




**D10.4**

**Best Practices and Policy  
Development Guidelines for  
Replicability and Wider Use**

<b>Project number:</b>	833683
<b>Project acronym:</b>	CyberSANE
<b>Project title:</b>	Cyber Security Incident Handling, Warning and Response System for the European Critical Infrastructures
<b>Start date of the project:</b>	1 <sup>st</sup> September, 2019
<b>Duration:</b>	36 months
	H2020-SU-ICT-2018

<b>Deliverable type:</b>	Report
<b>Deliverable reference number:</b>	DS-01-833683 / D10.4 / N.1
<b>Work package contributing to the deliverable:</b>	WP 10
<b>Due date:</b>	31 August 2022 (M36)
<b>Actual submission date:</b>	M36

<b>Responsible organisation:</b>	KU Leuven (KUL)
<b>Editor:</b>	Burcu Yasar (KU Leuven), Bengi Zeybek (KU Leuven), Ana Maria Corrêa Marcus (KU Leuven), Halid Kayhan (KU Leuven), Anton Vedder (KU Leuven)
<b>Dissemination level:</b>	PU
<b>Revision:</b>	N.1

<b>Abstract:</b>	This deliverable provides an assessment of legal and ethical considerations within the project. It further formulates policy recommendations for public authorities dealing with regulatory aspects of the fight against both cyber-attacks, risks and threats in critical information infrastructures.
<b>Keywords:</b>	Data protection, Cybersecurity, Critical Infrastructures, Fundamental Rights, Trustworthy Artificial Intelligence
	The project CyberSANE has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833683.

**Editor**

Burcu Yasar (KU Leuven)

Bengi Zeybek (KU Leuven)

Ana Maria Corrêa Harcus (KU Leuven)

Halid Kayhan (KU Leuven)

Anton Vedder (KU Leuven)

**Contributors** (ordered according to beneficiary numbers)

Luis Landeiro Ribeiro (PDMFC)

Yuste Guillermo (ATOS)

Thanos Karantjias (MAG)

Matej Kovacic (JSI)

Sophia Karagiorgou (UBI)

Manos Athanatos (FORTH)

Eva Papadogiannaki (FORTH)

Pablo Giménez (VPF)

Diarmuid O Connor (LSE)

Robert Bordianu (LSE)

Lena Raber (KN)

Thorstenn Engmann (KN)

## **Disclaimer**

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The content of this document reflects only the author’s view – the European Commission is not responsible for any use that may be made of the information it contains. The users use the information at their sole risk and liability.



## Executive Summary

To ensure sustainable use of new technologies, it is crucial to identify potential legal and ethical issues and mitigate them as early as possible in the development stage. Having this in mind, the CyberSANE project has adopted privacy- and data-protection-by-design approach and ethics-by-design approach. Against this background, this deliverable provides an assessment of the legal and ethical considerations within the project. It demonstrates how CyberSANE has integrated the identified requirements with a reference to relevant technical and organizational measures. Furthermore, this deliverable provides policy recommendations in the area of the regulation of Artificial Intelligence (AI) in Critical Infrastructures (CIs), notification of security incidents, data and digital evidence sharing with component authorities, and freedom of expression. It highlights that it is of utmost importance to provide clear and consistent rules for emerging technologies designed to be used in critical infrastructures. Clarity and the consistent application of rules and safeguards across sectors are considered important to provide legal certainty for manufactures, eliminate the regulatory burden for both businesses and supervisory authorities, and ensure effective protection for individuals.

## Contents

<b>Executive Summary</b> .....	<b>4</b>
<b>Contents</b> .....	<b>5</b>
<b>List of Tables</b> .....	<b>6</b>
<b>Chapter 1 Introduction</b> .....	<b>7</b>
<b>Chapter 2 Legal and Ethical Evaluation &amp; Best Practices</b> .....	<b>8</b>
2.1 Characteristics of the CyberSANE system.....	8
2.2 Methodology .....	9
2.3 Ensuring Cybersecurity through CyberSANE: Legal and Ethical Assessment....	10
2.4 Best practices.....	17
<b>Chapter 3 Policy Guidelines for Replicability and Wider Use</b> .....	<b>18</b>
3.1 Leveraging AI in Critical Infrastructures (CIs): Policy considerations .....	18
3.1.1 Ethics Guidelines for Trustworthy AI: A European Approach to AI .....	19
3.2.1.1 Policy considerations.....	22
3.1.2 Proposed AI Act and the implications for critical infrastructures .....	23
3.1.2.1 Policy considerations .....	28
3.2 Information sharing: Policy considerations.....	29
3.2.1 Notification of security incidents .....	30
3.2.1.1 Notification obligations under NIS and the revised framework .....	35
3.2.1.2 Policy considerations.....	36
3.2.1.3 The Revised NIS framework - NIS 2.0 .....	37
3.2.2. Barriers for information sharing: Access to data by component authorities.....	39
3.2.2.1 Access to data by component authorities .....	39
3.2.2.1.1 Policy considerations.....	41
3.2.2.2 Cross-border access to digital evidence .....	41
3.2.2.2.1. European Union: Proposal for E-Evidence Framework.....	42
3.2.2.2.2 Council of Europe: Cybercrime Convention and the additional protocol.....	43
3.2.2.2.3 Policy considerations.....	44
3.3. Apprehensions between (cyber)security and freedom of expression: Policy considerations.....	46
3.3.1 Setting the scene: Cybersecurity and Freedom of Expression .....	47



3.3.2	Legal and Policy Framework .....	48
3.3.2.1	United Nations Soft Law Instruments: Reports of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression .....	48
3.3.2.2	Council of Europe Human Rights Framework .....	50
3.3.2.3	Policy considerations .....	53
3.4	Lessons learned derived by the CyberSANE System .....	54
<b>4</b>	<b>Conclusion .....</b>	<b>56</b>
<b>4</b>	<b>List of Abbreviations .....</b>	<b>59</b>
<b>5</b>	<b>Bibliography .....</b>	<b>62</b>

## List of Tables

Table 1	Legal and ethical assessment of the CyberSANE platform .....	17
Table 2	List of safeguards for processing personal data provided by Article 14 of the Second Additional Protocol to the Cybercrime Convention .....	44

## Chapter 1 Introduction

This deliverable aims to provide an assessment of the legal and ethical considerations within the CyberSANE project. It demonstrates how CyberSANE has integrated the legal and ethical requirements identified by the **D2.2** (Legal and Ethical Requirements), which provided the framework and requirements relevant to the CyberSANE platform. Furthermore, this deliverable aims to formulate policy recommendations for public authorities dealing with regulatory aspects of the fight against cyber-attacks, risks and threats in Critical Infrastructures (CIs). To this end, it builds on the research carried out in earlier stage of the project. It further integrates the new legislative developments and case law that were introduced after the delivery of D2.2 in M7.

To this end, the second chapter presents the methodology and provides an assessment of the legal and ethical requirements with a reference to technical and organisation measures taken by the CyberSANE Team. The third chapter provides policy considerations and recommendations in the area of the regulation of Artificial Intelligence (AI) in Critical Infrastructures, notification of security incidents, data and digital evidence sharing with component authorities, and freedom of expression. It identifies legal barriers to the harmonized application of rules to deployment of the platform across different sectors. It further reflects on policy solutions with the aim of contributing to putting in place a consistent regulatory approach to the emerging technologies, and effective safeguards.

## Chapter 2 Legal and Ethical Evaluation & Best Practices

### 2.1 Characteristics of the CyberSANE system

The CyberSANE system combines active approaches that are used to detect and analyse anomaly activities and attacks in real-time with reactive approaches that deals with the analysis of the underlying infrastructure to assess an incident in order to provide a more holistic and integrated approach to incident handling. CyberSANE seeks to enhance the cybersecurity of critical infrastructures information systems through the collection, correlation and sharing of information by multiple sources.

Briefly put, the CyberSANE system consists of five technical components to collect, compile, process and fuse attack related data from multiple perspectives. A brief description of the technical components are as follows.<sup>1</sup> The **LiveNet** (Live Security Monitoring and Analysis) component is an advanced and scalable component capable of detecting threats in real time and, in case of a declared attack, capable of mitigating the effects of an intrusion. The component gathers network traffic data (such as data coming from a smart energy meter) and compares the collected data against the expected one to detect anomalies using advanced algorithms. The **DarkNet** component integrates two tools: MEDUSA Dark Web Intelligence Solution and Event Registry. The first one provides the capability to monitor focused collected texts referring to cybersecurity activities and find out for pawned email accounts to bring forth data breached data over the dark web marketplaces. In the example of transportation use case, it provided general purpose textual analytics and reports related with port activities, illegal cyber-activities concerning port transactions and operations. EventRegistry collects, harmonizes and analyses data from various multilingual non- structured data sources. More specifically, it collects media news and blog posts to retrieve content related to cyber incidents. In this way, it helps to analyse the big picture of global malware and cybersecurity activities, by retrieving information about similar attacks that happened in the past.

The **HybridNet** (Data Fusion, Risk Evaluation and Event Management) component receives security related information from both LiveNet and DarkNet. It generates anomaly detection/ security events and report it to the security expert team employing a set of machine learning algorithms so that any countermeasures can be taken by the expert team. The algorithms have been updated on a regular basis with new attack categories to successfully detect the abnormal traffic caused by the Denial of Service (DoS) attack. In the next step, the **ShareNet** component provides the necessary threat intelligence sharing features to allow data controllers to exchange

---

<sup>1</sup> Detailed description of the components with underlying tools can be found in D9.3.

information about cyber incidents with relevant parties in compliance with the relevant data sharing agreements. It allows sharing lessons learned with internal and external stakeholders, using the capabilities of the **PrivacyNet** component. The latter provides the necessary anonymization features to share information and threat intelligence with relevant parties in a secure and privacy-friendly way.

This brief description reveals the following characteristics of the CyberSANE platform, which are important for the purposes of this deliverable. Firstly, the platform relies on processing of large scale personal data emanating from structured and unstructured sources, including IP addresses and metadata. It is therefore necessary that the system provides the privacy and security friendly functions so that the deployers of the system can comply with the applicable framework. Furthermore, CyberSANE integrates data mining and machine learning techniques such as deep learning (the so-called artificial intelligence) to make predictions about anomalies and cyber incidents. In that sense, it is a support tool that informs human decision-making in handling security threats. From an ethical point of view, it is crucial that the system addresses the AI-specific needs such as explainability and traceability, and function safely and securely. In what follows, this deliverable offers the legal and ethical assessment followed by policy recommendations having these characteristics in mind.

## 2.2 Methodology

This chapter provides an evaluation of the legal and ethical requirements identified under D2.2 Legal and Ethical Requirements and delivered in M7. For the purposes of this chapter, KU Leuven (KUL) has prepared a survey, reflecting the legal and ethical requirements listed in D2.2 and the Assessment List for Trustworthy Artificial Intelligence (ALTAI)<sup>2</sup> of the independent High-Level Expert Group on Artificial Intelligence (HLEG AI) set up by the European Commission. The latter is a practical tool that helps businesses to self-assess the trustworthiness of AI systems.<sup>3</sup> As CyberSANE involves components integrating artificial intelligence technology, this tool has been integrated in the evaluation process in order to address ethics-related concerns in accordance with the guidance of the European Commission.<sup>4</sup> Further analysis on ALTAI from a policy perspective has been provided in Chapter 3.

It should be noted, however, that the assessment provided in this chapter is not limited to the components involving artificial intelligence. In fact, the CyberSANE platform has incorporated different components and tools with various technical features and procedures. Several other new tools may be incorporated to it after the end of the CyberSANE project. This deliverable aims to provide a general assessment of the platform, with reference to certain aspects and examples of

---

<sup>2</sup> HLEG AI, The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment, 17 July 2020, available at <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence>.

<sup>3</sup> For further information on this tool, see Chapter 3.1.

<sup>4</sup> European Commission, Ethics By Design and Ethics of Use Approaches for Artificial Intelligence, 25 November 2021, available at [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence\\_he\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf).

different tools developed during the project. It does not aim to provide a detailed account of how requirements were incorporated by each tool or component.

The survey has been presented to the consortium partners in WP10 meeting dated 18 February 2022. Answers to the survey were collected by e-mail between April-June 2022. Contributions from eight partners have been received, including all of the three end-users: ATOS, FORTH, JSI, KN, LSE, MAG, VPF and PDMFC. Partners' input has been also received through regular meetings, email exchange, and partner's presentations and deliverables.

