4. Cybersecurity of AI medical devices: risks, legislation, and challenges

Elisabetta Biasin, Erik Kamenjašević and Kaspar Rosager Ludvigsen

1. INTRODUCTION

Medical devices and artificial intelligence (AI) systems could rapidly transform the provision of healthcare. However, due to their nature, medical devices incorporating AI might be exposed to cyberattacks, leading to patient safety and security risks.[1] For example, an AI insulin pump under cyberattack could stop working correctly and provoke serious health risks for the patient using it.[2] Alongside the health-related consequences that might turn fatal, cyberattacks on AI medical devices could also provoke indirect effects, ranging from diminishing patient trust in the security of the healthcare system to hesitancy towards using these AI medical devices due to their cybersecurity-related vulnerabilities.[3]

After this introduction, the chapter is divided into three sections. Section 2 explains the role of cybersecurity in healthcare and, subsequently, defines AI classified as a standalone medical device or AI that operates as a component of a device.[4] To illustrate the risks posed by medical devices, we provide three examples: the poisoning of datasets, social engineering

[1] In this chapter, 'AI medical devices' refers to AI operating as a medical device or in a medical device following the caveats explained under Section 3. Contrary to the EU, the United States' Food and Drug Administration (FDA) uses definitions of 'Software as Medical Devices' as defined by the International Medical Device Regulators Forum (IMRDF), see IMDRF Software as a Medical Device (SaMD) Working Group, '"Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations' (International Medical Device Regulators Forum 2014). Definitions for when it is a medical device can be found in FDA, 'How to Determine if Your Product is a Medical Device' (*US Food & Drug Administration,* 29 September 2022) <https:// www .fda .gov/ medical -devices/ classify -your -medical -device/ how -determine -if -your -product-medical-device> accessed 29 April 2024. For critical analysis of this, see eg, Dhruv B Pai, 'Mapping the Genealogy of Medical Device Predicates in the United States' (2021) 16 PLoS ONE 1.

[2] Tamar Levy-Loboda and others, 'Personalized Insulin Dose Manipulation Attack and Its Detection Using Interval-Based Temporal Patterns and Machine Learning Algorithms' (2022) 132 Journal of Biomedical Informatics 1.

[3] Elisabetta Biasin and Erik Kamenjašević, 'Cybersecurity of Medical Devices: New Challenges Arising from the AI Act and NIS 2 Directive Proposals' (2022) 3 International Cybersecurity Law Review 163, 164.

[4] AI will be regulated separately, but can be considered a medical device or an accessory to one. See Kaspar Ludvigsen, Shishir Nagaraja and Angela Daly, 'When Is Software a Medical Device? Understanding and Determining the "Intention" and Requirements for Software as a Medical Device in European Union Law' (2021) 13 European Journal of Risk Regulation 78.

and data or source code extraction. Section 3 provides an overview of the European Union (EU) regulatory framework relevant to ensuring the cybersecurity of AI medical devices. That framework includes Regulation (EU) 2017/745 on medical devices (MDR),[5] Directive (EU) 2016/1148 concerning measures for a high common level of security of network and information systems across the Union (NIS Directive),[6] Regulation (EU) 2019/881 on ENISA and on information and communications technology cybersecurity certification (Cybersecurity Act),[7] and Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR).[8] Two EU laws that, once adopted, will impact the existing obligations concerning the cybersecurity of AI medical devices are also addressed as part of the framework. These are the Artificial Intelligence Act (AI Act)[9] and the proposed NIS 2 Directive.[10] Finally, section 4 examines possible challenges stemming from the EU regulatory framework – in particular, challenges deriving from the two laws and their interaction with the existing legislation concerning AI medical devices' cybersecurity. They are structured as answers to the following questions: (1) How will the AI Act interact with the MDR regarding the cybersecurity and safety requirements? (2) How should we interpret incident notification requirements from the NIS 2 Directive and the MDR? (3) What are the consequences of the evolving terms of critical infrastructure?[11]

[5] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017, on Medical Devices, Amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and Repealing Council Directives 90/385/EEC and 93/42/EEC [2017] OJ L 117/1 (MDR). For a soft law document relevant for the interpretation of different provisions, roles and obligations in the context of the medical devices' cybersecurity, see Medical Device Coordination Group (MDCG), 'MDCG 2019-16, Rev.1, Guidance on Cybersecurity for Medical Devices' (2020) (MDCG Guidance). The MDCG Guidance deals with the cybersecurity-related provisions embedded in the MDR. This non-binding document provides a comprehensive overview of cybersecurity-related requirements that manufacturers must implement to comply with the MDR and to ensure the appropriate level of cyber resilience of the medical device.

[6] Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016, concerning Measures for a High Common Level of Security of Network and Information Systems Across the Union [2016] OJ L 194/1 (NISD).

[7] Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on Information and Communications Technology Cybersecurity Certification and Repealing Regulation (EU) No 526/2013 (Cybersecurity Act) [2019] OJ L 151/15 (CSA).

[8] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1 (GDPR).

[9] European Parliament Legislative Resolution of 13 March 2024 on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD) (AI Act).

[10] Commission, 'Proposal for a Directive of the European Parliament and of the Council on Measures for a High Common Level of Cybersecurity Across the Union, Repealing Directive (EU) 2016/1148' COM (2020) 823 final (NIS 2 Directive).

[11] This chapter lists three legal challenges. There could be others insisting on medical device cybersecurity, such as cybersecurity certification, see Elisabetta Biasin and Erik Kamenjašević,

2. CYBERSECURITY IN HEALTHCARE

As the uptake of medical devices that are part of the Internet of Things (IoT) increases, it will be necessary for the cybersecurity landscape to improve.[12] In cybersecurity, anything with an operating system, and perhaps network access, is vulnerable. This includes anything with these attributes that is used in healthcare.[13] Like regular security, cybersecurity involves adversaries and defenders, and the risks presented by the failures to mitigate or prevent the consequences from arising can, in healthcare, also affect the physical and mental health of patients. One noteworthy future challenge is the potential use of quantum computers, which will make a majority of the past encryption and security schemes obsolete through the exponential increase in computational power. However, new quantum-proof defences may themselves be vulnerable to conventional cyberattacks.[14] In this sense, cybersecurity is a necessity to prevent the most trivial attacks and measures – placing a financial burden on manufacturers and healthcare authorities. Only rigorous testing can reveal weaknesses and potential new attack venues, which is why cybersecurity research is essential. Just because something has not yet been discovered in practice does not mean it cannot happen or has not already happened.

