuters.com/technology/google-facebook-twitter-will-have-tackle-deepfakes-or-risk-eu-fines-sources-2022-06-13/.

[34] Reuters.com, "China issues draft rules for fake in cyberspace", 28 January 2022, at https://www.reuters.com/world/china/china-regulator-issues-draft-rules-cyberspace-content-providers-2022-01-28/.

[35] Meta, "Enforcing Against Manipulated Media", 6 January 2020, at https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/.

facilitation and intensification of disinformation spreading and other forms of manipulation online.[36] Against this backdrop, this paper will first investigate a number of possible contributions that an AI-based detector for deepfakes could make to the challenge of responding to disinformation as a threat to democracy.[37] Additionally, this paper seeks to analyze and frame the implications of such a tool within the emerging European discourse on digital sovereignty in a global environment.[38] While disinformation is certainly not a new topic, recent technological developments relating to artificial intelligence ("AI")-generated deepfakes have increased the manipulative potential of video and audio-based contents available online, making it a specific, important current challenge in the global and interconnected information context.

One important contextual background element is the current global competition for leadership taking place in the field of artificial intelligence and machine-learning technologies. This primarily involves the two global leading technological poles—namely the United States and China—but also the European Union as well as other states such as Russia. This competition also intervenes in the AI/machine-learning ("ML") regulatory field, with the European Union and China being at the forefront of drafting non-sectorial AI regulatory frameworks.[39] Yet, there is currently no global consensus on AI/ML regulation, despite some important developments such as the adoption of the 2021 UNESCO recommendation on the ethics of artificial intelligence.[40]

In this context, with the European Commission's Proposal for an AI regulation, the European Union is currently trying to occupy this space to establish itself as a global standard-setter with a

[36] See for instance, Markus Appel, Fabian Prietzel, "The detection of political deepfakes", *Journal of Computer-Mediated Communication*, 2022, Vol. 27, No. 4, at https://academic.oup.com/jcmc/article/27/4/zmac008/6650406; Matthew Bodi, "The First Amendment Implications of Regulating Political Deepfakes", *Rutgers Computer and Technology Law Journal*, 2021, Vol. 47, No. 1, pp. 143-172; Marc Jonathan Blitz, "Deepfakes and Other Non-Testimonial Falsehoods: When is Belief Manipulation (Not) First Amendment Speech?", *Yale Journal of Law& Technology*, 2020, Vol. 23, No. 3, pp. 160-300; Bobby Chesney and Danielle Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, 2019, Vol. 107, No. 6, pp. 1753-1820.

[37] See for instance, Ido Kivovaty, "The international law of cyber intervention" *in* Nicolas Tsagourias and Russel Buchan (eds.), *Research Handbook on International Law and Cyberspace*, 2021, Cheltenham (UK)/Northampton (US)Edgar Elgar Publishing, 2nd ed., 2021, xxviii-634p., pp. 97-112, p. 104; Nicholas Tsagourias, "Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace" *in* Dennis Broeders and Bibi van den Berg (eds.), *Governing Cyberspace: Behavior, Power, and Diplomacy*, Lanham/Boulder/New York/London, 2020, Rowman & Littlefield, vii-327p., pp. 45-63.

[38] This contribution is based on the research project *FAKE-ID: Videoanalyse mit Hilfe künstlicher Intelligenz zur Detektion von falschen und manipulierten Identitäten (meaning "AI-based video analysis to detect false and manipulated identities")*, financed by the German Federal Ministry of Education and Research (BMBF) within the framework of the research programme *Künstliche Intelligenz in der zivilen Sicherheitsforschung ("AI in civil security research")* (FKZ: HWR/FÖPS 13N15737, OVGU 13N15736).

[39] See for instance, CBNC.com, "China and Europe are leading the push to regulate A.I. – one of them could set the global playbook" 6 May 2022, at https://www.cnbc.com/2022/05/26/china-and-europe-are-leading-the-push-to-regulate-ai.html.

[40] UNESCO, Recommendation on the ethics of artificial intelligence, November 2021, SHS/BIO/REC-AIETHICS/2021, at https://unesdoc.unesco.org/ark:/48223/pf0000380455.

particular emphasis on a human-centered, ethical, and trustworthy model of AI regulation.[41] Meanwhile, China has recently publicized two legislative drafts that aim to create an ethical framework for regulating AI, as well as one specific proposed bill about deepfakes (called "deep synthesis services" in an unofficial translation).[42] Therefore, there are already emerging and competing regulatory frameworks for AI at national and international level.

In this context, the FAKE-ID project looks at the use of deepfake detection tools by law-enforcement for video-based authentication.[43] Among other research objectives, it aims to react to harmful uses of deepfakes that run counter to EU laws and interests, while ensuring the protection of fundamental rights, democracy, and the rule of law. As we will demonstrate, the overall objective of the FAKE-ID research project is consistent with the nascent European approach on informational digital sovereignty, i.e., "to seek to assert [its] political, economic and social self-determination with regard to digital technology" and to develop its "institutional capacity to reign over developments that affect" societies in the EU.[44]

Various scandals have contributed to making the concept of digital sovereignty more appealing within the European Union, such as the Snowden revelations on US global intelligence practices, the Cambridge Analytica scandal, and the allegations that the 2016 US presidential elections took place under the influence of manipulative data-driven campaigns. We should also not forget the COVID-19 global pandemic. For instance, the presidency of the Council of the European Union stated in its October 2020 conclusions that:

> The COVID-19 pandemic has shown more clearly than ever that Europe must achieve digital sovereignty in order to be able to act with self-determination in the digital sphere and to foster the resilience of the European Union.[45]

This political statement exemplifies the European Union's growing openness towards the necessity to either establish, ensure, or defend its digital sovereignty, including the informational dimensions

[41] European Commission, White paper: On Artificial Intelligence – A European approach to excellence and trust, COM (2020) 65 final, 19. February 2020, at https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

[42] chinalawtranslate.com, Provisions on the Administration of Deep Synthesis Internet Information Services (Draft for solicitation of comments), 28 January 2022, at https://www.chinalawtranslate.com/en/deep-synthesis-draft/, Art. 2.