## 2.3 Ensuring Cybersecurity through CyberSANE: Legal and Ethical Assessment

The evaluation of the CyberSANE platform has been provided in the table below. On the left side of the table, the requirements have been briefly listed. On the right side of the table, an assessment of whether and how the relevant requirement was accomplished by CyberSANE is provided. The requirements have been divided in seven thematic areas established by the HLEG AI:

- 1) Data protection and governance
- 2) Human agency and oversight
- 3) Technical robustness and safety
- 4) Transparency
- 5) Societal wellbeing
- 6) Diversity, non-discrimination and fairness
- 7) Accountability

Sub-requirements are indicated under these general requirements in order to provide a more specific assessment of a relevant criteria or aspect. More detailed description of each requirement can be found in D2.2 Legal and Ethical Requirements.

	Requirement	Assessment
1	<b>Data protection &amp; governance</b>	
1.1	<b>Data protection by design and default</b>	The General Data Protection Regulation (GDPR) encourages producers of the products, services and applications to take into account the right to data protection when they develop, design, select and use applications, services and products which will necessitate processing of personal data ( <i>data protection by design</i> ). <sup>5</sup> In doing so, these producers are encouraged to give due consideration to the

<sup>5</sup> Recital 78, GDPR.

		<p>state of the art.<sup>6</sup> In other words, they should take account of the current progress in technology that is available in the market.<sup>7</sup></p> <p>The CyberSANE platform is such a service which will be provided to process various types of data, including personal data such as names, email addresses, IP addresses and traffic data (e.g. traffic data resulting from an online activity of an employee). In that sense, it is a 'means' of processing<sup>8</sup>.</p> <p>The end user organisation of the CyberSANE platform (e.g. a critical infrastructure operator) will be a 'data controller' and will be responsible for fulfilling data protection obligations. One of these obligations is the data protection by design and by default, which requires to put in place technical and organisational measures at the design stage with a view to implement data protection principles. The fulfilment of this obligation starts from the time when the 'means' of processing are determined (i.e. before any processing occurs). Product and service providers' roles in designing the system are, therefore, crucial to make sure that final users can comply with their obligation. At the same time, the way in which the CyberSANE platform will be used once it is put into market will be of utmost importance to comply with data protection principles. For instance, users or administrators responsible for onboarding of data ingestion feeds, which might contain personal data, should take the time to classify within CyberSANE the estimated impact of the breach of that data.</p> <p>In order to enable data controllers to fulfil by-design obligation and other data protection requirements, the developers of CyberSANE have taken into account the data protection principles in the design of the platform, giving due regard to the state of the art. These principles are examined more concretely and separately below.</p>
1.2	<b>Lawfulness</b>	<p>Any organisation that deploys the CyberSANE platform has to establish a legal basis for any processing of personal data.<sup>9</sup> <b>Legitimate interests</b> pursued by the data controller is one of such legal basis. The CyberSANE platform can be lawfully deployed on the basis of the legitimate interests of the data controller to ensure the security of its IT systems, provided that such interests are not overridden by the interests for fundamental rights and freedoms of the data subject.<sup>10</sup> Another potential legal basis for processing activities is compliance with a <b>legal obligation</b> emanating from EU law or domestic law (Art. 6/1-c). For instance, if the domestic law of the country in which the user organisation is established creates an obligation for that user to implement security measures and provide</p>

<sup>6</sup> Article 25(1), GDPR.

<sup>7</sup> European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default', 20 October 2020.

<sup>8</sup> Article 25(1), GDPR.

<sup>9</sup> Article 6, GDPR

<sup>10</sup> Article 6(1)(f), GDPR

D10.4 - Evaluation and Benchmarking Methodology Best Practices and Policy Development Guidelines for Replicability and Wider Use

		<p>relevant safeguards for data subjects, this would create a legal obligation.</p> <p>Another legal basis is <b>consent</b>; however this basis is not a feasible option for processing data of attackers. The use of CyberSANE would include the processing of employees data (e.g. email addresses) in order to prevent, identify and mitigate the breaches of the data of employees. However, employees' consent may be considered as an invalid basis for processing activities because there is a risk that such consent is not 'freely' given due to the power imbalance between the employer and the employee. Supervisory authorities recommend, therefore, to rely on a basis other than consent in the context of employment<sup>11</sup>. For that reason, relying on one of the basis mentioned in the previous paragraph would be a more appropriate option.</p> <p>When the processing also includes special categories of data (e.g. health, political opinion), one of the specific legal basis for this category of data should be additionally satisfied.<sup>12</sup> For instance, special categories data can be processed if personal data was manifestly made public by the data subject. CyberSANE incorporates the DarkNet component, which collects information from publicly available sources, and therefore could rely on this legal basis. When processing activity has one or more legal basis listed above, it should still demonstrate compliance with other principles and safeguards.</p>
1.3	<p><b>Purpose limitation (Only for certain purposes)</b></p>	<p>Purpose limitation requires that personal data is processed only for a specified, explicit and legitimate purposes. The purposes of the CyberSANE processing activities are ensuring network security, safety and security of relevant stakeholders (e.g. patients, port passengers, employees, consumers). Therefore, CyberSANE makes it possible to comply with this requirement. The compliance with this principle will depend on how the system will be used, and in particular, how data will be further processed. Any further processing (for instance, further sharing with a third party) should either be compatible with the initial purpose or should have a separate specified, explicit and legitimate purpose. Data collected for security purposes cannot be further processed for another purpose (e.g. marketing purposes) unless it is compatible with this initial purpose. One of the ways in which CyberSANE incorporated this principle is the integration of the PrivacyNet (Privacy and Data Protection and Orchestrator)<sup>13</sup>. PrivacyNet provides the necessary anonymization features that allow threat intelligence and information sharing capabilities with relevant parties, which will allow the processing of personal data only for limited purposes.</p>

<sup>11</sup> European Data Protection Board, 'Guidelines 05/2020 on consent under Regulation 2016/679' (2020), p. 9.

<sup>12</sup> Article 9, GDPR.

<sup>13</sup> See D7.1 Security & Privacy Algorithm Innovation Report; D7.2 Specification of the Privacy & Data Protection (PrivacyNet) Orchestrator.



1.4	<b>Data minimisation (Only adequate, relevant and necessary data)</b>	According to this principle, personal data should be processed as long as it is adequate, relevant and necessary for the purpose of processing. A natural consequence of this requirement is that data that is no longer necessary should be deleted or anonymized. Anonymisation is one of the important ways in which no personal data is used, stored or transferred beyond what is necessary for the processing purposes. CyberSANE integrates PrivacyNet which addresses, among others, this technique. Furthermore, compliance with this principle will depend on the use of the platform. Each organizational administrator will be able to delete users' information and assets from the CyberSANE environment, if required.
1.5	<b>Accuracy</b>	Accuracy principle obliges data controllers to keep personal data up-to-date. If data controller becomes aware that data regarding an individual is inaccurate, it should erase or correct such data. CyberSANE platform does not provide automatic correction of inaccurate data. The established ingestion pipelines and data processing algorithms strive to keep data accuracy. End-user organisations can make any appropriate changes, where necessary.
1.6	<b>Storage limitation</b>	According to the storage limitation principle, personal data must be deleted or anonymised as soon as they are no longer needed for the purposes for which they were collected. CyberSANE is designed to offer the possibility for the administrator of the end-user organization to delete any personal data that is no longer necessary. As a result, end-users can implement their data retention policies to comply with storage limitation. One of the ways in which CyberSANE incorporated this principle is by the integration of the PrivacyNet (Privacy and Data Protection and Orchestrator) <sup>14</sup> , which provides the necessary anonymization features that allow threat intelligence and information sharing capabilities with relevant parties.
1.7	<b>Fairness, transparency and data subject rights</b>	According to the principle of fair processing, data subjects should be made aware of the processing conducted on their data and understand what exactly is happening to it. Processing should not be performed in a secret manner. At the same time, the use of the CyberSANE platform will require a certain level of secrecy for security purposes. Even in that case, its use should remain transparent (in other words, affected parties should be aware of the existence of data processing for security purposes through privacy policies). One the ways in which CyberSANE incorporates this principle is by the integration of the PrivacyNet, which will enable the data controllers to enforce their privacy policies. In addition, the CyberSANE platform interfaces allow access to all information. It enables the provision of proper information and/or access to the data subjects, whenever appropriate (for instance, if an access request is made).

<sup>14</sup> See D7.1 Security & Privacy Algorithm Innovation Report; D7.2 Specification of the Privacy & Data Protection (PrivacyNet) Orchestrator.



1.8	<b>Accountability</b>	Data controllers (end-users of the CyberSANE platform) should be able to demonstrate accountability with all data protection principles. Regulatory compliance measures included putting in place measures listed in 1.9. Furthermore, to ensure compliance with the accountability principle, the design of the CyberSANE platform makes it possible for the data controller to give access to its Data Protection Officer (DPO) to oversight the use of the system, as well as implement privacy policies and obligations.
1.9	<b>Data security (Integrity and Confidentiality)</b>	<p>Data security measures have been implemented in order to keep the CyberSANE platform and the related components secure.</p> <p>Technical measures included:</p> <ul style="list-style-type: none"> <li>• Encryption and hash functions</li> <li>• Following international standards (ISO 27001<sup>15</sup> standard on information security, NIST 800-53 Security and Privacy Controls for Information Systems and Organizations<sup>16</sup>)</li> <li>• Regular backups</li> </ul> <p>Organizational measures included:</p> <ul style="list-style-type: none"> <li>• Training process is established, which is composed of the 'Train the trainers' phase and 'train the pilot end-users' phase. Cybersecurity training materials are produced. See D9.2 Training Materials and Report on Training Processes.</li> <li>• Strong access control system and restricted access to authorized persons</li> <li>• Use of credentials</li> <li>• Dissemination of confidential information is controlled and restricted</li> </ul>
1.10	<b>Data breach notifications</b>	Critical infrastructure operators often need to meet strict deadlines to notify component authorities about data breaches. CyberSANE offers an automated solution that provides support to detect and analyse anomalies. In that sense, it provides added value to speed up the notification processes. CyberSANE does not notify component authorities automatically. This is, in fact, desirable because security and/or compliance officers in the end-user organisation will be able to analyse whether conditions of a notification obligation are met and determine which particular information or data will be shared to make sure that all applicable law (e.g. data protection) is respected.
<b>2 Human agency and oversight</b>		
2.1		The principle of human agency and oversight derives from the moral principle of human autonomy. It requires that automated systems should support human agency and human decision-making and does not prevent him or her from taking his or her own actions. The