3. AI MEDICAL DEVICES

AI is defined in a plethora of ways. The AI Act proposes a legal definition premised on machine learning (ML), type of system and the incorporated statistical models.[15] This is fascinating, as some of the AI which are ML or purely statistics-based will not *per se* be considered AI by developers.[16] Regulating software or code that can be considered everything from classical

'Cybersecurity of Medical Devices: Regulatory Challenges in the European Union' in Carmel Shachar and others (eds), *The Future of Medical Device Regulation: Innovation and Protection* (Cambridge University Press 2022). Therefore, the list should be regarded as non-exhaustive.

[12] On the uptake of medical devices, IoT and AI, see eg, Roman V Yampolskiy, 'AI Is the Future of Cybersecurity, for Better and for Worse' (*Harvard Business Review*, 8 May 2017) <https://hbr.org/2017/05/ai-is-the-future-of-cybersecurity-for-better-and-for-worse> accessed 29 April 2024.

[13] Ross Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems* (3rd edn, John Wiley & Sons 2020) 15.

[14] On security and quantum computers, see Sofia Celi, 'The Quantum Solace and Spectre' (*The Cloudflare Blog*, 21 February 2022) <http://blog.cloudflare.com/quantum-solace-and-spectre/> accessed 29 April 2024 (describing how companies like Cloudflare have already started developing relatively future proof techniques against it); Wouter Castryck and Thomas Decru, 'An Efficient Key Recovery Attack on SIDH' (2022) Cryptology ePrint Archive Paper 2022/975 <https://eprint.iacr.org/2022/975.pdf> accessed 29 April 2024.

[15] AI Act (n 9) art 3(1) which defines an AI system as a 'machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.'

[16] See Peter Savadjiev and others, 'Demystification of AI-Driven Medical Image Interpretation: Past, Present and Future' (2019) 29 European Radiology 1616, 1616–18.

decision-making systems to very simple statistical analysis may be a competitive advantage for the EU, as it allows for tighter regulation than the general rules that apply.[17] Due to the network effects of strong European regulation, the AI Act would set the global standard for AI.[18] It would also impact cybersecurity, as the AI Act has a cybersecurity reference in Article 15, while other types of product legislation may not. It seems likely that tighter regulation will lead to less risk for the patient or others on whom AI is used, despite the patient not being the target of the MDR or the AI Act, as the security of AI would then be regulated literally and not through guidance.

The potential damage caused by breaches in cybersecurity on AI should dictate how the lifecycle of the systems is regulated. AI medical devices are an interesting case because of how they interface with patients and users. One can consider by analogy AI that is part of, or steers, industrial control systems or cyber-physical systems being regulated via the AI Act in a similar way to AI medical devices. AI that controls such systems, which do not always put humans at risk in such a bodily invasive manner as some medical devices, such as surgical robots and implantable medical devices, do, is wholly different. What makes AI medical devices unique is the kind of *risks* they pose, as the models and methods used to create and control them are not unique or novel. Examples of these physical risks could be physical harm to the patient, incorrect prescriptions and improper choice of surgery.[19] These failures in cybersecurity make using such medical devices more dangerous than if they possessed no hardware or software, and AI medical devices go further than this. Instead of just attacking a local network or device, the attacker can cause damage to the model, or the AI service used by the medical devices, causing damage at a larger scale instead of locally.

The Medical Device Coordination Group (MDCG) provides an overview of some severe but likely common types of cyberattacks and their consequences on patients in Annex II of their MDCG Guidance on Cybersecurity.[20] Many of these will apply directly to AI medical devices. Additionally, privacy risks could include leakage of special categories of personal data or profiles necessary for ML, which AI has created, and specific information about movement or whereabouts.[21] Physical risks can be due to indirect consequences, such as the wrong

[17] Instead of the ‘Brussels effect’, AI will be developed with EU sales in mind on a worldwide basis. For more, see eg, Lee A Bygrave, ‘The “Strasbourg Effect” on Data Protection in Light of the “Brussels Effect”: Logic, Mechanics and Prospects’ (2021) 40 Computer Law & Security Review 1.

[18] See eg, Michal S Gal and Oshrit Aviv, ‘The Competitive Effects of the GDPR’ (2020) 16 Journal of Competition Law and Economics 349.

[19] Assuming AI will assist with telerobotic surgery, see eg, Homa Alemzadeh and others, ‘Targeted Attacks on Teleoperated Surgical Robots: Dynamic Model-Based Detection and Mitigation’ (2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Toulouse, October 2016).

[20] See MDCG Guidance (n 5). For more information on the effect of the guidance, see Kaspar Rosager Ludvigsen and Shishir Nagaraja, ‘Dissecting Liabilities in Adversarial Surgical Robot Failures: A National (Danish) and EU Law Perspective’ (2022) 44 Computer Law and Security Review 1.

[21] On leakage of profiles necessary for ML, see Xinlei Pan and others, ‘How You Act Tells a Lot: Privacy-Leaking Attack on Deep Reinforcement Learning’ (Proceedings of the 18th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), Montreal, May 2019); leakage of information about movement and whereabouts is increasingly

calibration of instruments of a surgical robot or medicine recommender systems, or other types of communication disruption which may cause indirect injuries.[22] Essentially, normal accidents or non-adversarial risks exist in the same way for AI medical devices as they do for medical devices without them, but AI has an additional layer of complexity in the form of its coding and potentially ML, and because of its digital presence in the form of cybersecurity.[23]

4. SPECIFIC CYBERSECURITY ISSUES

AI presents specific risks within cybersecurity. One risk beyond the scope of this chapter is that of AI malfunctioning or making a wrong decision. In the ML context, this is often caused by an ‘adversarial sample/examples’, popularised through road signs that a car’s vision would misread or by causing a system to misidentify pictures.[24] These cause AI to fail, and while theoretically common, they do not represent the complexity that our three new concerns possess.

The three adversarial cyberattacks discussed below are not exhaustive and present overarching categories in themselves.[25] Poisoning is known as the biggest threat to ML-based systems, and if many types of AI deployed in healthcare use such decision-making mechanisms, examining this specific vulnerability is paramount. Social Engineering is the most common type and covers everything from shoulder-surfing to phishing, attacks that everyone should know about and whose risks can be mitigated through behavioural and organisational training. Extraction is another unique ML vulnerability, which leaves both the source code and training data vulnerable – with the latter causing both practical and data protection issues if leaked.