[43] Comp. about deepfake detection for law-enforcement purposes, Europol Innovation Lab, "Facing Reality? Law Enforcement and the Challenge of Deepfakes", 28. April 2022, at https://www.europol.europa.eu/media-press/newsroom/news/europol-report-finds-deepfake-technology-could-become-staple-tool-for-organised-crime; European Parliament, Artificial Intelligence and Law Enforcement: Impact on Fundamental Rights, Studied Requested by the LIBE committee, PE 656.295, July 2020, at https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf.

[44] Julia Pohle and Daniel Voelsen, "Centrality and Power. The struggle over the techno-political configuration of the Internet and the global digital order", *Policy & Internet*, 2022, Vol. 14, No. 1, pp. 13-27, pp. 20-21.

[45] EU Council, Presidency Conclusions – The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change, 11481/20, 21. Oktober 2020, at https://www.consilium.europa.eu/media/46496/st11481-en20.pdf, p. 3. See also, ibid., p. 5 and p. 7.

of the controls and powers that the EU and its member states can exercise over digital forms of information. There are, for instance, growing concerns about the necessity to safeguard the integrity of electoral processes against the rising digital means of influence over political processes. This is also evident in those EU Council conclusions:

> Direct, universal suffrage and free elections by secret ballots are the basis of the democratic process and a core element of our common values. They need to be preserved in the digital era. Cyberattacks and disinformation targeting electoral processes, campaigns and candidates have the potential to polarize public discourse and undermine the secrecy of the ballot, the integrity and fairness of the electoral process and citizens' trust in elected representatives. In this context, we stress the importance of safeguards and active measures to counter disinformation campaigns, the abuse of private data, hybrid threats and cyberattacks.[46]

The very concept of digital sovereignty remains controversial outside of the EU,[47] but it is especially controversial within the EU in terms of what it concretely entails.[48] Despite the lack of a unified European perspective on digital sovereignty, one emerging consensus in the EU equates digital sovereignty internally with the notion of *strategic autonomy*, and externally with the EU's agenda to establish itself as global leader on the basis of its regulatory powers for digital matters and its worldwide influence via the appeal of its standards in related matters: the so-called *Brussels Effect*.[49] Furthermore, there is a trend in the EU towards ensuring some forms of informational privacy and self-determination for peoples and individuals, especially given the increased technology-driven possibilities to exert manipulative influence over societies, in the global process of digitalization.[50]

[46] Ibid., p. 13, para. 26.

[47] See for some examples of the how conceptions of sovereignty diverge internationally, Anupam Chander and Haochen Sun, "Sovereignty 2.0", *Vanderbilt Journal of Transnational Law*, 2022? Vol. 55, No. 2, pp. 283-324.

[48] See for instance, Andrej Savin, "Digital Sovereignty and Its Impact on EU Policymaking", *Copenhagen Business School Law Research Paper Series No. 22-02*, 4. April 2022, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075106, p. 1; Huw Roberts, Josh Cowls, Frederico Casolari, Jessica Morley, Mariarosaria Taddeo, Luciano Floridi, "Safeguarding European values with digital sovereignty: an analysis of statements and policies", *Internet Policy Review*, 2021, Vol. 10, No. 3, at https://policyreview.info/articles/analysis/safeguarding-european-values-digital-sovereignty-analysis-statements-and-policies, p. 3.

[49] See generally on that concept: Anu Bradford, *The Brussels Effect: how the European Union rules the world*, New York, Oxford University Press, 2020, xix-404p.; Anu Bradford, "The Brussels Effect", *Northwestern University Law Review*, 2012, Vol. 107, No. 1, pp. 1-67. See also, Andrej Savin, "Digital Sovereignty and Its Impact on EU Policymaking", *Copenhagen Business School Law Research Paper Series No. 22-02*, 4. April 2022, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075106, p. 4-6, p. 12; Annegret Bendiek and Isabella Stürzer, "Die digitale Souveränität der EU ist umstritten", *SWP-Aktuell 2022/A 30*, April 2022, at https://www.swp-berlin.org/publikation/die-digitale-souveraenitaet-der-eu-ist-umstritten, pp. 5-6; Julia Pohle and Daniel Voelsen, "Centrality and Power. The struggle over the techno-political configuration of the Internet and the global digital order", *Policy & Internet*, 2022, Vol. 14, No. 1, pp. 13-27, pp. 20-21.

[50] See for instance, Anastasia Iliopoulou-Penot, "The construction of a European digital citizenship in the case law of the Court of Justice of the EU", *Common Market Law Review*, 2022, Vol. 59, No. 4, pp. 969-1006.

Indeed, this is one of the core motivations for the European Commission's Proposal for an AI Regulation.[51]

Against this backdrop, this paper focuses on one specific dimension in this overall global context of digital transformation as accelerated by the ongoing artificial intelligence "revolution": the phenomenon of deepfakes and their potential to exert manipulative influence in a way that can affect democratic and security issues. Deepfakes have been defined in a report by the European Parliamentary Research Service titled "Tackling Deepfakes in European Policy" as "manipulated or synthetic audio or visual media that seem authentic, and which feature people that appear to say or do something they have never said or done, produced using artificial intelligence techniques, including machine learning and deep learning."[52]

This paper relates to a research project led by an interdisciplinary research team of IT, law, social and cultural anthropology scholars, working together in a consortium in Germany funded by the German Federal Ministry of Education and Research.[53] This research project, titled "FAKE-ID," aims at researching AI-based detectors for deepfakes in order to contribute to identifying and reacting to deepfakes for security purposes, and also to foster the self-empowerment potential of a deepfake detector for a public willing to assess the authenticity of video or audio content.

One of the main outcomes of the Fake-ID project will be research demonstrators for conducting risk assessments of video-based deepfake threats in authentication applications that will be investigated for identity remote authentication scenarios.[54] Threats for identity proofing are elaborated. A risk and suspicion map as a basis for decision-making is proposed. One further challenge in prioritization and the application of metrics known as factors to compute a Common Vulnerability Scoring System (CVSS) score for a weakness, to include deepfakes in common vulnerability or weakness enumerations[55] are summarized and discussed.