<sup>15</sup> <https://www.iso.org/isoiec-27001-information-security.html>

<sup>16</sup> <https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final>

		<p>CyberSANE platform is a support system, and its interface allows the security expert (end-user) to receive alerts on security incidents and receive relevant information on incidents and attacks. Security experts (end-users) stay in control and are able to take necessary decision and action. Respondents to the survey considered that the platform would support the decision-making in a positive way to improve critical infrastructure protection. The majority of the respondents to the survey do not think that the CyberSANE system could create any confusion among end-users as to whether they interact with a human or AI system. One respondent thinks that it could create confusion only among non-expert users, however the system is expected to be used by experts. The existing mitigation measures include different user roles, visual interface of the system and training to users.</p>
3	<b>Technical robustness and safety</b>	
		<p>The technical robustness and safety of each component will have an impact on the overall robustness and safety of the platform. Each component deployed measures to ensure the integrity, robustness and overall security of the CyberSANE system. Relevant measures include:</p> <ul style="list-style-type: none"> <li>- Regular updates and improved detection tools.</li> <li>- Keeping events and alarm.</li> <li>- Keep training procedures up to date.</li> <li>- Following international standards (ISO 27001 Standard on information security, NIST 800-53 Security and Privacy Controls for Information Systems and Organizations)</li> <li>- Updating data sources for EventRegistry regularly</li> <li>- Each tool integrated to the CyberSANE platform have its own procedures and are updated regularly.</li> </ul>
4	<b>Transparency</b>	
4.1	<b>Explainability and traceability</b>	<p>At least a certain level of explainability is necessary to achieve trustworthy AI to make individuals aware of how and why an AI system acts in a certain way, and what kind of consequences it may have. Explainability ensures that individuals can recognize undesired impacts and object to the AI-based decision-making. Different level of explanation is possible depending on the circumstances and seriousness of potential impacts. Various explainability measures include traceability, auditability and transparent communication on system capabilities. CyberSANE has addressed these three aspects.</p> <p>In terms of traceability, the CyberSANE platform acts as a meta-tool in which many individual tools may integrate. All data transferred in the CyberSANE core are properly logged, while the transferring process requires the adaptation of a specific data model and rules. All actions on the data are properly logged allowing traceability at any time. Furthermore, most respondents to the survey acknowledged</p>

		that the CyberSANE provide information to users on how to adequately use the system and what the capabilities and technical limitations are. A responded noted that the CyberSANE provided descriptive training materials to its users on how to properly use CyberSANE services and functions. It should be noted that the algorithmic-decision making, and machine learning techniques require specialized knowledge. Lack of expertise in this area may potentially create limitation to the proper understanding of the functioning of certain components involving AI. The best way to mitigate this is to increase training in the long term during the lifetime of the AI system (including when it is put into use). The third aspect, auditability, is examined in 7.2.
<b>5</b>	<b>Diversity, non-discrimination and fairness</b>	
		<p>CyberSANE ensures diversity, inclusion and fairness through:</p> <ul style="list-style-type: none"> <li>- Testing of all releases before deploying them at the operation environment</li> <li>- Descriptive training material of how to properly use the CyberSANE system and services</li> <li>- Design and implementation of user-interface that significantly increases user experience and provision of advanced services</li> </ul>
<b>6</b>	<b>Societal well-being</b>	
<b>6.1</b>	<b>Impacts to the workforce</b>	In terms of impacts to the workforce, no risk of de-skilling of the workforce is identified. CyberSANE is designed and provide services to security professionals and experts that undertake the responsibility of properly securing their organisation. To this end, it does not require any new digital skill from these persons since they are very familiar with the services and processes provided. A respondent from an end-user organization responded that the use of a tool such as CyberSANE would require a company to employ and/or train security experts.
<b>7</b>	<b>Accountability and regulatory compliance</b>	
<b>7.1</b>	<b>Performance of notification duties</b>	Essential service operators such as energy, transportation and health service providers have the obligation to notify promptly the authorities or the CSIRT of incidents having a significant impact. To this end, they need to meet strict deadlines for notification. CyberSANE facilitates the compliance with notification obligations by automating certain tasks to meet these strict deadlines. CyberSANE integrates with various Malware Information Sharing Platform (MISP) instances to allow prompt dissemination of incidents. It allows security professionals to build alert configuration to allow the system to automatically send notification to the designated individuals within the organisation when an incident is identified. Security professional and other competent persons (e.g., compliance officer) will be able to examine the relevant information, and assess the relevant factors,

		including the nature of the incident (number of users, duration and geographical spread) and determine if conditions of a notification obligation are met. CyberSANE does not automatically notify institutions outside the end-user organisation. This will in fact ensure that all privacy protocols are complied with by the security professionals and/or compliance officers.
7.2	<b>Auditability</b>	CyberSANE facilitates auditability by internal and independent parties and risk management. All data stored and processed within CyberSANE adopt specific data models and rules in order for the system to be able to handle and maintain. Therefore, CyberSANE allows for providing data in a structure, commonly used and machine-readable format (data portability). In case GDPR & NIS Directive (NISD) <sup>17</sup> will apply together, CyberSANE will enable the implementation of both framework cumulatively, for example it is possible to keep a list of records for distinct obligations.

Table 1 Legal and ethical assessment of the CyberSANE platform

## 2.4 Best practices

Based on the evaluation of the CyberSANE platform, certain best practices can be derived regarding the operation and use of incident handling and response approaches in critical infrastructures. These best practices will be helpful for any organisation who might deploy the CyberSANE platform beyond the lifetime of the project. Putting in place these best practices will ensure that the CyberSANE results will successfully extend and apply to other critical information infrastructures.

WP10 and specifically Task 10.5 identified the following best practices:

- Adherence to international standards to put in place adequate technical means or measures.
- Establishment of privacy-friendly policies for the day-to-day organization's operation.
- Provision of training to employees on privacy, data protection and security aspects, as well as ethical aspects concerning the new technologies such artificial intelligence.
- Inclusion of trustworthiness assessment for the use AI systems in the organization's operations.
- Implementation of strong access control systems and provision of restricted access to authorized persons.

The next chapter also provided some best practices derived from the analysis of policy aspects as part of policy recommendations.

<sup>17</sup> <https://www.enisa.europa.eu/topics/nis-directive>

## Chapter 3 Policy Guidelines for Replicability and Wider Use

Building on the evaluation of the CyberSANE platform, and the research carried out in the framework of T10.4 and T10.5, this Chapter 3 aims to reflect on the policy implications of the relevant legal and ethical framework identified in **D2.2 Legal and Ethical Requirements** and provide relevant policy recommendations. For this purpose, it focuses on a wide range of instruments concerning (i) the development and use of artificial intelligence, (ii) information and digital evidence sharing, and (iii) freedom of expression. It gives particular attention to the developments in the law and policy-making since the delivery of the D2.2.

### 3.1 Leveraging AI in Critical Infrastructures (CIs): Policy considerations

Since the delivery of **D2.2 Legal and Ethical Requirements**, there has been new developments concerning the regulation of AI at the policy level. The HLEG AI has continued its work to improve and deliver its practical tool to assess ethics requirements for AI. Moreover, the European Commission has elaborated a proposal of a legal instrument to regulate AI (the so-called 'proposed AI Act'). Both developments play an important role for the CyberSANE project. As it will be further explained below, the CyberSANE platform, or some of its components, may fall under the application of the AI Act if it is approved by the European Parliament. At the same time, ethics requirements – which were incorporated in the assessment of the platform in the previous chapter- remain to be relevant for automated network security systems such as CyberSANE because ethics will continue to govern the development and use of AI until the AI Act is adopted, or even after it will be adopted because the AI Act will not govern all types of AI systems<sup>18</sup>. This is why, it is crucial to consider the policy implications of these instruments, which will shape any further development and use of the CyberSANE platform or similar platforms beyond the lifetime of the project.

The main aim of this sub-chapter (3.1) is to point out the gaps or areas of improvement in these instruments. This sub-chapter will first provide a brief description of the Ethics Guidelines for Trustworthy AI beyond what was covered by the D2.2 in order to allow a deeper understanding of the EU's AI policy and better shape the policy considerations. It then continues with policy recommendations with a particular focus on the practical implementation of the ethics guidelines. Lastly, this sub-chapter will introduce the proposed AI Act, and offer policy recommendations regarding its scope.

---

<sup>18</sup> For further discussion, see 3.1.2 below.

### 3.1.1 Ethics Guidelines for Trustworthy AI: A European Approach to AI

AI has a great potential to anticipate cyber threats and increase resilience of critical infrastructures. The promises of artificial intelligence may become a reality if it complies with all applicable norms. Ethical standards, in particular, have been an important regulatory tool to guide the development and use of AI in Europe. In 2017, the European Council pointed out to a ‘sense of urgency’ to address emerging issues such as AI, and the need to ensure a high level of data protection, digital rights and ethical standards. The Council invited the European Commission to put forward a European approach to artificial intelligence to strengthen the EU’s ‘risk-based’ innovation capacity in new markets.<sup>19</sup>

In 2018, European Commission set out a European initiative on AI with three main goals<sup>20</sup>:

- Boost the EU's technological and industrial capacity and AI uptake across the economy
- Prepare for socio-economic changes brought about by AI
- Ensure an appropriate ethical and legal framework based on the Union's values and in line with the Charter of Fundamental Rights of the EU

The AI initiative acknowledges that transformative technologies such as AI may raise new ethical and legal questions, such as the risk of biased decision-making arising from poor training data. This initiative highlights the need to establish trust and accountability around its development and use. The ambition of the European AI policy is to promote new technologies based on values. This ambition underpins a sustainable approach to AI which promotes innovation and competitive level playing field while respecting the fundamental rights and ethical principles. The approach to regulate AI aims to build on the existing standards on data protection, network and information systems security, safety, liability, and to strengthen them. In this context, European initiative announced to publish ethics guidelines in order to address issues such as algorithmic transparency, minimisation of the risk of bias or error, fairness and security. These guidelines are meant to build on the previous work carried out on ethical aspects of AI by the European Commission’s advisory group.<sup>21</sup>

In order to implement this initiative, the European Commission established the High-Level Expert Group on Artificial Intelligence (HLEG AI) in April 2018. The HLEG AI is an independent group mandated with the drafting of two deliverables: (1) AI Ethics Guidelines and (2) Policy and Investment Recommendations. “Ethics Guidelines for Trustworthy AI” is the first deliverable published on 8 April 2019.

---

<sup>19</sup> European Council, Cover Note from General Secretariat of the Council to Delegations, EUCO 14/17, 19 October 2017, available at <https://data.consilium.europa.eu/doc/document/ST-14-2017-INIT/en/pdf>.

<sup>20</sup> European Commission, Artificial Intelligence for Europe (SWD(2018) 137 final), 25 April 2018 COM(2018) 237 final, available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN>.

<sup>21</sup> European Group on Ethics in Science and New Technologies. See [https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/scientific-support-eu-policies/ege\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/scientific-support-eu-policies/ege_en).



The HLEG AI acknowledges the importance of AI for the society, and its promise to increase human flourishing. AI can enhance individual and societal well-being and the common good, as well as bringing progress and innovation. To achieve its ambitions, AI systems need to be human-centric. This means that it should serve the humanity and common good and improve human welfare and freedom. In that context, any potential risks should be handled appropriately and proportionately. Ethics Guidelines expects AI producers to embed in their products and services ‘trustworthy AI’. Trustworthy AI aims to prevent or minimize unwanted consequences while taking advantage of the benefits.

Trustworthy AI refers to the entire life cycle of AI that should incorporate three components:

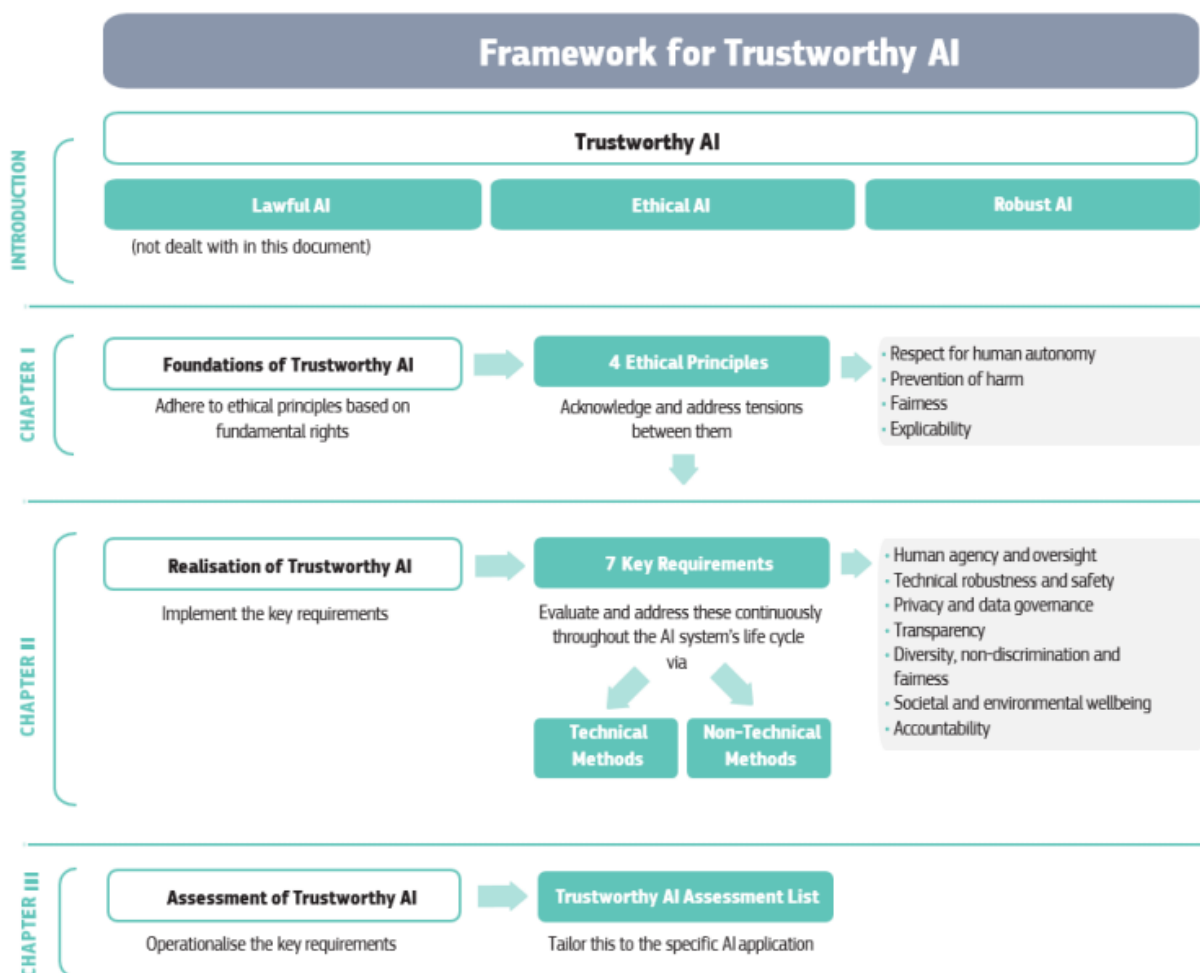
- 1) **Lawful AI:** AI should comply with all applicable laws and regulations.
- 2) **Ethical AI:** AI should adhere to ethical principles and values.
- 3) **Robust AI:** AI should be robust from a technical and social perspective to mitigate harms.