4.1 Poisoning of Datasets

If the AI medical device uses ML, it will use datasets to create its classifiers to make decisions. This can be further complicated with neural networks, but what matters is the distinction between whether it uses such datasets or not. If it does, it is vulnerable to adversarial cyber-

becoming a danger, and would lead to same risks as general stalkerware, see Cynthia Khoo, Kate Robertson and Ronald Deibert, ‘Installing Fear: A Canadian Legal and Policy Analysis of Using, Developing, and Selling Smartphone Spyware and Stalkerware Applications’ (*The Citizen Lab*, 12 June 2019) <https://citizenlab.ca/2019/06/installing-fear-a-canadian-legal-and-policy-analysis-of-using-developing-and-selling-smartphone-spyware-and-stalkerware-applications/> accessed 30 April 2024.

[22] For a discussion on how to prove and understand injuries caused by adversarial cyberattacks, indirectly or directly, see Ludvigsen and Nagaraja (n 20).

[23] From a terminological point of view, ‘adversarial’ can be translated to opponent, and refers to actors who want to attack a given software or hardware system. ‘Non-adversarial’ refers to failures that occur without them.

[24] On adversarial examples, see Richard Tomsett and others, ‘Why the Failure? How Adversarial Examples Can Provide Insights for Interpretable Machine Learning’ (2018 21st International Conference on Information Fusion (FUSION), Cambridge, September 2018); see also Dawei Zhou and others, ‘Removing Adversarial Noise in Class Activation Feature Space’ (2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, October 2021).

[25] Other authors use four categories, but these do not reflect the most common types of attacks, as these are always present, see Anderson (n 13) 876–77.

attacks, which can poison these datasets.[26] This is usually in the form of manipulation, where the data is changed and will be reflected in the decisions of the AI, but may take other forms such as inserting random numbers or deleting parts of it instead. Constructing and mounting generic ML-based attacks related to medical devices was shown to be possible as far back as 2015, when Mozaffari-Kermani et al. experimentally showed that generic poisoning, in the form of purely adding to datasets, either derived or known, can be carried out on a wide range of systems.[27] Most interestingly, the experiments were done on datasets used in healthcare settings, which currently but especially in the future will include ML-based systems in various forms. The attack consists of two algorithms: the first describes the insertion, while the second dictates the design of malicious insertions. As with all ML attacks, deception is key, and the effects of the attack should be shown in the outcome. In this case the accuracy of ML models decreased by as much as 24 per cent, which is likely to be replicated in a real-world setting.

At best, the attack causes no injury, and the AI merely malfunctions or does nothing. In the worst case, it could harm the patient. This can occur in very subtle ways that may not be detectable or in an obvious manner that will make the AI malfunction or otherwise fail.[28] Currently, any software using learning or datasets in general to function will be vulnerable to this attack.[29] Unlike simpler systems, AI that uses learning is wholly reliant on this data being usable and worthwhile unless a human fully controls it. However, at that point, it may not qualify as AI in the first place.[30] While there are no known examples of this attack being successful, it is widely regarded as the most significant vulnerability for ML-based systems.[31]

4.2 Social Engineering

Social engineering involves attacking a system either purely with human attributes or as the first step in a series of attacks that use other techniques after one has penetrated the defences. Ontologically, social engineering attacks have six factors: target, compliance principles,

[26] For a general overview, see Matthew Jagielski and others, ‘Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning’ (2018 IEEE Symposium on Security and Privacy, San Francisco, May 2018).

[27] Mehran Mozaffari-Kermani and others, ‘Systematic Poisoning Attacks on and Defenses for Machine Learning in Healthcare’ (2015) 19 IEEE Journal of Biomedical and Health Informatics 1893.

[28] See Efrat Shimron and others, ‘Subtle Data Crimes: Naively Training Machine Learning Algorithms Could Lead to Overly-Optimistic Results’ (2021) <http://arxiv.org/abs/2109.08237> accessed 30 April 2024; for examples, see Sara Kaviani, Ki Jin Han and Insoo Sohn, ‘Adversarial Attacks and Defences on AI in Medical Imaging Informatics: A Survey’ (2022) Expert Systems With Applications 1.

[29] This is a feature of anything which makes use of these techniques, and there are no means to generally prevent it. See Andrew Ilyas and others, ‘Adversarial Examples Are Not Bugs, They Are Features’ (33rd Conference on Neural Information Processing Systems (NeurIPS), Vancouver, 2019).

[30] This is not to be confused with `Supervised Learning’, which will still enable AI to fulfil the legal definition in Article 3(1) of the AI Act, regardless of whether it is changed drastically.

[31] Anderson (n 13) 875–79.

techniques, goal, medium and social engineer.[32] Target is an individual or organisation, and compliance principles refer to what one is trying to abuse, be it friendship, internal order or rules. Techniques concern anything from phishing to baiting or directly cheating individuals through conversation, while goal is either financial, access or disruption. Medium is anything from shoulder-surfing to online, phone or snail mail, while social engineer concerns whether the attack is by an individual or a group.

Social engineering is a general threat to most systems, either through phishing, guessing passwords or other means to access systems.[33] In recent years, the focus has been on general or spear-phishing attacks, which continue to occur in massive numbers, and involve mimicking the accounts of senior staff or others, to persuade victims to either click links, or transfer funds or documents to the adversary.[34] The technique enabling trustworthy-looking emails or messages is sometimes referred to as spoofing, as the receiver must believe that the email is from the right server and the right domain.[35] They do so to identify what is required to commit to the spear-fishing attack – in this case, whether the targeted mail server could tell the real domain from a fake with one which it failed to do.