[51] See for instance, European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, COM(2021) 206 final, 2021/0106(COD), 21 April. 2021, at https://eur-lex.europa.eu/legal-content/EN-DE/TXT/?from=EN&uri=CELEX%3A52021PC0206, p. 21, (15): "[a]side from the many beneficial uses of artificial intelligence, that technology can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices."

[52] European Parliamentary Research Service, Tackling deepfakes in European policy, PE 690.039, Juli 2021, at https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf, p. I.

[53] This contribution is based on the research project *FAKE-ID: Videoanalyse mit Hilfe künstlicher Intelligenz zur Detektion von falschen und manipulierten Identitäten (standing for "AI-based video analysis to detect false and manipulated identities"),* financed by the German Federal Ministry of Education and Research (BMBF) within the framework of the research programme *Künstliche Intelligenz in der zivilen Sicherheitsforschung ("AI in civil security research")* (FKZ: HWR/FÖPS 13N15737, OVGU 13N15736).

[54] See for an example of non-deepfakes based threat to remote ID proofing system, Chaos Computer Club, "Chaos Computer Club hackt Video-Ident", 8 August 2022, at https://www.ccc.de/de/updates/2022/chaos-computer-club-hackt-video-ident.

[55] See for further information on the Common Vulnerability Scoring System (CVSS): first.org, "FIRST is the global Forum of Incident Response and Security Teams", at https://www.first.org/cvss/ and the similar cwe.mitre.org, "CWE

2 Deepfake detection and remote ID proofings, as part of the emerging Union approach on digital sovereignty

Some of the application cases that the FAKE-ID project is looking at concern the use of AI-based tools for detecting deepfakes that can fall under the generic category of remote identity controls or proofing methods (hereafter referred to as "remote ID proofing"). Indeed, remote identity proofing methods "are a way to identify individuals without relying on physical presence."[56] A diverse set of techniques and processes are covered by the formulation that they "can be used in a variety of contexts where trust in the identity of a natural or legal person is essential—such as financial services, e-commerce, travel industry, human resources [and] public administrations".[57]

Remote ID proofing or verification are thus not only performed by public administrations or officials. This is clearly manifest in the growing appeal of such processes of remote ID proofing among private operators such as banks, financial institutions—and in some circumstances, digital service providers.[58] Private actors are also important in this constellation, over and above situations in which they are entrusted to perform remote ID verification, because it is primarily with the help of their datasets that those verification processes are developed and implemented. Indeed, remote ID proofing can be based on several data categories that are collected from various sources and third-party databases—be they private or publicly-owned—which serve as templates or references to verify individuals' identities.[59] This has important implications for data protection and privacy[60] that can have consequences for the compliance of a deepfake detector for law-enforcement purposes with requirements relating to the protection of fundamental rights and the rule of law.[61]

Several methods exist to conduct remote ID proofing. But the approach that is currently most reliable is a combination of several methods, including the use of AI and human intervention to

approach ("Common Weakness Enumeration: A Community-Developed List of Software & Hardware Weakness Types")", Page Last Updated on 20 May 2022, at https://cwe.mitre.org/.

[56] ENISA, Remote Identity Proofing: How to spot the Fake from the Real?", 16 July 2021, at https://www.enisa.europa.eu/news/enisa-news/remote-identity-proofing-how-to-spot-the-fake-from-the-real.

[57] Ibid.

[58] Ibid., p. 21.

[59] Ibid., p. 2.

[60] See ibid., pp. 39-40. See also Julia Pohle and Daniel Voelsen, "Centrality and Power. The struggle over the techno-political configuration of the Internet and the global digital order", *Policy & Internet*, 2022, Vol. 14, No. 1, pp. 13-27, p. 22.

[61] Such a deepfakes-detector besides to be subjected to the requirements of Art. 52(3) of the European Commission Proposal for an AI Regulation, will have to respect the requirements applicable to high-risk AI systems under this future Regulation that are set forth under Title III of the Proposal, including data governance practices that must be met by AI providers or also requirements applicable to training-datasets that constitute a core aspect of the development of AI systems. See about that, European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, COM(2021) 206 final, 2021/0106(COD), 21 April. 2021, at https://eur-lex.europa.eu/legal-content/EN-DE/TXT/?from=EN&uri=CELEX%3A52021PC0206, pp. 26-28, (32), (38) and (40) as well as Ibid., Annex III, Sect. 6(d).

operate verifications or final controls. Such mixed approaches are sometimes named "breed methods."[62] The FAKE-ID research project follows a similar mixed approach for detecting deepfakes. The typical outcome of remote ID proofing is the issuance of a proof of authenticity for a person's identity. This can take several forms, such as a confirmation of identity with attribution of an absolute score (YES/NO), a confidence level as a percentage, a likelihood ratio, or the assignment of identification credentials.[63] Similarly, the tools that the Fake-ID research project is investigating to detect deepfakes will generate a score to establish a confidence level that a picture or a video does not constitute a deepfake.

The trend to develop and implement remote ID proofing is rapidly taking off, triggered by the COVID-19 pandemic crisis, which helped establish identity verification without physical presence.[64] Furthermore, this trend is likely to continue in the EU context, given the plan to establish a European digital identity "wallet" common to all EU citizens.[65] The European Commission has publicized a proposal that aims, among other things, to harmonize remote ID proofing of EU identities, both online and offline.[66] In that context, detecting deepfakes will be of increased importance within the EU, given the EU objective to provide "access to highly secure and trustworthy electronic identity solutions" for cross-border activities, so that "that public and private services can rely on trusted and secure digital identity solutions." Another objective of the EU relevant to mention here is empowering and facilitating the use of digital identity solutions by natural and legal persons, including for secure business transactions and access to public services.[67] In its capacity as president of the EU Council in the first half of 2022, France has more recently submitted a second compromise text for the European digital identity wallet, with the primary aim of "prevent[ing] fragmentation of the internal market" and "to define a pan-European legal framework that allows for the cross-border recognition of trust services for the recording of [identity] data in electronic ledgers."[68] One future practical use-

[62] ENISA, Remote ID Proofing Analysis of methods to carry out identity proofing remotely, March 2021, at https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing, p. 25.