All three components ideally work in harmony to achieve trustworthiness. In practice, however, this is often more difficult because different norms are often in conflict and require making a choice of a norm over another (for instance, increasing accuracy would require collecting more data, and can be in conflict with privacy).

The Ethics Guidelines do not address lawful AI and does not provide advice on how the legal norms will be implemented in the context of AI. They rather build a framework based on ethical and robust AI. The rationale for focusing on ethics result from the fact that legal norms can be outdated to keep up with the technological developments, and law-making requires time to catch up with new technologies. Four ethical principles underpin the trustworthy AI:

- 1) **Respect for human autonomy:** AI should aim at enhancing and complementing human cognitive and social skills. Design options that condition individuals to make a particular choice (for instance, through default settings) may hinder them to determine their own actions. Human-centric design should secure human oversight over AI work processes and create meaningful work for humans.
- 2) **The principle of prevention of harm:** Lack of robustness or malicious use can cause physical or mental harm, for instance by making a hospital vulnerable to cyber threats, and causing malfunctioning to medical devices. Developers and users should make sure that the AI systems operate in a safe and secure way. In particular, human-centric design must pay attention to the vulnerabilities (e.g. disability), and power asymmetries between different interest groups such as business and consumers, employees and employers, and citizens and government.
- 3) **Principle of fairness:** Use and development of technologies bring both benefits and costs for the society. These costs and benefits should be fairly distributed between different actors in the society. Economic efficiency and security at the expense of unreasonably depriving individuals from their privacy or making them subject to biased decision-making would, for instance, lead to unfair circumstances. In addition, fairness requires to put in place procedures in which individuals can challenge the AI-based decision-making and seek effective remedies.

4) **The principle of explicability:** The problem known as ‘black box’ can arise if it is not possible to know how an algorithmic model generated a particular output (e.g. risk assessment). Without at least a certain level of explainability, individuals cannot know how and why AI system acts in a certain way, and what kind of consequences it may have. This would make it difficult or even impossible for individuals to recognize unwanted impacts and object to the AI-based decision-making. The Ethics Guidelines does not expect from AI practitioners to achieve absolute explainability. Different level of explanation is possible depending on the circumstances and seriousness of potential impacts. Various explainability measures include traceability, auditability and transparent communication on system capabilities.



Building on these ethical principles, the Ethics Guidelines establish specific requirements for trustworthy AI, including human oversight, transparency, data governance and accountability. Importantly, the Ethics Guidelines are accompanied with an Assessment List, which is meant to provide a practical tool to operationalize Trustworthy AI. Similar to the Ethics Guidelines, this



practical tool does not deal with the legal compliance of an AI system ('legal' AI). The CyberSANE project has incorporated this tool in its evaluation phase.<sup>22</sup> What follows is policy considerations on the assessment list.

### 3.2.1.1 Policy considerations

The Ethics Guidelines and their practical assessment tool are important steps forward to mitigate potential risks. Studies show that applicable legal rules may be inadequate to address certain aspects of AI systems, including in the areas of human oversight, transparency and traceability.<sup>23</sup> The European Union aims to address this legislative gap by adopting an AI-specific legislation generally referred to as the 'AI Act'. The policy considerations on the proposal of this legislation will be examined in the remaining of this sub-chapter. At this stage, it is worth to briefly note that, if adopted, the AI Act will not regulate all AI systems, and will potentially leave certain AI systems used in critical infrastructures involving fundamental right risk outside its scope.<sup>24</sup> Therefore, the Ethics Guidelines will remain as an important regulatory tool for the development and use of AI systems at least in near future.

There exist some limitations to the assessment to this framework. Firstly, it is rather a general tool that does not take into account a sector or a context in which the system will be deployed. To have a practical impact, the assessment tool should be first tailored to the specific use case. There is research going on how to tailor the tool for the use in healthcare, human resources and public sector.<sup>25</sup> Other sectors would also benefit from such a sector-specific focus studies. Secondly, most questions involve the so-called yes or no questions, meaning that a question expects a reply that either affirms it or denies it. The Assessment List does not expect an explanation or justification to the answer provided. No matter whether the answer is yes or no, the assessment process would benefit from explanation or justification to the answer provided. For instance, if the assessor thinks that a risk is non-existent, it is crucial to know why and how the assessor come to that conclusion (what factors were taken into account, what is the risk level). Similarly, if a measure was taken or not taken, it would be important to know which measures were taken or why a certain measure was not taken in order to be able to check whether there is a justification for doing so (That measure may not be necessary or effective in that particular case.) Otherwise, it would be a tool that offers ticking or not ticking boxes without giving too much consideration to what each question entails. Furthermore, it would be crucial to provide examples or further guidance to the assessors, if they recognize that a certain issue is not yet addressed or needs more attention.<sup>26</sup> The Assessment List can better achieve its purpose to make stakeholders aware of the potential risks of AI and take necessary measures, if policy makers provide an amendment of the assessment tool or further guidance or clarification on these practical limitations.

---

<sup>22</sup> See Chapter 2 above.

<sup>23</sup> European Commission, White Paper on Artificial Intelligence, 19 February 2020, COM (2020) 65 final.

<sup>24</sup> See 3.1.2

<sup>25</sup> Nathalie Smuha, Towards a Practical Assessment Tool for Trustworthy AI, Presentation at the European AI Week 2022, 15 March 2022, available at <https://www.youtube.com/watch?v=tb47bUIKPec&t=858s>.

<sup>26</sup> Nathalie Smuha, Towards a Practical Assessment Tool for Trustworthy AI, Presentation at the European AI Week 2022, 15 March 2022, available at <https://www.youtube.com/watch?v=tb47bUIKPec&t=858s>.

In addition, a practical challenge could arise from lack of expertise in the field of ethical issues involved in AI. The HLEG AI stresses that the 'best' way to complete the Assessment List for Trustworthy AI is to involve a multidisciplinary team of people 'with specific competences or expertise on each of the 7 requirements and related questions'. It is the ideal scenario to bring together professionals, such as designers, developers, front-end staff and legal officers, to deliver, overall, a good understanding and assessment of all the relevant aspects. In reality, a challenge that businesses could face is that not all of them will have employees with the AI-specific skills, especially considering that technologies move fast, and the most requirements involve a technical aspect. In particular, the end-users (e.g. a hospital) may lack staff with specific expertise on the technological development in order to assess the ethical use. Therefore, it is recommended that policy initiatives support the training and education of involved stakeholders.

#### **Policy recommendations**

- It is recommended to amend the Assessment List for Trustworthy AI to eliminate the so-called 'yes or no questions', and to include questions that ask an explanation or justification to the answer provided.
- It is recommended to provide further guidance or clarification to those who will be involved in the assessment process, by providing examples or giving further information about the next steps.
- It is recommended to put in place policy initiatives to support the training and education of involved stakeholders. It is a best practice for businesses that will deploy the CyberSANE platform or similar automated network security systems to train their staff on the ethical assessment of AI systems.

### **3.1.2 Proposed AI Act and the implications for critical infrastructures**

As already noted in the previous sub-section, the European Union aims to fill in the legislative gap on issues surrounding emerging technologies. For that purpose, on 21 April 2021, the European Commission proposed the AI Act to establish harmonized legal requirements and conformity assessment procedures for AI systems, which also concern critical infrastructure protection. In line with the AI strategy established at an earlier stage, the proposal of the Act follows a 'risk-based' approach to artificial intelligence. Importantly, the AI Act is designed as a 'regulation' which means that it will be directly applicable in all EU member states after it is adopted. As it is a proposal, it is not yet in force, and some of its provisions can still change. This sub-section will briefly describe the novelties brought by the proposed AI Act and analyse policy implications for critical infrastructures.

## Scope of application

The AI Act establishes obligations for providers, importers, distributors of AI systems. Businesses that provide AI products to the EU market are covered by the legislation regardless of their country of establishment. Users also have obligations under the AI Act. They include all individuals, businesses and public institutions located in the EU who uses AI in their authority.<sup>27</sup> Military or non-commercial use, as well as, the use of AI by non-EU public authorities are left out of the scope of application.

## Definition of AI

The AI Act aims to establish a clear definition of AI to ensure legal certainty for those who develop AI, on the one hand, and not to hinder the development of future technologies, on the other.<sup>28</sup> Nevertheless, the AI Act provides a rather broad definition of AI systems. Article 3 defines AI systems as any software that is developed with certain techniques and approach, which can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with. AI techniques and approaches include:

- Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning
- Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems
- Statistical approaches, Bayesian estimation, search and optimization methods

In essence, any software that provides outputs based on a human-set input will qualify as AI, which is wide enough to cover almost all algorithms.<sup>29</sup> The listed techniques and approaches are meant to specify the definition.<sup>30</sup> However they include a wide range of approaches used by computer scientists. This may create uncertainty as to whether software or automation that are not typically considered as AI could also fall under this definition.<sup>31</sup>

## Types of AI systems

The AI Act proposal establishes different categories of AI, and make them subject to different requirements. In the future legal landscape, AI systems could fall under one of the four risk categories<sup>32</sup>:

---

<sup>27</sup> Article 2, AI Act proposal.

<sup>28</sup> Recital 6, AI Act proposal.

<sup>29</sup> Smuha et al., How the EU can achieve legally trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence, LEADs Lab, University Birmingham, 5 August 2021, p. 14.

<sup>30</sup> Recital 6, AI Act proposal.

<sup>31</sup> Nathalie Smuha, Emma Ahmed-Rengers, Adam Harkens, Wenlong Li, James MacLaren, Riccardo Pisellif and Karen Yeung, 'How the EU can achieve legally trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence', LEADs Lab, University Birmingham, 5 August 2021, p. 14.

<sup>32</sup> European Commission, Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence, 21 April 2021, available at [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_21\\_1682](https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682).