There have been many successful social engineering attacks on healthcare providers and companies caused by phishing, but of particular note is one that successfully hit Magellan Incorporated in 2020.[36] The number of affected individuals was significant, and the attackers had the opportunity to not only deploy ransomware but also take control of any deployed software – such as medical devices – run or supported by AI, or even poison datasets or manipulate the ML models. There is no report indicating that this was the case, but it remains one of the more serious breaches due to the unique services that Magellan offers, through IoT or CPS – the latter specifically to surgical robots. The number of individuals who can be affected by privacy breaches only grows, and so does the potential of these rather simple attacks, if more sophisticated or widespread technology is deployed by these same entities.[37] What is vital here are the risks associated with access to such AI through these attacks. For example, through social engineering an adversary could get access to the system with the intent to harm a specific patient or a group. These attacks can be harmful enough via financial or reputational

[32] See figure 2 in Francois Mouton and others, 'Social Engineering Attack Framework' (2014 Information Security for South Africa, Johannesburg, August 2014).

[33] Jan-Willem Bullée and Marianne Junger, 'Social Engineering' in Thomas J Holt and Adam M Bossler (eds), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (Palgrave Macmillan 2020).

[34] Concerning spear-phishing, there exists a myriad of good human behaviour-focused literature, as well as technical, see eg, Yuosuf Al-Hamar and others, 'Enterprise Credential Spear-Phishing Attack Detection' (2021) 94 Computers & Electrical Engineering 1; see also Hossein Abroshan and others, 'COVID-19 and Phishing: Effects of Human Emotions, Behavior, and Demographics on the Success of Phishing Attempts During the Pandemic' (2021) 9 IEEE Access 121916.

[35] Al-Hamar and others (n 34).

[36] See Steve Alder, 'Healthcare Data Breaches Due to Phishing' (*The HIPAA Journal*, 6 January 2024) <https://www.hipaajournal.com/healthcare-data-breaches-due-to-phishing/> accessed 30 April 2024; Sergiu Galtan, 'Healthcare Giant Magellan Health Hit by Ransomware Attack' (*Bleeping Computer*, 12 May 2020) <https://www.bleepingcomputer.com/news/security/healthcare-giant-magellan-health-hit-by-ransomware-attack/> accessed 30 April 2024.

[37] Successful adversarial attacks like these focus on leeching information and reselling it at a later point.

losses, and unless the AI can control what harm humans directly or indirectly suffer, the risk from these types of attacks persists. Mitigation exists but relies on the very humans in the loop within the system where the AI resides.[38] These attacks are often used to either escalate and gain control of the system or perhaps enable the other two examples of attack examined below.[39]

4.3 Extraction of Data or Source Code

AI medical devices make decisions. This is a core reason to use them, and it is one of the advantages they have over non-deciding systems. But with this comes the unique weakness that there exists a myriad of adversarial attacks which can lure out the classifiers or the very core of the AI.[40] The adversary asks or tricks the AI into revealing its classifiers, or reverse-engineers the model. These attacks can be black-box (no knowledge of the model) or white-box (partial knowledge of the model), illustrated by the two classic papers by Tramer et al.[41] In the black-box attack, analysis of what information the providers of ML provide is used to predict what the model does with the data through queries, particularly through what the queries indirectly or inversely say about the model.[42] This is then used to reconstruct a model which provides the same functionality as the models that the attacker does not know, through depictions of the decision tree, leaves of the decision tree, categories, labels and other relevant information. The white-box paper goes further, and while it requires some knowledge of the models, it discusses the idea of transferable extraction or otherwise adversarial example attacks, to cause the same consequences as the past paper, but on a myriad of different models.[43] Surprisingly, it finds proof of transferability, but also the opposite – that there are limits to how easily attacks that work on one type can be transferred to others. Through the extraction of either data or the algorithms, there is, at the very minimum, a privacy risk or other potential physical threats, as the data could contain information about the patient. The other special risk comes from re-building of the model, as the adversary can use the same techniques to recreate the targeted AI and either uses this information to find other vulnerabilities or simply sells it to a competitor of the manufacturer. Unlike social engineering attacks, there are no major

[38] For an overview in general, which will apply to AI directly, see Walter Fuertes and others, ‘Impact of Social Engineering Attacks: A Literature Review’ in Álvaro Rocha, Carlos Hernan Fajardo-Toro and José María Riola Rodríguez (eds), *Developments and Advances in Defense and Security* (Springer Singapore 2022).

[39] Kang Leng Chiew, Kelvin Sheng Chek Yong and Choon Lin Tan, ‘A Survey of Phishing Attacks: Their Types, Vectors and Technical Approaches’ (2018) 106 Expert Systems with Applications 1.

[40] Nicolas Papernot and others, ‘Practical Black-Box Attacks Against Machine Learning’ (Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security, Abu Dhabi, April 2017).

[41] Florian Tramer and others, ‘Stealing Machine Learning Models via Prediction APIs’ (Proceedings of the 25th USENIX Security Symposium, Austin, August 2016); Florian Tramer and others, ‘The Space of Transferable Adversarial Examples’ (2017) <http://arxiv.org/abs/1704.03453> accessed 30 April 2024.

[42] Tramer and others, ‘Stealing Machine Learning Models via Prediction APIs’ (n 41).

[43] Tramer and others, ‘The Space of Transferable Adversarial Examples’ (n 41).

mitigation measures, and the risks will always exist if the AI is built to respond to inputs.[44] Indeed, it is through these inputs that the attacks happen, and it is a systemic vulnerability of any ML-based system. Finally, there is also a unique immaterial risk, as competitors of AI manufacturers could use these same attacks to gain a competitive advantage, making the type of adversaries more diversified than in the other examples. There are no known examples of these attacks, but they remain a serious threat for the reasons given above.

5. THE EU LEGAL FRAMEWORK RELEVANT FOR THE CYBERSECURITY OF AI MEDICAL DEVICES

The EU legal framework, having provisions relevant to the cybersecurity of medical devices, consists of the Regulation (EU) 2017/745 on medical devices (MDR),[45] Directive (EU) 2016/1148 concerning measures for a high common level of security of network and information systems across the Union (NIS Directive),[46] Regulation (EU) 2019/881 on ENISA and on information and communications technology cybersecurity certification (Cybersecurity Act)[47] and Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR).[48] Additionally, there is the AI Act[49] and the proposed NIS 2 Directive.[50] For each law, the relevant provisions concerning cybersecurity of medical devices are examined below.