[63] Ibid., p. 15.

[64] See for instance, ENISA, Remote ID Proofing Analysis of methods to carry out identity proofing remotely, March 2021, at https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing, p. 4.

[65] European Commission, European Digital Identity, 28 May 2021, at https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-digital-identity_en#digital-identity-for-all-europeans.

[66] European Commission, Proposal for a Regulation of European Parliament and of the Council amending Regulation (EU) No 910/2014 as regards establishing a framework for a European Digital Identity, 2021/0136(COD), COM(2021) 281 final, 3 June 2021, p. 2.

[67] Ibid., p. 1.

[68] European Commission, Proposal for a Regulation of European Parliament and of the Council amending Regulation (EU) No 910/2014 as regards establishing a framework for a European Digital Identity, 2021/0136(COD), COM(2021) 281 final, 3 June 2021, p. 19, (34) and pp. 23-25. See also for a slightly revised definition of Digital Identity Wallet as amended by the French Presidency of the EU Council, Conseil de l'Union européenne, Proposition de règlement du Parlement européen et du Conseil modifiant le règlement (UE) n°910/2014 en ce qui concerne l'établissement d'un cadre européen relatif à une identité numérique – Deuxième proposition de compromis, 2021/0136(COD), 9200/22, at https://aeur.eu/f/1v7, pp. 19-20, Article 6A.

case of secure access to public services could be to ensure the cyber-security conditions of election infrastructures, including online identity verification for possible forms of e-vote in the future.

The examples show the importance of ensuring secure digital means of identification and identification controls without the need for an individual's physical presence. This will continue to grow in importance and will be taken into account in future regulatory frameworks. In this context, the detection of deepfakes in the overall EU regulatory framework for AI systems will constitute an important aspect of the European digital identity wallet, and will be indirectly relevant for a European conception of digital sovereignty.

3 DEEPFAKES AS POTENTIAL INFORMATIONAL THREATS FOR THE EUROPEAN UNION'S DIGITAL SOVEREIGNTY

The phenomenon of deepfakes potentially affects digital sovereignty as it offers technological means to manipulate digitalized or digital means of identity, including official means of identification.[69] Deepfakes also contribute to contemporary discussions about how informational digital sovereignty can be ensured in light of external "informational threats," with the official aim in the EU to follow a different path than existing illiberal or authoritarian conceptions of informational digital sovereignty. This is a pressing issue due to rapid technological developments that are enabling to generate more and more elaborate deepfakes, but also due to their "democratization" and increasing spread within societies globally.[70]

As mentioned in the introduction, despite the disputed contours of the EU perspective over digital and informational sovereignty, one current minimum consensus within the EU equates sovereignty in the digital context with the objective of ensuring strategic autonomy, mostly against external threats and including informational threats. Against this backdrop, one important constellation for reaching strategic autonomy translates into ensuring cybersecurity, therefore connecting sovereignty and cybersecurity.[71] Indeed, several EU digital policy milestones have emerged in connection with cyber security issues, as shown by the strengthened role attributed to the European Union Agency for Cybersecurity ("ENISA") after the adoption of the EU Cybersecurity

[69] See for some examples in a security-oriented perspective, Europol's European Cybercrime Centre, United Nations Interregional Crime and Justice Research Institute (UNICRI) and Trend Micro, Report on Malicious Uses and Abuses of Artificial Intelligence (AI), 19. November 2020, at https://eucrim.eu/news/report-on-malicious-uses-and-abuses-of-artificial-intelligence/, pp. 54-65.

[70] See for instance in the scientific context, Chemistryworld.com, "AI-generated images could make it almost impossible to detect fake papers", 24 May 2022, at https://www.chemistryworld.com/news/ai-generated-images-could-make-it-almost-impossible-to-detect-fake-papers/4015708.article.

[71] Andrej Savin, "Digital Sovereignty and Its Impact on EU Policymaking", *Copenhagen Business School Law Research Paper Series No. 22-02*, 4. April 2022, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075106, p. 4.

Regulation in 2019.[72] Since the adoption of the 2019 EU Cybersecurity Act, ENISA has been entrusted among others with the task of "contribut[ing] to the development and implementation of Union policy and law," by "supporting […] the development and implementation of Union policy in the field of electronic identity and trust services."[73]

In this context, one important question is whether deepfakes can constitute a cybersecurity threat to an emerging EU digital sovereignty conception. This question requires a nuanced reflection, for three reasons. First, various application cases already exist for deepfakes, but it is important to note that deepfakes can also be used for artistic,[74] political,[75] educational,[76] entertainment,[77] or medical purposes.[78] Second, there are two constellations wherein deepfakes can clearly deepen "cybersecurity threats," namely disinformation and identity manipulations/thefts. Third, both constellations can interrelate in the particular context of elections, where digital identity plays a crucial role: Either (i) they can be used in formal contexts, to verify the authenticity of nationals allowed to vote for an election, or (ii) in an informal way, to prevent online contents from being spread online via accounts using fake identities to manipulate electoral processes in the increasingly digitalized dimensions of public debates.[79] Admittedly, the second constellation poses delicate issues given the anonymity that users of online platforms can often enjoy.

Furthermore, protection against identity manipulation and thefts refers to the general concept of sovereignty, insofar as it deeply relates to the concept of nationality and to the emerging legal concept of digital citizenship in EU law. Under general international law, one of the traditional core prerogatives of states is to attribute nationality to individuals and verify it. However, there is "no

[72] Annegret Bendiek and Isabella Stürzer, "Die digitale Souveränität der EU ist umstritten", *SWP-Aktuell 2022/A 30*, April 2022, at https://www.swp-berlin.org/publikation/die-digitale-souveraenitaet-der-eu-ist-umstritten, pp. 3-4.

[73] European Parliament and EU Council, Regulation (EU) 2019/881 on ENISA (the European Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), PE/86/2018/REV/1, Art. 5.