1. Unacceptable risk
2. High-risk
3. Limited risk
4. Minimal risk

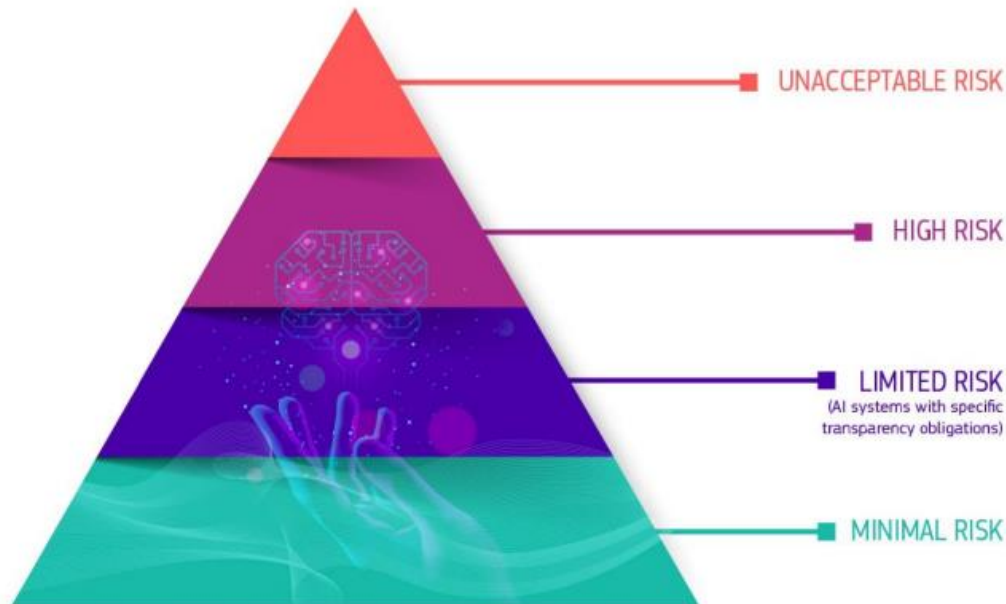
The first category involves unacceptable risk level. Certain AI practices are so intrusive and have such a high impact on the freedom and rights of individuals that they cannot be tolerated. These practices include manipulation of human behaviour through subliminal techniques, exploitation of vulnerable groups such as children (e.g. toys leading children to harmful behaviour), social scoring for public purposes, and real time biometric identification in public areas. The development and use of such practices are prohibited.

The second and third category concerns AI systems with high-risk and limited risk levels. These systems are allowed as long as they comply with the specific requirements. High-risk AI include AI-based products and safety components that are intended to be used in specified areas such as critical infrastructure and/or fall under the specified legislation (e.g. medical device legislation). Limited risk AI, or the so-called AI with transparency obligations are the ones that could pose a threat to the autonomy and free will of the individuals if they are not aware that they are interacting with AI. The users of the AI systems resembling humans (deepfakes, bots), emotion recognition or biometric systems are expected to inform individuals interacting with them so that they are aware of what they are engaging with before they decide their own actions<sup>33</sup>.

All other AI systems that do not fall under categories 1-3 are called 'minimal' risk, and are not regulated by the AI Act. In other words, they can be developed and used without any AI-specific restrictions, as long as they comply with other applicable norms. As further explained below, high-risk AI is particularly relevant for the critical infrastructure protection and the CyberSANE project. This is why, it will be the main focus of the remaining sub-chapter.

---

<sup>33</sup> Article 52(2-3), AI Act proposal.



The Pyramid of Criticality for AI Systems

Image: The categories of AI systems under the AI Act proposal (Kop,2021)<sup>34</sup>

### High-risk AI systems

The AI Act does not provide a definition of high-risk AI. Instead, it labels certain AI systems as high-risk AI based on the legislation under which they fall or the area in which it will be used. AI applications are considered as high-risk if (i) they are specified products that fall under the enumerated legislation (Annex II), or (ii) they are intended to be used in certain areas (Annex III).

Annex III lists the following areas, which will trigger the application of AI Act requirements:

1. Real time or post remote biometric identification and categorization of natural persons
2. Management and operation of critical infrastructure
3. Access to and assessment in education and vocational training
4. Employment, workers' management and access to self-employment
5. Access to public assistance benefits and services, emergency first response services, evaluation of creditworthiness
6. Field of law enforcement for the purposes of crime prevention, investigation or probation, including individual risk assessment, emotion detection, detecting deepfakes
7. Migration, asylum and border control management

---

<sup>34</sup> Mauritz Kop, 'EU Artificial Intelligence Act: The European Approach to AI', Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University 2(2021).

## 8. Justice sector, including the application of law in courts and judicial institutions

What is especially relevant for the critical infrastructure protection is that Annex III covers AI systems in the area of management and operation of critical infrastructure. More specifically, AI systems would be considered as high-risk AI if they are 'intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity'. The AI Act proposal does not intend to label all AI systems used in the critical infrastructures as high risk. It introduces two criteria:

1. AI system should be a 'safety component'.
2. AI system should be used in the management and operation of road traffic or supply of water, gas, heating and electricity.

In terms of the first criterion, Article 3 defines safety component as 'a component of a product or of a system which fulfils a safety function for that product or system or the failure or malfunctioning of which endangers the health and safety of persons or property'. This definition limits the application of the AI Act to the critical infrastructure AI in two ways. Firstly, the requirement of performing a safety function excludes the AI systems that does not have this function. For instance, certain AI-based components integrated or can be integrated in the CyberSANE platform will not necessarily have a safety function, but their function will be to monitor data to identify cyber threats. Secondly, AI that does not have safety function can still be high-risk if its malfunctioning may have dangerous impacts, however the impacts beyond the 'health and safety' of the individuals and property are not considered. For instance, AI systems that collect and aggregate unstructured data for network security may have impacts on the right to privacy, data protection and freedom of expression of individuals. Yet, manufacturers could consider that these systems are not high risk because they do not fulfil a safety function or endanger health and safety of individuals. AI systems that could have significant adverse impacts on fundamental rights could consequently be considered as non-high risk AI and would not be subject to a conformity assessment.

The second criterion for the critical infrastructure AI significantly restricts the application of the AI Act in certain sectors. Annex III does not mention many sectors covered by the NIS Directive (NISD) and the respective domestic legislation, for instance, maritime sector, banking and finance, health sector and digital infrastructure. Hence, there is a sectoral discrepancy between the AI Act and cybersecurity legislation. Because of this discrepancy, an AI system that is considered as high-risk in one sector may be qualified as non-high risk in another sector.

Nevertheless, this does not automatically mean that the sectors that are explicitly mentioned in the Annex III (e.g. maritime, health sector) should not concern themselves with the AI Act. AI systems that fall under other seven area listed above can still qualify high risk AI. This could be the case, for instance, if the critical infrastructure will use 'real time' or 'post' remote biometric identification and categorization of natural persons (for instance, based on location) in order to identify threats to its network security. Another example of high-risk AI for critical infrastructure could be AI systems used to deliver 'emergency first response services'<sup>35</sup>. For instance, if a

---

<sup>35</sup> Annex III.5(c).



hospital is tasked with establishing priority in the provision of medical aid in emergency situations, and the AI system is used for that purpose, this would be considered as a high-risk AI scenario.

In addition, the AI Act will be relevant for the sectors that use the products that fall under the listed legislation in Annex II as specified in Article 6(1). Annex II includes a wide range of legislation that covers products including medical devices, radio equipment (e.g. IoT devices, Apps) and equipment used in marine or civil aviation sector. AI systems that qualify as one of the specified products (or their safety component), for instance a medical device, will be considered as high-risk AI even if health sector is not mentioned among critical infrastructure as explained above. However, AI Act is silent when it comes to the AI-based products and services which do not fall under the listed legislation but is used in the operation and management of non-listed critical sectors. For example, an AI-based component of CyberSANE may be used in the management of a health institution but may not itself constitute a medical device or a safety component of a medical device. In that case, it could become less clear whether it would be covered by high-risk AI.

### **3.1.2.1 Policy considerations**

It is a welcome development that the proposed AI Act has also included systems that are used in critical infrastructures. Critical infrastructure AI that complies with the high-risk requirements such as data governance, accuracy and cybersecurity would expectedly contribute to the resilience of the critical infrastructure, on the one hand, and protection of individuals, on the other.

Nevertheless, it is regrettable that the AI Act's application to critical infrastructure AI is limited, which could create uncertainty for operators of critical services and manufacturers as to whether they will be bound by it. Thus far, the AI Act leaves many sectors covered by the NISD out of its scope, and creates a sectoral discrepancy between the AI regulation and cybersecurity legislation. This could create a regulatory obstacle for the application of harmonious rules across all critical sectors. Furthermore, manufacturers could face a legal uncertainty and implementation costs to determine whether an AI system used in one sector would be subject to the same requirements if the same system is used in another sector. The same or similar technology used in one sector can be subject to different requirements if used in another sector. As a result, such uncertainty could be a challenge for the replicability and wider use of AI-based systems in different sectors.<sup>36</sup> It is therefore recommended to clarify the sectoral inconsistencies in the area of critical infrastructure protection.

In addition, the AI Act' proposal's application to critical infrastructure is limited in terms of the definition of safety component. Critical infrastructure AI will be labelled as high risk if it is a safety component used in the management and operation in some critical sectors. Safety components cover components which have a safety function, or which can cause a threat to the health and safety of persons or property if they fail or malfunction. As explained in the previous section (3.1.2), this definition is not wide enough to cover all AI systems that involve a fundamental right risk. AI systems may have fundamental rights impacts beyond health and safety, such as the

---

<sup>36</sup> Inconsistent or unclear application of rules can cause regulatory uncertainty for the manufactures. For an analysis focusing on the regulatory challenges for the cybersecurity of medical devices, see Elisabetta Biasin and Erik Kamenjasevic, 'Cybersecurity of medical devices: new challenges arising from the AI Act and NIS 2 Directive proposals', forthcoming (2022).

impacts to the right to privacy and freedom of expression, nevertheless, manufacturers could consider that these systems are not high risk because they do not fulfil a safety function or endanger health and safety of individuals. As a result, AI systems that have significant fundamental rights impacts beyond health and safety may not be labelled as high-risk and may not need to comply with the requirements.

Relevant to the CyberSANE platform, D2.2 has already discussed that the CyberSANE platform may have impacts on fundamental rights such as the right to privacy, data protection and freedom of expression. One or more components of the platform could qualify as high-risk AI if they fall under Annex II (e.g. listed products or their safety components), or Annex III (e.g. listed sectors, biometric identification) as described above. However, despite the fundamental rights risks, the possibility cannot be ruled out that the platform or its components cannot be considered as high-risk AI if they do not satisfy any high-risk criteria, for instance in terms of sectors, or being a safety component of listed products. In that case, it will remain unclear, whether and how the fundamental right risks will be addressed and mitigated.

The AI Act could benefit from a harmonious and consistent application of high-risk requirements to all AI systems, which are used in the critical infrastructure, and which could pose an unjustified interference to fundamental rights. It is therefore recommended to take legislative or policy measures- through expanding the scope of the AI Act or through another way- to ensure that fundamental rights risks beyond health and safety are not overlooked in the in the development and use of AI systems.

#### **Policy recommendations with regard to application of AI in CIs**

- It is recommended to eliminate the sectoral inconsistencies to ensure a consistent application of AI-related requirements across all critical infrastructures.
- It is recommended to expand the list of high-risk AI systems in the proposed AI Act or through delegated acts that can be adopted after the entry into force of the proposed AI Act to cover all AI systems with significant fundamental rights impacts.
- It is recommended to take any other legislative or policy measures to ensure that all fundamental rights risks are taken into account in the development and use of AI systems.

### **3.2 Information sharing: Policy considerations**

The EU's critical infrastructure and essential services are increasingly interdependent and digitised. Information sharing about incidents is crucial to understand causes of cyberattacks, their cross-sector impacts, to establish simpler reporting processes for authorities and companies and to create a more resilient cyberspace. This is, in fact, important to ensure accountability for the wrongdoings committed in cyberspace, considering that cyberattacks and cybercrime are on the rise. Sharing information between private and public entities are crucial to identify those who are responsible for these wrongdoings and to ensure redress for the victims.

The CyberSANE project has contributed to the efforts to ensure accountability for a secure cyberspace. The project has included the integration of the ShareNet component to provide



intelligence sharing features to allow data owners to exchange information about cyber incidents with relevant parties in a privacy-friendly and secure manner<sup>37</sup>. Moreover, CyberSANE has researched on the techniques to maintain and store all recovered forensic information and evidential data to be able to link the chain of events with the attacker's data in a trusted way.<sup>38</sup> This research will be valuable to be able to use vital information in a legal, reliable and secure way to identify and prosecute those who are responsible for (cyber)wrongdoings, and to shape cybersecurity and cybercrime policies in the long term.

For the purposes of the Task 10.5, this sub-chapter will focus on the policy implications of the legal framework on information sharing. Firstly, it will focus on the notification obligations under a wide range of legal instruments concerning different sectors in the area of critical infrastructure protection. In this way, it covers the policy considerations concerning the sectors in which CyberSANE end-users operates but also other sectors, taking into account the wider use of the CyberSANE platform beyond the project's lifetime. Secondly, it will focus on sharing of information with public authorities for the purposes of investigation and prosecution of crime, giving particular consideration to the developments in the recent data retention case law, and cross-border access to digital evidence. This sub-chapter will provide relevant policy recommendations on the examined framework.