5.1 Current Laws

The Cybersecurity Act aims to ensure proper functioning of the EU internal market by targeting a high level of cybersecurity of network and information systems, communication networks, services and devices within the Union. One of the Act's objectives is to create a new framework for European Cybersecurity Certificates of ICT products, processes and services. Certification schemes established under the Act's framework are voluntary, and vendors can decide whether to certify their products. The Act mainly impacts AI medical device manufacturers since these devices may fall under the definition of an ICT product, being an 'element […] of a network or information systems', which consequently implies the application of the NIS Directive.[51] Healthcare providers may also fall within the scope of the Act since they use ICT processes or ICT services to carry out their activities.[52] To obtain cybersecurity certifica-

[44] Alice Hutchings, Sergio Pastrana and Richard Clayton, 'Displacing Big Data: How Criminals Cheat the System' in Rutger Leukfeldt and Thomas J Holt (eds), *The Human Factor of Cybercrime* (1st edn, Routledge 2019).
[45] MDR (n 5).
[46] NISD (n 6).
[47] CSA (n 7).
[48] GDPR (n 8).
[49] AI Act (n 9).
[50] NIS 2 Directive (n 10).
[51] CSA (n 7) art 2(12).
[52] Furthermore, see ibid art 56(3) which sets healthcare as a focus priority for the European Commission.

tion, manufacturers or healthcare providers may voluntarily (when not prescribed by national or EU law) apply to the conformity assessment bodies of their choice established in the EU. Such a body produces a formal evaluation of the device against a defined set of criteria and standards, and issues a certificate that should assure users and maintain trust and security of the device.

The NIS Directive is relevant to cybersecurity considerations concerning network and information systems. Specifically, the Directive lays down obligations for EU Member States to reach a minimum harmonisation level across the EU concerning network and information systems' security. Network and information systems are defined as 'any device or group of interconnected or related devices, one or more of which, pursuant to a program, perform automatic processing of digital data'.[53] Applied to the healthcare sector, this definition of network and information systems may encompass hospitals' IT networks and AI medical devices. The NIS Directive sets security and notification requirements for operators of essential services (OES). Healthcare providers are concerned by the NIS Directive when identified as OES by the respective Member State. As OES, healthcare providers must ensure a minimum level of security for their networks and information systems and must notify security incidents to competent authorities without delay. To reach that level, they must take appropriate and proportionate technical and organisational measures to manage the risk posed to the NIS security, which they use in their operations. Security measures should be taken to prevent and minimise the impact of incidents affecting the security of the NIS used for the provision of essential services to ensure the continuity of those services.[54] These measures and modalities for communication of security incidents are defined at a national level by each Member State, which is required to adopt national strategies on network and information security.[55]

The MDR sets general requirements that may implicate cybersecurity considerations. They primarily address manufacturers of medical devices.[56] Article 5(1) MDR obliges manufacturers to ensure that the device complies with the MDR obligations when used following its intended purpose. According to Article 5(2) MDR, a medical device must meet the general safety and performance requirements set out in Annex I MDR, taking into account the intended purpose. As part of the general requirements set in Annex I MDR, 'devices shall achieve the performance intended by the manufacturer' and be designed in a way suitable for the intended use.[57] They shall be safe and effective, and associated risks shall be acceptable when weighed against the benefits to the patients and the level of protection of health and safety while taking into account the state of the art.[58] Moreover, 'manufacturers shall establish, implement, document,

[53] NISD (n 6) art 4.

[54] See NIS Cooperation Group, 'Reference Document on Security Measures for Operators of Essential Services' (2018) 4.

[55] NISD (n 6) art 1. According to the NISD, a national strategy is a 'framework providing strategic objectives and priorities on the security of network and information systems at national level' (NISD, art 4).

[56] MDR (n 5) art 2(30) defines a manufacturer as 'the natural or legal person who manufactures or fully refurbishes a device or has a device designed, manufactured, or fully refurbished and markets that device under its name or trademark'.

[57] ibid Annex I req 1.

[58] ibid. The intended purpose is defined in Article 2(12) of the MDR as 'the use for which a device is intended according to the data supplied by the manufacturer on the label, in the instruc-

and maintain a risk management system'.[59] Part of this system also includes risk control measures to be adopted by manufacturers for the design and manufacture of a device, and they shall conform to safety principles and state of the art.[60] A medical device designed to be used with other devices or equipment as a whole (including the connection system between them) must be safe and must not impair the specified performance of the device.[61] Furthermore, a medical device shall be designed and manufactured to remove, as far as possible, risks associated with possible negative interaction between software and the IT environment within which it operates.[62] If a medical device is intended to be used with another device, it shall be designed, so that 互操作性和相容性 interoperability and compatibility are reliable and safe.[63] A medical device incorporating electronic programmable systems, including software or standalone software as a medical device, 'shall be designed to ensure repeatability, reliability, and performance according to the intended use'.[64] '[A]ppropriate means shall be adopted to […] reduce […] risks or impairment of the performance.'[65] A medical device should be developed and manufactured according to the state of the art and should respect the development life cycle principles, risk management (including information security), verification and validation.[66] Finally, manufacturers must 'set out minimum requirements concerning hardware, IT network characteristics and IT security measures, including protection against unauthorized access'.[67] Concerning information to be supplied together with the device, manufacturers must inform about residual risks, provide warnings requiring immediate attention on the label and, for electronic programmable system devices, give information about minimum requirements concerning hardware, IT networks' characteristics and IT security measures (including protection against unauthorised access) necessary to run the software as intended.[68]

The GDPR is relevant to AI medical devices' cybersecurity because these devices function by processing a vast amount of personal (and non-personal) data. The Regulation, thus, lays down rules for protecting natural persons regarding the processing of their personal and sensitive data, including data concerning health. Among the many requirements, the GDPR sets obligations to ensure the security of the processing of data which takes place in the lifecycle of AI medical devices.[69] Parties involved in the processing of personal data must put in place technical and organisational measures that are adequate to the risk of processing. Security measures and obligations include the performance of risk assessments and the notification of a personal data breach to competent authorities. In the case of AI medical devices, security

tions for use or in promotional or sales materials or statements and as specified by the manufacturer in the clinical evaluation'.

[59] ibid Annex I req 3.

[60] ibid Annex I req 4.

[61] ibid Annex I req 14.1.

[62] ibid Annex I req 14.2(d).

[63] ibid Annex I req 14.5.

[64] ibid Annex I req 17.1.

[65] ibid.

[66] ibid Annex I req 17.2.

[67] ibid Annex I req 17.4.

[68] ibid Annex I see reqs 23.1.(g), 23.2.(m), 23.4(ab).