[74] See for instance, wired.co.uk, "These historical artefacts are totally faked", 24 October 2021, at https://www.wired.co.uk/article/fake-artefacts-ai.

[75] See for instance, Umur A. Ciftci, Gokturk Yuksek, Ilke Demir, "My Face My Choice: Privacy Enhancing Deepfakes for Social Media Anonymisation", 2 November 2023, at https://arxiv.org/pdf/2211.01361v1.pdf.

[76] See for instance, Wired.com, "Deepfakes Are Becoming the Hot New Corporate Training Tool", 7 July 2020, at https://www.wired.com/story/covid-drives-real-businesses-deepfake-technology/.

[77] See for instance, Ft.com, "Deepfakes: Hollywood's quest to create the perfect digital human", 10 October 2019, at https://www.ft.com/content/9df280dc-e9dd-11e9-a240-3b065ef5fc55.

[78] European Parliamentary Research Service, Tackling deepfakes in European policy, PE 690.039, Juli 2021, at https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf.

[79] See the eight scenarios developed for illustrating ethical harms that could be generated by the use of deepfakes in the context of elections, Nicholas Diakopoulos and Deborah Johnson, "Anticipating and addressing the ethical implications of deepfakes in the context of elections", *New Media & Society*, 2021, Vol. 23, No. 7, pp. 2072-2098. See also Tom Dobber, Nadia Metoui, Damian Trilling, Nathalie Helberger, Claes de Vreese, "Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?", *The International Journal of Press/Politics*, 2021, Vol. 26, No. 1, pp. 69-91. See about allegations regarding the use of false identity online in the context of the 2016 U.S. presidential elections, Michael N. Schmitt, ""Virtual" Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law", *Chicago Journal of International Law*, 2018, Vol. 19, No. 1, pp. 30-67, p. 36.

coherent, accepted definition of nationality in international law and only conflicting descriptions under the different municipal laws of states." Furthermore, "the rights and duties attendant upon nationality vary from state to state."[80] That said, nationality still constitutes a core link between states and "their" peoples and therefore directly involves the concept of sovereignty.[81] This principle was confirmed by the Permanent Court of International Justice in its 1923 *Nationality Decrees in Tunis and Morocco* case, in which it was decided that "it is for each state to determine under its own law who are its nationals," before adding that this "law shall be recognized by other states in so far as [applicable international law] with regard to nationality."[82] In its 1955 *Nottebohm* case, the International Court of Justice established that under international law, nationality with respect to the State granting it, is "a legal bond having as its basis a social fact of attachment, a genuine connection of existence, interests and sentiments, together with the existence of reciprocal rights and duties."[83] If international law—and in particular international human rights law—has restricted the freedom of States to attribute and control the nationality of persons under its jurisdiction since the 1950s, this still remains the position of principle under currently applicable international law:

> Even if the freedom of States to regulate their nationality is much more restricted today, considering the development of international law since 1923, that [Permanent International Court of Justice's] statement is essentially still valid: each State is in principle still entitled to determine under its own law who are its nationals (cf. Art. 3 (1) European Convention on Nationality). International law limits that discretion, but it neither contains nor prescribes certain criteria for acquisition and loss of nationality.[84]

Neither processes of digitalizing traditional forms of identification documents and establishing digital forms of identification documents and controls fundamentally disturb this core relation between a

[80] Malcom N. Shaw, *International Law*, 2014, Cambridge, CUP, 7th Edition, 1063p., p. 479.
[81] Matthias Leese, "Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU", *Geopolitics*, 2022, Vol. 27, No. 1, pp. 113-133, p. 114.
[82] Permanent Court of International Justice, *Nationality Decrees in Tunis and Morocco Case*, Series B, No. 4, 1923; 2 AD, p. 24.
[83] International Court of Justice, Nottebohm Case (second phase) (Lichtenstein v. Guatemala), Judgment of April 16th, 1955, I.C.J. Reports, p. 4, p. 23.
[84] Oliver Dörr, "Nationality", *Max Planck Encyclopedia of International Law*, August 2019, para. 4.

state's sovereignty and personal jurisdiction[85] as exercised over nationals from that state,[86] subject to restrictions imposed by international law.

Identity manipulation is increasingly perceived in the EU as a potential threat to its sovereignty, because it could endanger EU laws, interests, and values if it materialized at a general level.[87] Such a claim might appear exaggerated in the contemporary context, but as digital forms of identification processes and related verification mechanisms gradually grow, so too will such threats, especially deepfake-based threats. Even if such major deepfake-based threats have not yet made their presence felt,[88] there is a serious possibility, evidenced in the many regulatory efforts to counter them, that security threats may rapidly increase, since deepfakes are increasingly easier to create and use.[89] The logic increasingly at play here that links digital sovereignty and identity control in the process of digitalization[90] can indeed be broadly compared to the phenomenon of smart borders,[91] despite the fact that remote ID proofing is meant to apply independently of any proper physical or digitalized borders and does not focus on the control and identification of foreigners. Several legal instruments have already been adopted with that aim at the EU level, that is, in order to regulate how digital forms of identity can be established and controlled in the EU without the individuals in question being physically present. That said, these developments can also be critically assessed, as they for instance arguably participate to frame the figure of the "foreigner" as an indirect threat to the EU sovereignty, especially given wide-spread securitization discourses in EU policies and legislations.[92]

[85] In this sense, see for instance Onuma Yasuaki, *International Law in a Transcivilizational World*, Cambridge, Cambridge University Press, 2017, 666p., p. 333: "This bond constitutes the basis for a state's jurisdiction over its members (so-called personal sovereignty or jurisdiction). A state can apply its law over tis nationals even when they are outside of its territory, including within the territory of a foreign state […]. Furthermore, while nationality is fundamentally a concept concerning natural persons, it plays an intermediary function in the application of jurisdiction over juridical persons, whose activities transcend national borders."

[86] Comp. with Ewa Michalkiewicz-Kadziela, Ewa Milczarek, "Legal boundaries of digital identity creation", *Internet Policy Review*, 2022, Vol. 11, No. 1, at https://policyreview.info/articles/analysis/legal-boundaries-digital-identity-creation.