### 3.2.1 Notification of security incidents

Reporting security incidents is a central part of the European cybersecurity framework. Incident reporting schemes stem from different pieces of legislation, such as the GDPR and the NIS Directive (NISD). Reporting obligations are coupled with sharing of information about vulnerabilities. Cybersecurity incident reporting is an important part of supervision. Although it does not solve incidents by itself, it helps national authorities and agencies to understand cybersecurity trends, issues and weaknesses within the sector and across sectors. Without incident reporting authorities may have to rely on media, which may not always give a balanced and accurate view of cybersecurity incidents.

This section briefly discusses notification requirements set out in different legal frameworks. More detailed information on reporting formats and procedures under different EU instruments can be found in the Cooperation Group (CG) Publication 04/20.<sup>39</sup> Because the CyberSANE system might be used in different areas, it is useful to lay out briefly what these requirements are and their interplay. The following section discusses the NIS framework and how the revised NIS framework responds to the fragmented set of obligations.

In the EU, there are several laws that provide for incident reporting obligations. The NISD introduces cybersecurity incident reporting for operators of essential services in a range of critical sectors such as energy, transport, finance and health. The GDPR requires data controllers to notify personal data breaches. Other reporting and information sharing requirements in variety of sectors and aim at different entities and include apart from the already mentioned NISD and GDPR: Directive (EU) 2018/1972 (EECC) (replacing and integrating in one code the previously

---

<sup>37</sup> See D8.1 Integration and Validation of the CyberSANE system.

<sup>38</sup> See D5.1 Prevention and Responses to Advanced Treats, and D5.2 Cyber Fusion Models.

<sup>39</sup> NIS Cooperation Group, Synergies in Cybersecurity Incident Reporting (CG Publication 04/2020), available at [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=72147](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=72147).

existing framework for electronic communications networks and services spread across several legislative acts),<sup>40</sup> Directive 2002/58/EC (ePrivacy Directive),<sup>41</sup> Regulation (EU) No 910/2014 (eIDAS Regulation)<sup>42</sup> Directive (EU) 2015/2366 (PSD2 Directive)<sup>43</sup> as well as Regulation (EU) 2017/745 (MDR)<sup>44</sup>. The reporting schemes of these interventions have in common that they are aimed to understanding (cyber-)security threats as well as identifying vulnerabilities.

### **Telecommunications Sector**

By the end of 2020, the European Electronic Communications Code (EECC)<sup>45</sup> came into effect across the EU but was only implemented into national legislation in some EU countries. Member States except for Spain, Croatia, Latvia, Lithuania, Ireland, Poland, Portugal, Romania, Slovenia and Sweden successfully transposed the EECC into their national laws.<sup>46</sup> While none of the CyberSANE partners operate in the telecommunications sector (and thus, are not bound by the mentioned EECC), this deliverable aims to cover the wider use of the platform across different sectors. Therefore, reporting requirements concerning various sectors will be briefly outlined below. Under Article 40 of the EECC the incident reporting requirements have a broader scope, including not only outages but also breaches of confidentiality, for instance.

Providers of public electronic communications networks or of publicly available electronic communications services are bound with notification obligation. There are more services within the scope of the EECC, including not only traditional telecom operators but also, for example, over-the-top providers of communications services such as WhatsApp.

Article 13a of the Framework Directive and Article 40 of the EECC, provide for three types of incident reporting: 1) National incident reporting from providers to NRAs, 2) Ad-hoc incident reporting between NRAs and ENISA, and 3) Annual summary reporting from national authorities to the European Commission and ENISA.

Note that in this setup ENISA acts as a collection point, anonymizing, aggregating and analysing the incident reports. In the current setup, NRAs can search incidents in the reporting tool (CIRAS),

---

<sup>40</sup> Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code [2018] OJ L 321/36 (EECC)

<sup>41</sup> Directive 2009/136/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector [2009] OJ L 201/37 (ePrivacy Directive).

<sup>42</sup> Regulation (EU) No. 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC [2014] OF L 257/73 (eIDAS Regulation).

<sup>43</sup> Directive (EU) of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC [2015] OJ L 337/35 (PSD2 Directive).

<sup>44</sup> 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 [2017] OJ L 117/1 (MDR).

<sup>45</sup> Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code [2018] OJ L 321/36 (EECC)

<sup>46</sup> European Commission, EU Electronic Communications Code: Commission refers 10 Member States to the Court of Justice of the EU, 6 April 2022, available at [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_22\\_1975](https://ec.europa.eu/commission/presscorner/detail/en/IP_22_1975).

but the incident reports themselves do not refer to countries or providers, making the overall summary reporting process less sensitive.<sup>47</sup>

### **Security incident reporting for trust services**

The eIDAS Regulation<sup>48</sup> (The regulation for electronic identification and trust services) came into force in 2016, sets up rules for trust services and e-ID schemes across the EU. Article 19 of the Regulation requires, among other requirements, that providers of trust services to assess risks, take appropriate security measures to mitigate the risks, and notify the supervisory body about significant incidents/breaches.

Trust service providers within the meaning of the Regulation are obligated to notify supervisory body and where applicable, other relevant bodies such as the competent national body for information security or data protection authority without undue delay but in any event within 24 hours having become aware of it. The notification requirement applies where there is a breach of security or loss of integrity that has significant impact of the trust service provided or on the personal data maintained.

Where there is a cross border element to the breach of security, the notified supervisory body informs the supervisory bodies in other Member States and ENISA. In addition, national supervisory bodies send annual summary reports about the notified breaches to ENISA and the Commission.<sup>49</sup>

### **Personal data breach reporting for telecom providers**

Article 4 of the E-Privacy Directive<sup>50</sup> requires the providers of publicly available electronic communications services to notify, without undue delay, the personal data breach to the competent national authority.

Cross-border reporting scheme is laid down in Article 2 of the Implementing Regulation 611/2013<sup>51</sup>. In case of a cross-border personal data breach, meaning a personal data breach that affects subscribers or individuals from Member States other than that of the competent national authority to which a breach has been notified, the competent authorities are obliged to notify other national authorities.

---

<sup>47</sup> ENISA, Telecom Security Incidents 2020 Annual Report 26 July 2021, available at <https://www.enisa.europa.eu/publications/telecom-annual-incident-reporting-2020>.

<sup>48</sup> Regulation (EU) No. 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC [2014] OF L 257/73 (eIDAS Regulation).

<sup>49</sup> ENISA, Trust Services Security Incidents 2020 - Annual Report, 26 July 2021, available at <https://www.enisa.europa.eu/publications/trust-services-security-incident-2020-annual-report>.

<sup>50</sup> Directive 2009/136/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector [2002] OJ L 201/37 (ePrivacy Directive).

<sup>51</sup> Commission Implementing Regulation (EU) 2018/151 of 30 January 2018 laying down rules for application of Directive (EU) 2016/1148 of the European Parliament and of the Council as regards further specification of the elements to be taken into account by digital service providers for managing the risks posed to the security of network and information systems and of the parameters for determining whether an incident has a substantial impact [2018] OJ L26/48.

## Payment Services Providers

Another instance where breach reporting is required concerns payment services. Under Article 96 of the Payment Services Directive 2,<sup>52</sup> payment service providers are to report major operational or security incidents without delay to the competent authority where the payment service provider is located. Where the incident has or may implicate the financial interests of payment service users, then payment service providers must inform its payment service users of the incident and of all measures that they can take to mitigate the adverse effects of the incident.<sup>53</sup> The competent authority of the home Member State is then obliged to provide relevant information about the incident to the European Banking Authority (EBA) and to the European Central Bank (ECB).<sup>54</sup> EBA and the ECB shall, in cooperation with the competent authority of the home Member State, assess the relevance of the incident to other relevant Union and national authorities and shall notify them accordingly. The ECB shall notify the members of the European System of Central Banks on issues relevant to the payment system.

On the basis of that notification, the competent authorities shall, where appropriate, take all of the necessary measures to protect the immediate safety of the financial system.

EBA shall, in close cooperation with the ECB and after consulting all relevant stakeholders, including those in the payment services market, reflecting all interests involved, issue guidelines on the implementation of Article 96.<sup>55</sup>

## Personal data breach reporting under GDPR

Article 33 of the GDPR requires from data controllers to (i) notify a personal data breach to the supervisory authority within 72 hours after becoming "aware" of a breach that poses risks for the privacy of one or more citizens, and (ii) communicate the personal data breach to the data subject without undue delay in accordance with Article 55 of the GDPR, unless the data breach is unlikely to result in a risk for the privacy of individuals.

For data breaches with cross-border relevance Article 60 of the GDPR sets up a cooperation mechanism between the lead authority (the authority where the notifying data controller is based) and other supervisory authorities, for example, in other Member States where citizens might be impacted). Article 60 of the GDPR states "The lead supervisory authority and the supervisory authorities concerned shall exchange all relevant information with each other."

---

<sup>52</sup> Directive (EU) of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC [2015] OJ L 337/35 (PSD2 Directive).

<sup>53</sup> Article 96(1) para. 2, PSD2.

<sup>54</sup> Article 96(2), PSD2.

<sup>55</sup> European Banking Authority, 'Guidelines on major incidents reporting under PSD2', (10 June 2021) available at [https://www.eba.europa.eu/sites/default/documents/files/document\\_library/Publications/Guidelines/2021/Guidelines%20on%20major%20incident%20reporting%20under%20PSD2%20EBA-GL-2021-03/1014562/Final%20revised%20Guidelines%20on%20major%20incident%20reporting%20under%20PSD2.pdf](https://www.eba.europa.eu/sites/default/documents/files/document_library/Publications/Guidelines/2021/Guidelines%20on%20major%20incident%20reporting%20under%20PSD2%20EBA-GL-2021-03/1014562/Final%20revised%20Guidelines%20on%20major%20incident%20reporting%20under%20PSD2.pdf).

Further, according to Article 59 of the GDPR, each supervisory authority is to draw up an annual report on its activities which may also consist of a list of infringements notified and measures taken.

### Medical Devices Incident Reporting

Article 87 of the EU Medical Devices Regulation (MDR) also requires manufacturers of devices available in the EU to report incidents. Manufacturers of devices made available in the Union market shall report any serious incident involving devices made available in the EU market.

In addition, under Article 92 of the MDR, the Commission, in collaboration with Member States, is designated to set up and manage an electronics system to collate and process the reports by manufacturers on serious incidents.

### A Brief Analysis of Different Reporting Schemes

The CG Publication 04/20 of the NIS cooperation group provides a comparison table of the incident reporting and information sharing schemes explained above on the basis of the following criteria: services in scope, incident definitions, notification timing, notification criteria, cross-border information sharing and annual summary reporting at EU level. A brief analysis of these different schemes provided below is useful to illustrate the divergences between them and to better substantiate the policy considerations explained in the following section.

One important question, given the **multiplicity of reporting schemes**, is how these schemes interact with each other. While the NISD introduces a cross-sectoral cybersecurity incident reporting scheme, in contrast Article 40 EEC, Articles 10 and 19 eIDAS Regulation, and Article 96 PSD2 have a very limited and clearly defined scope of application. In order to avoid a **duplication** of reporting obligations, Article 1(7) NISD sets forth that where a sector-specific Union act foresees security or notification requirements of at least equivalent effect, these *lex specialis* provisions shall prevail. Article 96 PSD2 is, for instance, considered as more specific law to the NISD with regard to the provision of payment services by credit institutions, and thus applies instead of the corresponding provisions of Article 14 NISD. According to Article 1(3) NISD, the same applies regarding pre-existing sector-specific legislation, namely, the reporting schemes of the EECC framework for the electronic communications networks and services and the eIDAS Regulation.

The GDPR does not constitute a *lex specialis* to the NISD in the sense of Article 1(7) NISD:<sup>56</sup> the GDPR applies to all data controllers and requires breach notification where personal data is at stake. Differing from that, the aforementioned schemes mandate breach notification if there is a significant disruption to the provision of the service. As such, although the notification obligations are very similar, they are not duplications and therefore do not exclude one another.

On a final note, information sharing schemes are generally coordinated between competent authorities and coordinated by EU-level institutions (i.e. ENISA or NIS Cooperation Group or the EBA), usually but not always – see PSD2 - where a cross-border incident is in question.