[69] GDPR (n 8) arts 5 and 32.

measures may concern healthcare providers, healthcare professionals and manufacturers of medical devices, among others.

5.2 New Laws

The AI Act may impact the cybersecurity of AI medical devices in four ways. The AI Act contains provisions prohibiting certain AI systems and practices, proposes a risk-based mechanism for governing those AI systems that pose a high risk for individuals or society, stipulates fines for providers' non-compliance with the Act and establishes an EU body responsible for the harmonised application of the Act among the Member States. The AI Act explicitly includes medical devices in the scope of its application in Article 6 and Annex I. The definitions of AI and risk classification provided therein imply that any medical device software could fall within the scope of the AI Act and be considered a high-risk AI system since most medical device software needs a conformity assessment by a notified body.[70] Consequently, this could imply parallel application of the two pieces of legislation and pose additional challenges for those implementing them (further elaborated in the following section). Article 13(1) of the AI Act concerns the cybersecurity of AI medical devices. The article requires that the design and development of high-risk AI systems is done in a way that ensures their transparent operation so deployers can interpret the system's output and use it appropriately. They should also be designed and developed to achieve an appropriate level of accuracy, robustness and cybersecurity, and perform consistently throughout their lifecycle.[71] In the instructions for use,[72] providers shall specify the level against which cybersecurity of the system has been tested and validated and which can be expected, and any known and foreseeable circumstances that may impact that level of cybersecurity. Article 15(4) of the AI Act requires that the technical solutions aimed at ensuring the cybersecurity of high-risk AI systems are appropriate to the relevant circumstances and the risks. To this end, high-risk AI systems certified according to the Cybersecurity Act shall be presumed to comply with the cybersecurity requirements set out in the AI Act.[73]

The proposed NIS 2 Directive is set to reform the currently applicable NIS Directive due to the varying level of harmonisation of the NIS Directive among the EU Member States. Concerning cybersecurity of AI medical devices, and in comparison with the applicable NIS Directive, the proposal removes the Member States' requirement to identify OES and Digital Service Providers (DSP) in their territories. The proposal also replaces OES and DSPs with new categories: 'essential' and 'important' entities (enlisted in Annexes I and II of the proposal). They are ordered per sector and sub-sector (for example, 'health' and 'manufacturing' sectors; 'manufacture of medical devices and in-vitro medical diagnostic medical devices' sub-sectors). Every sub-sector contains a list of 'types of entities'. The proposal broadens its scope of application. For instance, healthcare providers are now considered 'essential entities'

[70] MedTech Europe, 'Proposal for an Artificial Intelligence Act (COM/2021/206) - MedTech Europe response to the open public consultation' (2021) 1 <https://www.medtecheurope.org/wp-content/uploads/2021/08/medtech-europe-response-to-the-open-public-consultation-on-the-proposal-for-an-artificial-intelligence-act-6-august-2021-1.pdf> accessed 30 April 2024.

[71] AI Act (n 9) art 15(1).

[72] ibid arts 15(2) and 15(3).

[73] ibid art 42.

(Annex I). In addition, the proposal adds new types of entities relevant to the healthcare sector.[74] Like the NIS Directive, the proposal mandates the Member States to establish security measures for the entities under their scope. Chapter IV of the proposal contains obligations on cybersecurity, risk management and reporting. Article 18 of the proposal on cybersecurity risk management measures states that essential and important entities shall 'take appropriate and proportionate technical and organisational measures to manage the risks posed to the security of network and information system'. As examples of measures, the article includes, among others, incident handling (prevention, detection and response to incidents) and measures to ensure supply chain security and vulnerability handling and disclosure. Article 20 of the proposal on reporting obligations introduces a two-step procedure to report significant security breaches, which could also be reported to the recipients of the services. Article 21 of the proposal concerns cybersecurity certification schemes. Enforcement and supervision of essential and important entities are delegated to competent authorities. Competent authorities shall supervise them and ensure their compliance with the security and incident notification requirements. An ex-ante supervisory regime is in place for essential entities and an ex-post one for important entities.[75]

6. CURRENT CHALLENGES IN THE REGULATORY FRAMEWORK

6.1 The AI Act and Its Interaction with the MDR: Cybersecurity and Safety Requirements

Both the MDR and the AI Act requirements are relevant from a cybersecurity perspective. To demonstrate this, we propose the example of an anaesthesia device.[76] In this example, an unauthorised user with physical access to an anaesthesia device, thanks to social engineering techniques, guesses the weak password of the device and manipulates its settings so that the machine may supply a wrong anaesthetic concentration.

Following this case, both the MDR and the AI Act could be relevant. It could be a serious incident under the MDR, as the wrong anaesthetic concentration might (even indirectly) lead to the serious deterioration of a patient's health. For the same reason, the event could also be relevant under the AI Act.

In some circumstances, the AI Act requires providers of high-risk AI systems to report serious incidents to the relevant authority.[77] Similarly, the MDR requires the reporting of

[74] For further analysis, see Biasin and Kamenjašević (n 3).

[75] NIS 2 Directive (n 10) 27.

[76] See also MDCG Guidance (n 5) 40–41.

[77] See AI Act (n 9) art 73. In the AI Act, serious incidents are any 'incident or malfunctioning of an AI system that directly or indirectly leads to any of the following: (a) the death of a person, or serious harm to a person's health; (b) a serious and irreversible disruption of the management or operation of critical infrastructure; (c) the infringement of obligations under Union law intended to protect fundamental rights; (d) serious harm to property or the environment' (AI Act, art 3(49)).

serious incidents to the medical devices competent authorities.[78] A consequent first challenge concerns the overlapping of cybersecurity requirements between the MDR and the AI Act, particularly on serious incident notification. This overlapping is not necessarily a problem *per se*. Nevertheless, it remains essential to understand the interaction between the safety requirements of the AI Act with the MDR and whether the MDR should be considered as a *lex specialis* to the AI Act in that circumstance.

The AI Act and the MDR present similarities but also divergences concerning serious incidents. Elsewhere, we demonstrated that the definition and some other aspects concerning serious incident reporting might differ.[79] For example, the regulated entities are different. In the first case, the AI Act covers high-risk systems providers, while the MDR concerns medical device manufacturers. These may or may not overlap. Manufacturers of medical devices can also be providers if their product is software or hardware with software, but they may also license or buy these parts separately, which would not make them providers under the AI Act.