[87] See for instance the concept of Foreign Information Manipulation and Interference (FIMI) used by some European agencies, ENISA, Foreign information manipulation and interference (FIMI) and cybersecurity – Threat Landscape, December 2022, at https://www.enisa.europa.eu/news/cybersecurity-foreign-interference-in-the-eu-information-ecosystem.

[88] See for instance, Trend Micro, "How Underground Groups Use Stolen Identities and Deepfakes", 27 September 2022, at https://www.trendmicro.com/en_us/research/22/i/how-underground-groups-use-stolen-identities-and-deepfakes.html.

[89] See for instance, U.S. Congressional Research Service, "Deep Fakes and National Security", Updated on 3 June 2022, at https://crsreports.congress.gov/product/pdf/IF/IF11333.

[90] See on that interrelation, Matthias Leese, "Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU", *Geopolitics*, 2022, Vol. 27, No. 1, pp. 113-133, esp. pp. 116-120.

[91] See on that phenomenon, Ayelet Shachar, *The Shifting Border: Legal Cartographies of Migration and Mobility*, Manchester University Press, 2020, 328 p.; Jonas Püschmann, Book Review: The Shifting Border: Legal Cartographies of Migration and Mobility, 18 March 2022, at https://blogs.law.ox.ac.uk/research-subject-groups/centre-criminology/centreborder-criminologies/blog/2022/03/book-review.

[92] See for instance, euronews.com, "Joseph Borell apologises for controversial 'garden vs jungle' metaphor but defends speech", 20 November 2022, at https://www.euronews.com/my-europe/2022/10/19/josep-borrell-apologises-for-controversial-garden-vs-jungle-metaphor-but-stands-his-ground.

Remote ID verification increasingly relates to the emerging EU approach for digital sovereignty, as understood in its minimalistic conception that relate to strategic autonomy and cyber security concerning so-called informational threats. Art. 24(1) of Regulation (EU) 910/2014 on electronic identification and trust services ("eIDAS")[93] is a good example of how remote ID proofing *already relates* to the exercise of states' sovereignty through digital means, by relativizing the traditional importance of physical powers exercised territorially by sovereign states in the international society. This provision mandates that:

> [w]hen issuing a qualified certificate for trust service, a qualified certificate for a trust service provider shall verify, by appropriate means and in accordance with national law, the identity and, if applicable, any specific attributes of the natural or legal person to whom the qualified certificate is issued.[94]

The link between remote ID proofing and sovereignty is strengthened in some use cases for which EU law requires a verification of the identity of persons in online transactions for the purposes of countering money laundering or terrorism financing, as foreseen by the Anti-Money Laundering/Counter Financing Terrorism (AMT/CFT) directives. The fifth AML/CFT Directive was adopted to also strengthen the possibilities for the EU to monitor financial transactions, including regarding the identity of persons involved in those transactions, especially with respect to third countries that are regarded as a source of risk due to an insufficient level of control over money laundering and terrorism financing.[95] Art. 9 of the fifth AMD/CFT Directive seeks, for instance, to protect the integrity of the European financial system, which also relates to an emerging minimal understanding of European digital sovereignty.[96] More generally, the exploitation of cyber security vulnerabilities via the manipulation of identities[97] can lead to the emergence of threats that are not

[93] European Parliament and EU Council, Regulation (EU) 910/2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC, *Official Journal of the European Union L 257/73*, 28 August 2014.

[94] European Parliament and EU Council, Regulation (EU) 910/2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC, *Official Journal of the European Union L 257/73*, 28 August 2014, p. 26, Art. 24(1).

[95] European Commission, "Strengthened EU rules to prevent money laundering and terrorism financing" (Fact sheet), 9 July 2018, at https://ec.europa.eu/info/files/factsheet-main-changes-5th-anti-money-laundering-directive_en.

[96] European Parliament and EU Council, Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, amending Regulation (EU) No 648/2012 of the European Parliament and of the Council, repealing Directive 2005/60/EC, 20 May 2015, p. 18, Art. 9(1). Comp. with Andrej Savin, "Digital Sovereignty and Its Impact on EU Policymaking", *Copenhagen Business School Law Research Paper Series No. 22-02*, 4. April 2022, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075106, p. 2.

[97] See for some real examples of cyber security threats based on manipulated identities, Trend Micro, "How Underground Groups Use Stolen Identities and Deepfakes", 27 September 2022, at https://www.trendmicro.com/en_us/research/22/i/how-underground-groups-use-stolen-identities-and-deepfakes.html.

only damaging for parties involved in an online transaction or communication but more broadly for a country, if a sufficient gravity threshold is reached.[98]

The fact that consensus over digital sovereignty within the EU is mostly found in relation to the notion of strategic autonomy and cybersecurity threats is illustrated by several recent EU Council conclusions that all put the emphasis on the importance of cybersecurity and informational self-determination for the EU approach on digital sovereignty.[99] Developments at the EU level clearly indicate a willingness to move forward with a common cybersecurity strategy serving the whole European Union's digital and informational sovereignty. Indeed, they underscore the fact that securing digital means of identity as well as cybersecurity processes aiming specifically at the protection of the integrity of decision-making processes are increasingly influential in the emergence of a minimal European understanding of digital sovereignty.

4 Conclusion

In this context, deepfake detection exerts a narrow but nonetheless important role. Indeed, deepfakes are increasingly perceived as being able to threaten decision-making processes in the global context of digitalization, while posing threats to the security of persons and societies within the European Union. For this reason, deepfake detection and its use for remote ID verification integrates the emerging Union approach over digital informational sovereignty, which is for the moment mostly focused on ensuring security and strategic autonomy, while protecting fundamental rights, democracy and the rule of law.

5 Acknowledgments

The work in this paper is funded in part by the German Federal Ministry of Education and Research (BMBF) under FAKE-ID project: grant numbers OVGU FKZ: 13N15736 and HWR/FPÖS FKZ: 13N15737. We would like to thank all project partners for the fruitful discussion and exchange in the project.