---

<sup>56</sup> A detailed comparison of the reporting schemes under the NISD and the GDPR can be found in: Sandra Schmitz-Berndt and Fabian Anheier, 'Synergies in Cybersecurity Incident Reporting – The NIS Cooperation Group Publication 04/20 in Context' European Data Protection Law Review (2021) 7(1).



### 3.2.1.1 Notification obligations under NIS and the revised framework

The NISD is a minimum harmonisation tool in the area of NIS security, allowing for stricter rules to be adopted or maintained at the national level.<sup>57</sup> The NISD legal framework can be summarised as below: It ;

- Requires the Member States to adopt a national strategy on the security of networks and information systems
- Creates a Cooperation Group in order to support and facilitate strategic cooperation and the exchange of information among Member States and to develop trust and confidence amongst them. The Group's overall mission is to achieve a high common level of security for network and information systems in the European Union. It supports and facilitates the strategic cooperation and the exchange of information among EU Member States. The NIS Cooperation Group's tasks are explicitly described in Article 11 of the NISD.
- Sets up a Computer Security Incident Response Teams network (CSIRTs network) to facilitate trust, collaboration and information exchange among Member States;
- Provides security and notification obligations for Operators of Essential Services (OES) and Digital Service Providers (DSP);
- Requires Member States to designate national competent authorities, single points of contact and CSIRTs with tasks related to the security of network and information systems.<sup>58</sup>

The NISD establishes an incident reporting framework covering the notification of significant incidents as well as requiring the implementation of security measures. Under the NISD, two regimes are provided for the obligation to report an incident (i.e. 'any event having an actual adverse effect on the security of' NIS): (1) operators of essential services ("OESs")<sup>59</sup> and (2) digital service providers ("DSPs").<sup>60</sup> Member States shall ensure that OESs and DSPs notify, 'without undue delay,' the national competent authority ("NCA") or the computer security incident response team ("CSIRT") of incidents having a significant impact on the continuity of the essential services they provide (in case of an OES), or incidents having a substantial impact on the provision of a digital service (in case of a DSP).<sup>61</sup>

Under the NISD, there is no clear information-sharing obligation between authorities handling security incidents. A clear mandate to share information is also lacking for DPAs under the GDPR. Recital 63 NISD specifies that in the context of compromised personal data, competent authorities under the NISD "should cooperate and exchange information on all relevant matters to tackle any personal data breaches resulting from incidents". But a framework to support this

---

<sup>57</sup> Article 3, NISD.

<sup>58</sup> *ibid.*

<sup>59</sup> Operators of Essential Services are defined as 'public or private entities of a type referred to in Annex II' that should also fulfil the cumulative criteria set out in Article 5(2) of the NISD. OES are identified by Member States.

<sup>60</sup> Annex III to the NISD lists as DSPs within the scope of the NISD only three types of services: online marketplaces, online search engines, and cloud computing services.

<sup>61</sup> Article 14(3) NISD for OES and Article 16(3) for DSP.

cooperation is lacking. As pointed out by Sandra Schmitz-Berndt and Fabian Anheier's report, similar issues also persist with regard to cross-border information sharing.<sup>62</sup>

### 3.2.1.2 Policy considerations

With regard to information sharing mechanisms, a number of issues stands out. Ineffective use of notification sharing mechanisms, lack of cooperation and coordination in cross-border and cross-sectoral incidents are the main impediments to building a fully resilient European cyber-shield. As observed above, a number of incident reporting schemes exist in different pieces of legislation, but a structured and harmonized knowledge sharing is missing. This prevents a meaningful and effective analysis of the information provided in incident reports to increase the resilience of critical information infrastructures.

As it can be observed in the NIS CG's report, most pressing challenges to handling incidents, especially given the cross-sectoral consequences of cyber incidents, is the lack of **mandatory cooperation** and information exchange across sectors at the national and EU levels.<sup>63</sup> But it is necessary to keep in mind that notification requirements in different legislations serve different aims. "Legal silos" (information about serious incidents stays in sectoral silos and not shared with other authorities), differences in supervision style and the competent authorities, *lex specialis* overriding requirements in Article 14 of the NISD and operational considerations are obstacles to standardised and effective information sharing mechanisms.<sup>64</sup> This also prevents the sharing of experience, cross-sector analysis and aggregation.

Still, the NIS CG identifies possibilities for synergies at national and EU-levels. At the national level, joint work by national authorities on taxonomies and reporting tool may further align reporting processes.<sup>65</sup> The NIS Cooperation Group also identifies potential for collaboration in other supervision areas in case of a common reporting tool.<sup>66</sup> At the EU level, the potential for the exchange of aggregated and anonymised information about certain incidents is identified while at the same time recognizing that different supervision approaches at national level may hinder such information sharing.<sup>67</sup>

---

<sup>62</sup> Directive 2009/136/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector [2002] OJ L 201/37 (ePrivacy Directive).

<sup>63</sup> NIS Cooperation Group, Synergies in Cybersecurity Incident Reporting (CG Publication 04/2020), available at [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=72147](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=72147).

<sup>64</sup> NIS Cooperation Group, Synergies in Cybersecurity Incident Reporting (CG Publication 04/2020), available at [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=72147](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=72147), 19.

<sup>65</sup> NIS Cooperation Group, Synergies in Cybersecurity Incident Reporting (CG Publication 04/2020), available at [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=72147](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=72147), 20.

<sup>66</sup> Ibid.

<sup>67</sup> Ibid.

### Policy recommendations

- It is recommended to facilitate sustained cooperation and collaboration between different stakeholders and sectoral, national and European bodies.
- It is recommended to focus on creating incentives for cross-border and cross-sectoral information sharing of cyber-incidents through law and policy making. These incentives can focus on aligning the economic incentives for information sharing and incident reporting, providing a harmonised structure for knowledge sharing such as a common reporting tool, improving the quality of information.
- It is recommended to private sector actors that they engage in ad-hoc information sharing. Ad-hoc information sharing constitutes a best practice for businesses.

### 3.2.1.3 The Revised NIS framework - NIS 2.0

The NIS review is considered as a crucial opportunity to bring forth closer alignment of reporting requirements. Some stakeholders during the consultation process stressed the need for an overall cross-legislative alignment of reporting authorities, thresholds, timeframes and penalties in EU legislation to eliminate ‘persisting redundancies in terms of incident reporting and double notification requirements under different legal regimes’.<sup>68</sup> Following the review of the NISD, the European Commission adopted a proposal for a revised NISD on 16 December 2020 (“Proposal for NIS 2.0”).<sup>69</sup>

The NIS 2.0 Proposal is noteworthy as it aims to introduce a higher level of **harmonisation** of reporting obligations to eliminate divergences in implementation.<sup>70</sup> The proposal addresses the concern of fragmentation of incident reporting by expanding the scope of application of the NISD to providers of public electronic communications networks or publicly available electronic communications services, and trust service providers. In doing so, it aims to streamline the legal obligations imposed on those providers in relation to the security of their network and information systems with the obligations imposed on OES and DSPs.<sup>71</sup> Further the respective competent authorities shall be enabled to benefit from the legal cooperation framework established by the NISD.<sup>72</sup>

The NIS 2.0 will also repeal the corresponding provisions laid down in the eIDAS Regulation and EECR about the imposition of security and reporting requirements.<sup>73</sup> The NIS 2.0 proposal further

---

<sup>68</sup> Roadmap NIS-Review (Position Paper, 2020), 5, available at <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12475-Cybersecurity-review-of-EU-rules-on-the-security-of-network-and-information-systems/F542104>.

<sup>69</sup> European Commission, Proposal for a Directive of the European Parliament and of the Council on measure for a high common level of cybersecurity across the Union, repealing Directive (EU) 2016/1148 (COM(2020) 823 final).

<sup>70</sup> Recitals 4 and 8, NIS 2.0 Proposal.

<sup>71</sup> Recital 48, NIS 2.0 Proposal.

<sup>72</sup> *ibid.*

<sup>73</sup> *ibid.*



introduces a two-stage approach to incident reporting in order to prevent early vulnerability disclosure when reporting an incident.<sup>74</sup> Finally, the Proposal recognises that entities may often be in a situation where an incident needs to be reported to various authorities as a result of notification obligations included in various legal instruments.<sup>75</sup> In order to alleviate these additional burdens and uncertainties with regard to format and procedures, Member States should establish a **single joint reporting body** for all notifications required under the NISD and other Union law such as in particular the GDPR.<sup>76</sup> Such a single joint reporting body becomes ever more important in light of recently enacted legislation that regulate closely related areas such as the Cybersecurity Act<sup>77</sup> and related initiatives introducing notification obligations including the proposal for the Regulation on digital operational resilience for the financial Sector (DORA)<sup>78</sup> and the Proposal for a Directive on the resilience of critical entities.<sup>79</sup>

#### Policy recommendations

- It is recommended to put in place an integrated incident handling process that serves both cybersecurity obligations and others that stem from other laws such as data protection.<sup>80</sup>
- It is recommended to establish a single joint reporting body for all notifications required under the NISD and other European Union law, in particular, the GDPR to mitigate the complications that arise from the different supervision styles embodied in different legislative frameworks.
- It is recommended to enhance cross-sector collaboration on supervision, and exchange good practices about incident reporting, for instance by harmonizing and aligning incident reporting formats across sectors and different pieces of legislation.

---

<sup>74</sup> Recital 55, NIS 2.0 Proposal.

<sup>75</sup> Recital 56, NIS 2.0 Proposal.

<sup>76</sup> *ibid.*

<sup>77</sup> 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 [2019] OJ L 151/ 15.

<sup>78</sup> European Commission, Proposal for a Regulation of the European Parliament and of the Council on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014 and (EU) No 909/2014) (COM(2020) 595 final).

<sup>79</sup> European Commission, Proposal for a Directive of the European Parliament and of the Council on the resilience of critical entities (COM(2020) 829 final).

<sup>80</sup> See, in this regard, European Data Protection Supervisor, 'Opinion 8/2022 on the Proposal for a Regulation laying down measures for a high common level of cybersecurity at the institutions, bodies, offices and agencies of the Union' 17 May 2022 available at [https://edps.europa.eu/system/files/2022-05/2022-05-17\\_opinion\\_cybersecurity\\_regulation\\_en.pdf](https://edps.europa.eu/system/files/2022-05/2022-05-17_opinion_cybersecurity_regulation_en.pdf).

## 3.2.2. Barriers for information sharing: Access to data by component authorities

### 3.2.2.1 Access to data by component authorities

The use of the CyberSANE platform will enable to collect and store vast amount of data that can reveal vital information that can help those responsible for cybercrime. The further transfer of such data with component authorities, such as law enforcement authorities will need to be performed in a lawful way.

European supranational courts make it clear in their established case law that the transmission of the obtained data from private to public entities constitute itself an interference to the right to privacy<sup>81</sup> and right to personal data protection.<sup>82</sup> Such interference is allowed if it is prescribed by law. In recent years, Court of Justice of the EU (CJEU) scrutinized the laws allowing the retention of data for the purpose of allowing the competent national authorities to have possible access to those data and applied strict requirements for such transfers.<sup>83</sup> Importantly, the law allowing data transfer from private to public entities needs to satisfy certain qualities such as foreseeability, and the interference needs to be necessary and proportionate to its purposes.

In *Digital Rights Ireland*<sup>84</sup> judgement of 2014, the CJEU scrutinized the validity of the Directive 2006/24/EC<sup>85</sup> ('Data Retention Directive'). The Directive required from EU member states to oblige communication service providers to keep **traffic data and location data** of their users for a period of at least six to 24 months, and to make available such data to national authorities for the purposes of fighting against serious crime. The Court noted that the data covered by the directive enable deducing very precise conclusions regarding people's private lives, such as the habits of everyday life, places of residence, daily movements and social relationships. The Court also confirmed the fact that subscribers or registered users could have the feeling of constant surveillance because their data are retained and subsequently used without their knowledge.<sup>86</sup> As such, data retention covered by the directive constituted a particularly serious interference with the right to privacy and right to data protection, guaranteed by Article 7 and 8 of the Charter of Fundamental Rights (CFR), respectively.