The AI Act states that the 'complementarity between this Regulation and existing sectoral Union law should also be taken into account in future standardisation activities or guidance adopted by the Commission'.[80] This formulation lacks specificity because it is unclear how this 'standardisation' will happen between the different laws.

In conclusion, a possible challenge for the AI Act is understanding if and how the safety requirements will be integrated into medical device sector-specific legislation. The references provided in the AI Act seem not to be articulated enough to help understand how safety requirements will interact with the MDR. If these integration aspects are not addressed, the lack of a coordinated framework could lead to regulatory uncertainty.[81]

6.2 The NIS 2 Directive and Its Interaction with the MDR: Incident Notification Requirements

The NIS 2 Directive and the MDR also include relevant requirements from a cybersecurity perspective. Serious incidents under the MDR may also qualify as incidents under the NIS 2 Directive. To demonstrate, we offer an example of an implantable sensor used to monitor pulmonary artery pressures in heart failure patients.[82] In our example, an adversary modifies through data poisoning or creates patient data in transit to or from the external electronics unit, causing misdiagnosis that affects patient care. Due to this security breach, the physician could fail to provide the treatment based on incorrect low pulmonary artery pressure readings. This could lead to the worsening of the patient's heart failure condition.[83] In this case, worsening a patient's heart failure condition could be seen as a deterioration of their health caused by the event initiated by the attacker (relevant under the MDR's serious incident requirements). The

[78] See MDR (n 5) art 87. In the MDR, serious incidents are 'any incident that directly or indirectly led, might have led or might lead to […] the death of the patient, user or other person, the temporary or permanent serious deterioration of a patient's, user's or other person's state of health, a serious public threat' (MDR, art 2(65)).

[79] Biasin and Kamenjašević (n 3) 171–74.

[80] AI Act (n 9) Recital 81.

[81] ibid.

[82] The example is taken from the MDCG Guidance (n 5) 40.

[83] ibid.

misdiagnosis of patient care could also be considered as impacting the provision of healthcare services, thus relevant from the NIS 2 Directive's perspective.

As illustrated above, the MDR sets obligations concerning serious incident reporting. The NIS 2 Directive requires that any incident significantly impacting the provision of essential or important entities' services be notified to the national competent authority or CSIRT.[84]

The NIS 2 Directive entails similar challenges in terms of regulatory convergence with the MDR.[85] The situation described above could imply the application of both the MDR and NIS 2 Directive requirements. Such a parallel application of requirements is where the challenges might begin. In the past, the European Commission assessed the possible 'synergies' of the NIS Directive with sector-specific regulation, including in the medical devices field concerning incident notification reporting.[86] The actual text of the NIS 2 Directive suggests that where provisions of sector-specific acts of Union law require medical device manufacturers to notify incidents or significant cyber threats, the provisions of the NIS 2 Directive should not apply *if* those requirements are 'at least equivalent'.[87] The provision may turn problematic, for example, for manufacturers assessing which piece of legislation should apply. In this regard, it is questionable whether the MDR's safety requirements should be considered *at least equivalent* to the NIS 2 Directive's requirements concerning incident notification. In fact, safety incidents do not always equate to security incidents.[88]

We could take the example of a warming device for premature babies.[89] The MDCG provides the following example: an unauthorised user with physical access to the device guesses the weak password for the service account and exports therapy and patient data via the USB interface. In the view of the MDCG, there could be security harm (that is, the unauthorised access), which could not result in safety harm (such as the serious deterioration of a patient's health) in terms of the MDR's serious incident notification rules.[90] In other words, medical device manufacturers would have to notify the incident of the unauthorised access to the NIS 2 Directive competent authority, but not to the national relevant authority under the MDR.

Given that security incidents do not always equate to safety incidents, one may wonder which direction should be taken in interpreting the MDR and the NIS 2 Directive rules. We envisage at least three possible approaches for the legislator. A first, simplistic approach would mean considering the MDR as 'at least equivalent' to the NIS 2 Directive. In this case, the NIS 2 Directive requirements on incident notification would not apply, while the MDR

[84] For a definition of essential and important entities, see NIS 2 Directive (n 10) art 4(25)–(26). Concerning incident notification, it is worth noting that the current formulation of the proposal also includes references to 'any significant cyber threat that those entities identify that could have potentially resulted in a significant incident' (NIS 2 Directive, art 20(2)).

[85] Medical device manufacturers are subject to both obligations. For the NIS 2 Directive, it is worth noting that, as a Directive, its requirements will have to be set by national legislation.

[86] NIS Cooperation Group, 'Synergies in Cybersecurity Incident Notification Reporting' (2020).

[87] NIS 2 Directive (n 10) art 2(6).

[88] Safety will matter when cybersecurity causes injury or damage, see Anderson (n 13) 1044–45. But many situations will not involve safety, especially when the cyberattack considers privacy.

[89] MDCG Guidance (n 5) 41.

[90] ibid.

would prevail. A second approach could, on the contrary, suggest the parallel application of both legal acts and their requirements. Such a parallel application would require manufacturers to consider the different requirements and apply both. A third approach could consider the MDR as a *lex specialis*, leaving the NIS 2 Directive as possible general legislation to cover those hypotheses that are not strictly covered by the MDR but are still relevant to the NIS 2 Directive.

The solutions we propose come with advantages and disadvantages.[91] It remains essential that the legislator takes a stance and resolves this interpretative issue to ensure regulatory certainty for all stakeholders in healthcare.

6.3 The Evolving Term 'Critical Infrastructures': Serious Incidents for AI Systems in the Healthcare Sector

A further challenge within the AI Act stems from the 'serious incident' definition. Serious incidents are any incident that directly or indirectly lead to a 'serious and irreversible disruption of the management and operation of critical infrastructure'.[92]

However, the identification of critical infrastructure is delegated to the EU Member States, following the principle of subsidiarity.[93] In simple terms, the division of competencies between the EU and its Member States entails that the identification of a hospital, healthcare entity or health process as critical infrastructure shall ultimately depend on every national approach adopted in critical infrastructure regulation.[94] Until recently, the EU Member States have adopted various approaches to identify their own critical infrastructure. Usually, the

[91] For a prior analysis of the solutions, advantages and disadvantages, see also Biasin and Kamenjašević (n 3).

[92] AI Act (n 9) art 3(49)(b).