[98] ENISA, Remote ID Proofing Analysis of methods to carry out identity proofing remotely, March 2021, at https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing, pp. 45-46.

[99] EU Council, Council conclusions on Foreign Informational Manipulation and Interference (FIMI), 11429/22, 18 July 2022, at https://data.consilium.europa.eu/doc/document/ST-11429-2022-INIT/en/pdf; EU Council, Council conclusions on a Framework for a coordinated EU response to hybrid campaigns, 10016/22, 21 June 2022, at https://data.consilium.europa.eu/doc/document/ST-10016-2022-INIT/en/pdf; EU Council, Council conclusions on the Special Report of the European Court of Auditors No 05/2022 entitled 'Cybersecurity of the EU Institutions, bodies and agencies: Level of preparedness overall not commensurate with the threats, 10504/22, 21 June 2022, at https://data.consilium.europa.eu/doc/document/ST-10016-2022-INIT/en/pdf.

We thank the participants in our panel during the 2022 Weizenbaum Institute Conference and we would like furthermore to thank Prof. Harmut Aden (HWR) for his comments and suggestions, as well as Mario Petoshati (HWR) for his thorough review and editing.

6 REFERENCES

1. Appel M., Prietzel F. (2022). The detection of political deepfakes. Journal of Computer-Mediated Communication, 27(4), at https://academic.oup.com/jcmc/article/27/4/zmac008/6650406 (visited on 7 February 2023).
2. Bendiek A., Stürzer I. (2022). Die digitale Souveränität der EU ist umstritten. SWP-Aktuell 2022/A 30. At https://www.swp-berlin.org/publikation/die-digitale-souveraenitaet-der-eu-ist-umstritten (visited on 29 August 2022).
3. Bodi M. (2021), The First Amendment Implications of Regulating Political Deepfakes. Rutgers Computer and Technology Law Journal, 47(1), 143-172.
4. Blitz M. J. (2020), Deepfakes and Other Non-Testimonial Falsehoods: When is Belief Manipulation (Not) First Amendment Speech? Yale Journal of Law & Technology, 23(3), 160-300.
5. Bradford A. (2020). The Brussels Effect: how the European Union rules the world. New York, Oxford University Press, xix-404p.
6. Bradford A. (2012). The Brussels Effect. Northwestern University Law Review, 107 (1), 1-67.
7. CBNC.com, "China and Europe are leading the push to regulate A.I. – one of them could set the global playbook"6 May 2022, at https://www.cnbc.com/2022/05/26/china-and-europe-are-leading-the-push-to-regulate-ai.html.
8. Chander A., Sun H. (2022). Sovereignty 2.0. Vanderbilt Journal of Transnational Law, 55 (2), 283-324.
9. Chesney B., Citron D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review, 107(6), 1753-1820.
10. Chaos Computer Club (2022). Chaos Computer Club hackt Video-Ident, 8 At https://www.ccc.de/de/updates/2022/chaos-computer-club-hackt-video-ident (page visited on 28 August 2022).
11. Chemistryworld.com (2022). AI-generated images could make it almost impossible to detect fake papers. At https://www.chemistryworld.com/news/ai-generated-images-could-make-it-almost-impossible-to-detect-fake-papers/4015708.article (page visited on 28 August 2022).
12. chinalawtranslate.com, Provisions on the Administration of Deep Synthesis Internet Information Services (Draft for solicitation of comments), 28 January 2022, at https://www.chinalawtranslate.com/en/deep-synthesis-draft/.
13. Ciftci U. A., Yuksek G., Demir I (2023). My Face My Choice: Privacy Enhancing Deepfakes for Social Media Anonymisation, at https://arxiv.org/pdf/2211.01361v1.pdf.
14. Conseil de l'Union européenne (2021), Proposition de règlement du Parlement européen et du Conseil modifiant le règlement (UE) n°910/2014 en ce qui concerne l'établissement d'un cadre européen relatif à une identité numérique – Deuxième proposition de compromis. 2021/0136(COD), 9200/22.
15. cwe.mitre.org (2022). CWE approach ("Common Weakness Enumeration: A Community-Developed List of Software & Hardware Weakness Types"). At https://cwe.mitre.org/ (page visited on 28 August 2022).
16. Diakopoulos N., Johnson D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. New Media & Society, 23(7), 2072-2098.
17. Dobber T., Metoui N., Trilling D., Helberger N., de Vreese C. (2021). Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?. The International Journal of Press/Politics, 26(1), 69-91.

18. Dörr (2019). Nationality. Max Planck Encyclopedia of International Law.

19. ENISA (2021). Remote Identity Proofing: How to spot the Fake from the Real?. At https://www.enisa.europa.eu/news/enisa-news/remote-identity-proofing-how-to-spot-the-fake-from-the-real (page visited on 28 August 2022).

20. ENISA (2022). Foreign information manipulation and interference (FIMI) and cybersecurity – Threat Landscape. At https://www.enisa.europa.eu/news/cybersecurity-foreign-interference-in-the-eu-information-ecosystem.

21. euronews.com (2022), Joseph Borell apologises for controversial 'garden vs jungle' metaphor but defends speech. At https://www.euronews.com/my-europe/2022/10/19/josep-borrell-apologises-for-controversial-garden-vs-jungle-metaphor-but-stands-his-ground.

22. European Commission (2021)(a). Proposal for a Regulation of European Parliament and of the Council amending Regulation (EU) No 910/2014 as regards establishing a framework for a European Digital Identity. 2021/0136(COD), COM(2021) 281 final.

23. European Commission (2021)(b). European Digital Identity. At https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-digital-identity_en#digital-identity-for-all-europeans (page visited on 28 August 2022).

24. European Commission (2021)(c). Proposal for a Regulation of the European Parliament and of the Council Laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. COM(2021) 206 final, 2021/0106(COD).

25. European Commission (2018). Strengthened EU rules to prevent money laundering and terrorism financing" (Fact sheet). At https://ec.europa.eu/info/files/factsheet-main-changes-5th-anti-money-laundering-directive_en (page visited on 28 August 2022).

26. European Commission (2020). White paper: On Artificial Intelligence – A European approach to excellence and trust. COM (2020) 65 final. At https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (visited on 28 August 2022).

27. EU Council (2020) Presidency Conclusions – The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change. 11481/20. At https://www.consilium.europa.eu/media/46496/st11481-en20.pdf.

28. European Parliament (2020). Artificial Intelligence and Law Enforcement: Impact on Fundamental Rights, Studied Requested by the LIBE committee, PE 656.295. At https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf.

29. European Parliament and EU Council (2019). Regulation (EU) 2019/881 on ENISA (the European Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act). PE/86/2018/REV/1.

30. European Parliament and EU Council (2015). Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, amending Regulation (EU) No 648/2012 of the European Parliament and of the Council, repealing Directive 2005/60/EC.

31. European Parliament and EU Council (2014). Regulation (EU) 910/2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive *1999/93/EC. Official Journal of the European Union L 257/73*.

32. European Parliamentary Research Service (2021). Tackling deepfakes in European policy. PE 690.039. At https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf (page visited on 28 August 2022).

33. Europol’s European Cybercrime Centre, United Nations Interregional Crime and Justice Research Institute (UNICRI) and Trend Micro (2020). Report on Malicious Uses and Abuses of Artificial Intelligence (AI). At https://eucrim.eu/news/report-on-malicious-uses-and-abuses-of-artificial-intelligence/.

34. Europol Innovation Lab (2022. Facing Reality? Law Enforcement and the Challenge of Deepfakes. At https://www.europol.europa.eu/media-press/newsroom/news/europol-report-finds-deepfake-technology-could-become-staple-tool-for-organised-crime.

35. first.org (2022). FIRST is the global Forum of Incident Response and Security Teams. At https://www.first.org/cvss/ (page visited on 28 August 2022).

36. Ft.co (2019). Deepfakes: Hollywood’s quest to create the perfect digital human. At at https://www.ft.com/content/9df280dc-e9dd-11e9-a240-3b065ef5fc55 (page visited on 28 August 2022).

37. Iliopoulou-Penot A. (2022). The construction of a European digital citizenship in the case law of the Court of Justice of the EU. Common Market Law Review, 59 (4), 969-1006.

38. International Court of Justice (1955). Nottebohm Case (second phase) (Lichtenstein v. Guatemala). Judgment of April 16th, 1955, I.C.J. Reports, p. 4.

39. Kivovaty, I. (2021). The international law of cyber intervention *in* Tsagourias N. and Buchan R. (eds.), *Research Handbook on International Law and Cyberspace* (Cheltenham (UK)/Northampton (US); Edgar Elgar Publishing: 2021), 97-112.

40. Leese (M.) (2022). Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU. Geopolitics, 27(1), 113-133.

41. Meta (2020). Enforcing Against Manipulated Media. At https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/ (page visited on 28 August 2022).

42. Michalkiewicz-Kadziela E., Milczarek E. (2022). Legal boundaries of digital identity creation. Internet Policy Review, 11 (1).

43. reseach.google.com, “Colaboratory: Frequently Asked Questions”, at https://research.google.com/colaboratory/faq.html (page visited on 28 August 2022).

44. Roberts H., Cowls J., Casolari F., Morley J., Taddeo M., Floridi F. (2021). Safeguarding European values with digital sovereignty: an analysis of statements and policies. Internet Policy Review, 10 (3), at https://policyreview.info/articles/analysis/safeguarding-european-values-digital-sovereignty-analysis-statements-and-policies (visited on 28 August 2022).

45. reuters.com, “Exclusive: Google, Facebook, Twitter to tackle deepfakes or risk EU fines”, 14 June 2022, at https://www.reuters.com/technology/google-facebook-twitter-will-have-tackle-deepfakes-or-risk-eu-fines-sources-2022-06-13/.

46. Permanent Court of International Justice (1923). Nationality Decrees in Tunis and Morocco Case, Series B, No. 4, 1923; 2 AD.

47. Pohle J., Voelsen D. (2022). Centrality and Power. The struggle over the techno-political configuration of the Internet and the global digital order. Policy & Internet, 14 (1), 13-27.

48. Savin A. (2022). Digital Sovereignty and Its Impact on EU Policymaking. Copenhagen Business School Law Research Paper Series No. 22-02, , at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075106 (page visited on 28 August 2022).

49. Schmitt M. N. (2018). “Virtual” Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law. Chicago Journal of International Law, 19 (1), 30-67.

50. Shaw M. N. (2014) International Law. 2014, 7th Edition, Cambridge, Cambridge University Press, 1063p..

51. TechRadar.com, “Google is cracking down hard on deepfakes”, 31 May 2022, at https://www.techradar.com/news/google-is-cracking-down-hard-on-deepfakes.

52. Trend Micro (2022). How Underground Groups Use Stolen Identities and Deepfakes. At https://www.trendmicro.com/en_us/research/22/i/how-underground-groups-use-stolen-identities-and-deepfakes.html.

53. Tsagourias N. (2020). “Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace” in Broeders D. and van den Berg B. (eds.), Governing Cyberspace: Behavior, Power, and Diplomacy (Lanham/Boulder/New York/London; Rowman & Littlefield: 2020), 45-63.

54. UNESCO, Recommendation on the ethics of artificial intelligence, November 2021, SHS/BIO/REC-AIETHICS/2021, at https://unesdoc.unesco.org/ark:/48223/pf0000380455.

55. U.S. Congressional Research Service (2022). Deep Fakes and National Security. At https://crsreports.congress.gov/product/pdf/IF/IF11333 (page visited on 28 August 2022).

56. wired.co.uk (2021). These historical artefacts are totally faked. At https://www.wired.co.uk/article/fake-artefacts-ai (page visited on 28 August 2022).

57. Wired.com (2020).Deepfakes Are Becoming the Hot New Corporate Training Tool At https://www.wired.com/story/covid-drives-real-businesses-deepfake-technology/ (28 August 2022).

58. Yasuaki O. (2017). International Law in a Transcivilizational World, Cambridge, Cambridge University Press.