[93] From a conceptual perspective, the notion of 'critical infrastructures' has been defined as an evolving term. See Dimitra Markopoulou and Vagelis Papakonstantinou, 'The Regulatory Framework for the Protection of Critical Infrastructures against Cyberthreats: Identifying Shortcomings and Addressing Future Challenges: The Case of the Health Sector in Particular' (2021) 41 Computer Law & Security Review 1; for the purpose of this chapter, we refer to the definition of Rinaldi, Perenboom and Kelly: infrastructures are considered 'critical' when their disruption could have an impact on the functioning of the society (in terms of economy, security and people's wellbeing). See Steven M Rinaldi, James P Peerenboom and Terrence K Kelly, 'Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies' (2001) 21 IEEE Control Systems Magazine 11, 12; for an overview of the principles guiding the implementation of critical infrastructure protection, see also, Commission, 'Communication from the Commission on a European Programme for Critical Infrastructure Protection' COM (2006) 786 final, 3.

[94] There exists an EU-level definition of 'European Critical Infrastructure', which is determined in the European Critical Infrastructure Directive (ECI Directive). See Council Directive 2008/11/EC of 8 December 2008 on the Identification and Designation of European Critical Infrastructures and the Assessment of the Need to Improve their Protection [2008] OJ L345/75 (ECI Directive). Nevertheless, national critical infrastructures do not belong to the ECI Directive scope, and the Member States have autonomy in determining their approaches towards them.

trends followed by the Member States see the definition of critical infrastructure based on their defence strategies, national emergency management and long-term national traditions.[95]

Studies concerning the identification of critical infrastructure in the healthcare sector have shown how heterogeneous the situation could be. For instance, France includes the healthcare sector within the scope of critical infrastructure protection legislation, while the Netherlands has embraced a process-oriented approach that does not include healthcare assets.[96] This conceptual issue may entail some tangible consequences. We can use the example of a cyberattack carried out through social engineering techniques or data extraction on a healthcare provider. Some Member States might identify the event as an attack on critical infrastructure, while others may not because healthcare is not considered a critical infrastructure sector. In the first scenario the healthcare provider would possibly consider the event as a serious incident and thus activate the obligation of its reporting under the AI Act. The second case would be unlikely to activate such an obligation.[97]

AI systems are used in healthcare critical infrastructure. If disrupted by a cyberattack, the provision of healthcare services to individuals may affect the continuity of their services and, ultimately, the quality of healthcare systems. The Member States not considering healthcare within the sectors or processes of critical infrastructure may lower the level of protection of individuals, compared to the Member States who do, from the perspective of the AI Act incident notification rules. In the ultimate analysis, this kind of heterogeneous level of individuals' protection across the EU may be seen as a possible regulatory challenge, which could lead to fragmentation risks in the EU internal market.[98]

As a possible step forward, these uneven impacts could be mitigated by shifting the focus of Article 3 AI Act from 'critical infrastructures' to 'critical entities'.[99] In fact, in recent years, the many issues surrounding the European legislation concerning the protection of critical infrastructures led the European legislator to revise the existing legislation concerning the European Critical Infrastructures to issue a new proposed Directive on 'critical entities'. While critical infrastructures do not meet the same criteria in terms of sectors at the national level, as outlined above, the CER Directive proposal could fill this gap because it includes the health sector within its scope.[100] In other words, the CER Directive, once approved, will enlist the

[95] Elisabetta Biasin and others, 'SAFECARE D3.9 – Analysis of Ethics, Privacy, and Confidentiality Constraints' (2019) 48.

[96] ibid 49; see also Ordonnance no 2004-1374 du 20 décembre 2004 Code de la Défense (France).

[97] Ultimately, this would also depend on whether Article 3(49)(a) of the AI Act is met.

[98] Biasin and Kamenjašević (n 3) 177–78.

[99] Like the NIS Directive, the CER Directive proposes Member States identify critical entities based on common criteria for national risk assessments. For the definition of critical entities, see Commission, 'Proposal for a Directive of the European Parliament and of the Council on the Resilience of Critical Entities' COM (2020) 829 final (CER Directive proposal), art 2(1).

[100] See ibid Annex. There could be counterarguments, however, against this proposal, which, for reasons of space, are not possible to analyse with due detail in this chapter. Further discussions about this proposal may question the more expanded scope of critical entities' sectors against critical infrastructures following the ECI Directive. We limit ourselves to observe that this argument may be countered by observing that the scope of critical infrastructure protection can be nevertheless expanded by the Member States at the national level. Comparison and illustration of the differences between critical entities and critical infrastructures would also require more space.

sectors under which the Member States will have to identify the critical entities where healthcare is part of them.[101] Thus, the Directive's text would mitigate the risk of not considering healthcare as part of the scope of protection. In conclusion, considering 'critical entities' instead of 'critical infrastructure' could minimise the problem of uneven consideration of the healthcare sector for serious incidents concerning AI systems in the EU.

7. CONCLUSION

This chapter has examined essential aspects concerning cybersecurity of AI medical devices from integrated security and legal perspectives. Section 2 of this chapter explained the continuous evolution of cybersecurity threats. To underscore the problem's relevance, we selected three examples: dataset poisoning, social engineering and data or source code extraction. We then integrated the security analysis with the description of the legal aspects surrounding medical device cybersecurity. We analysed the MDR, the NIS Directive, and the Cybersecurity Act. Regulatory challenges arise as the legal framework concerning AI and cybersecurity evolves. Therefore, we illustrated some of the core challenges of two new laws: the proposed NIS 2 Directive and the AI Act. The AI and cybersecurity regulations are recent matters of regulation, and future research will have to closely monitor their future regulatory challenges.

Critical Infrastructure
Refers to the systems, assets, and facilities that are essential for the functioning of a society and its economy. This includes both physical systems (like transportation and utilities) and digital systems (like cybersecurity and data management).

Critical Entities
This broader term can encompass not only the physical infrastructure but also the organizations and institutions that provide essential services, governance, and support to society. This can include government agencies, private companies, non-profit organizations, and any entities deemed essential for maintaining societal functions. The emphasis here is on the role and importance of the organization rather than just the physical infrastructure.

Thus, we refer to the historical and conceptual analysis by Markopoulou and Papakonstantinou (n 93).

[101] The identification of critical entities would have to follow specific criteria and procedures, see CER Directive proposal (n 99) art 5.