This interference would have been justified if the limitations to these rights are provided by law, respect the essence of the rights, meet objectives of general interest and fulfil the principle of proportionality.<sup>87</sup> Although the obligation to retain data was particularly serious interference with the right to privacy, it did not adversely affect the essence of this right because the directive did

---

<sup>81</sup> *Leander v. Sweden*, 26 March 1987, Series A no. 116, para. 48; *Rotaru v. Romania [GC]* (App no. 28341/95, ECHR 2000-V, para. 46; *Weber and Saravia v. Germany (dec.)*, (App. no. 54934/00, ECHR 2006-XI, para. 79.

<sup>82</sup> C-293/12 *Digital Rights Ireland*.

<sup>83</sup> Plixavra Vogiatzoglou, 'Mass Surveillance, Predictive Policing and the Implementation of the CJEU and ECtHR Requirement of Objectivity', *European Journal of Law and Technology* (2019) 10(1).

<sup>84</sup> C-293/12 *Digital Rights Ireland*.

<sup>85</sup> Directive 2006/24/EC on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks.

<sup>86</sup> C-293/12 *Digital Rights Ireland*, para. 37.

<sup>87</sup> *ibid.*, para. 38.

not allow national authorities' access to content of the electronic communications.<sup>88</sup> Data retention also satisfied an objective of general interest considering that the objective was to investigate, detect and prosecute of serious crime.<sup>89</sup>

However, the directive did not satisfy the criteria of proportionality and exceeded the limits of what is appropriate and necessary in order to achieve this objective. This is because it lacked clear and precise rules governing the extent of the interference with the right to privacy and data protection. Firstly, data retention was required in a generalized manner without any distinction being made in relation to traffic data or persons whose data can be accessed. Even electronic communications data of individuals who have no link to serious crime could be retained and accessed by authorities.<sup>90</sup> Secondly, retention periods were not distinguished based on objective criteria. Thirdly, sufficient safeguards for the protection of fundamental rights were not provided. To name a few, the directive did not lay down any objective criterion to limit the number of persons authorised to access and subsequently use the data retained; it did not require an independent review for access; it did not establish objective criteria to guarantee that data will be accessed and used only for the purposes of prevention, detection or prosecutions of (serious) crime.<sup>91</sup>

Two years later, the CJEU ruled in the same direction in *Tele 2 Sverige*<sup>92</sup>, clarifying that general, indiscriminate retention of traffic and location data of all subscribers and registered users is unlawful, whereas **targeted retention** restricted in terms of time period, geographical area and persons could be allowed. The judgement highlighted that safeguards set out earlier in Digital Rights Ireland must be satisfied. While this case law provides strong protection for fundamental rights, it may be seen as less favourable from the perspective of the effective prevention and prosecution of crimes. In practice, it would be difficult to establish objective criteria for targeted surveillance. For targeted surveillance, it would be necessary to know in advance -before data collection - which geographical area or group of persons could potentially commit crime. In practice, this would create a dilemma because one can become aware of a crime after data is collected. On the other hand, it may be easier to set objective criteria for the subsequent use of collected data. However, it has been argued that aiming for targeted subsequent use while ignoring the untargeted (initial) data gathering in private databases can create confusion in the interpretation and implementation of the objective criteria for data retention.<sup>93</sup>

In recent judgements<sup>94</sup>, the CJEU confirmed its position that national legislation cannot oblige providers of electronic communications services to carry out an untargeted retention of traffic data and location data. On the other hand, the CJEU acknowledged that general and indiscriminate data retention could be exceptionally allowed for a genuine, foreseeable and serious threat to national security, for instance to prevent a terror attack, provided that sufficient safeguards are established.<sup>95</sup>

---

<sup>88</sup> *ibid.*, para. 39.

<sup>89</sup> *ibid.*, para. 41.

<sup>90</sup> *ibid.*, para. 57-58.

<sup>91</sup> *ibid.*, para. 57-65.

<sup>92</sup> C-698/15 *Tele2 Sverige*.

<sup>93</sup> Plixavra Vogiatzoglou, 'Mass Surveillance, Predictive Policing and the Implementation of the CJEU and ECtHR Requirement of Objectivity', *European Journal of Law and Technology* (2019) 10(1).

<sup>94</sup> *Joined Cases C-511/18 La Quadrature du Net and others; C-623/17 Privacy International*.

<sup>95</sup> *Joined Cases C-511/18 La Quadrature du Net and others*, para. 139.

### 3.2.2.1.1 Policy considerations

Recent cases make it clear that general and indiscriminate retention of traffic and location data by private entities with a view to make them available to public entities are generally prohibited, and can be allowed in very exceptional case of national security. In other cases, the emphasis is put on targeted surveillance in order to restrict the surveillance measures with objective criteria such as time and space. However, difficulties involved in establishing such objective criteria will create an obstacle for the collection and further transfer of data. In that context, further policy guidance would be beneficial to ensure that the criteria of objectivity as required by the supranational courts are satisfied while ensuring accountability in the fight against cybercrime.

The cases discussed in this sub-chapter primarily concern retention laws in the telecommunications sector. As cyber threats target many other sectors, an automated system such as CyberSANE will be expectedly used in many other sectors. It will be crucial to put in place respective laws that satisfy the strict requirements established by the case-law, in order to ensure not only the data collection but also the subsequent transfer and processing will be lawful. Importantly, lawmakers and policy makers need to establish necessary safeguards. In particular, data access by component authorities should be limited to precisely defined catalogue of serious crimes, should be limited in terms of the number of persons who can access to data, and should be subject to prior independent review.

#### Policy recommendations

- It is recommended to provide further guidance on how to establish and implement objective criteria for the purposes of limiting access and subsequent use of personal data, especially in the context of the automated network security tools.
- For businesses that will deploy the CyberSANE platform, it is best practice to retain and transfer personal data to component authorities (such as law enforcement) only if there is a law in the country in which it resides, which allows such transfers, and which provide safeguards (e.g., independent review, time limits).

### 3.2.2.2 Cross-border access to digital evidence

Information collected and shared with CyberSANE will provide vital information that can reveal the identity of individuals who threaten the confidentiality and integrity of information systems. This information can potentially be used as evidence in courts in order to hold these individuals criminally liable for their wrongdoings. Evidence can constitute any information used in a criminal investigation, or presented in court in support of fact-finding.<sup>96</sup> Digital evidence --evidence stored or transmitted in digital form - can also have probative value and be used for fact-finding purposes.

---

<sup>96</sup> Sabine Gless and Pauline Pfirter, 'Cross-Border Access and Exchange of Digital Evidence: Cloud Computing Challenges to Human Rights and the Rule of Law', in *Surveillance and Privacy in the Digital Age: European, Transatlantic and Global Perspectives*, (eds) Valsamis Mitsilegas and Niovi Vavoula, Oxford: Hart Publishing, 2021, p. 10.

Information provided through malware, IP addresses, metadata embedded in documents, access log files from a service provider are examples of digital evidence.<sup>97</sup>

A significant majority of investigations today involve digital evidence. Moreover, cyber threats have very often a cross-border aspect. 85% of investigations into serious crime necessitate access to digital evidence.<sup>98</sup> National authorities often need to access to evidence held in the territory of another country in order to effectively investigate or prosecute cybercrime. For example, if a hospital in Germany suffers from a cyberattack that was originated from a server in France, German authorities would need to seek the assistance of French authorities, who can seize computer data or user information located in France and share with German authorities.

At the European and international level, there is a number of legal instruments that establishes mutual legal assistance mechanisms that facilitate judicial cooperation between countries to share digital evidence.<sup>99</sup> The **D2.2 Legal and Ethical Requirements** includes one of these important instruments in Europe, which is the Cybercrime Convention. The Convention lays down principles and procedures to ensure that the states that are party to this convention engage in mutual legal assistance concerning computer-related criminal offences, particularly in the collection of electronic evidence.<sup>100</sup>

A main feature of this classical type of mutual legal assistance mechanisms is that data located in another state is received only if that state consents to it. Only in exceptional situations, such as in the case of open-source data, consent is not required. This creates a barrier for law enforcement authorities to lawfully seize the data stored, cloud-based or otherwise, in a server in another country, even if the computer in which data can be found is in their territory.<sup>101</sup> These authorities need to seek assistance of foreign authorities to access to digital evidence, which can significantly slow down the procedures. Existing cooperation mechanisms have been subject to criticism, especially for being time-consuming and complex, thus inadequately satisfying the realities of digital age where data is constantly and rapidly on the move.<sup>102</sup> In what follows, this sub-chapter will look at the initiatives at the EU and the Council of Europe level, which aim to address the difficulties involved in the existing cooperation mechanisms.

### 3.2.2.2.1. European Union: Proposal for E-Evidence Framework

The European Commission's proposal for a new e-evidence framework was introduced in 2018 and is composed of a directive and a regulation. The aim of this framework is to make it easier and faster for law enforcement and judicial authorities to obtain the digital evidence needed for

---

<sup>97</sup> *ibid.*

<sup>98</sup> European Commission, Joint Communication to the European Parliament and the Council, The EU's Cybersecurity Strategy for the Digital Decade, 16 December 2020, JOIN(2020) 18 final.

<sup>99</sup> For an overview of the mechanisms in Europe: See Smuha, 'Towards the EU Harmonization of Access to Cross-Border E-Evidence: Challenges for Fundamental Rights & Consistency', *European Criminal Law Review* (2018) Vol. 8(1), p. 14.

<sup>100</sup> Chapter III, Cybercrime Convention.

<sup>101</sup> Nathalie Smuha, 'Towards the EU Harmonization of Access to Cross-Border E-Evidence: Challenges for Fundamental Rights & Consistency', *European Criminal Law Review* (2018) 8(1), p. 14.

<sup>102</sup> *ibid.*, p. 13-17.



investigation and potential prosecution of criminals in other EU countries. A brief description of the proposed framework has been provided in **D2.2 Legal and Ethical Requirements**.

A notable feature of this framework is that it aims to empower authorities in one EU country to ask for user information directly from certain service providers in another EU country. It aims to establish the **European Production Order**, which allows a judicial authority in one country to obtain e-evidence such as emails, text or messages in apps, as well as subscriber information directly from providers of electronic communications services, information society services, and internet domain name and Internet Protocol (IP) numbering services. It also aims to establish the **European Preservation Order** to allow direct request to a service provider in another country preserve specific data.

### 3.2.2.2 Council of Europe: Cybercrime Convention and the additional protocol

The Council of Europe - a European regional organization different than the EU<sup>103</sup>- has also taken an important step forward to tackle the challenges posed by the proliferation of cybercrime and complexities involved in obtaining digital evidence stored remotely. In May 2022, in the 20th anniversary of the Cybercrime Convention<sup>104</sup>, the Second Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence ('Second Additional Protocol') was opened for signature.

As data has no borders and can be stored in multiple or unknown jurisdictions in any given time, it is important to have enhanced international cooperation beyond the EU borders. The Additional Protocol has a potential to have a broad impact on this cross-border aspect because the Council of Europe includes, in addition to all EU member states, neighbouring countries such as the UK and Turkey. The Additional protocol is also open for signature beyond the member states, and has already been signed by Chile, Colombia and the United States.

Similar to the EU's e-evidence proposal, the Second Additional Protocol establishes direct cooperation between requesting states and service providers and registrars. For instance, component authorities in one state party will be empowered to directly request from registrars and service providers located in another state party domain name registration information<sup>105</sup> and subscriber information<sup>106</sup>. **Traffic data**, on the other hand, can be only accessed indirectly. For this type of data, national authorities will need to lodge their request to the state in which the service provider resides, who can then compel the service provider to share the requested data.<sup>107</sup> Furthermore, state parties have some discretion to apply a stricter protection regime, by leaving access to traffic data out of the scope of application of the protocol, and/or making certain type of access numbers available only through indirect access procedure.<sup>108</sup>

---

<sup>103</sup> For information on the Council of Europe, see D2.2 Legal and Ethical Requirements, p. 3.

<sup>104</sup> See D2.2 Legal and Ethical Requirements.

<sup>105</sup> Article 6, 2<sup>nd</sup> Additional Protocol.

<sup>106</sup> Article 7, 2<sup>nd</sup> Additional Protocol.

<sup>107</sup> Article 8, 2<sup>nd</sup> Additional Protocol.

<sup>108</sup> Article 7(9) and 8(13), 2<sup>nd</sup> Additional Protocol.

Importantly, state parties should establish conditions and safeguards that will adequately protect human rights and liberties.<sup>109</sup> The Second Additional Protocol provides specific conditions for the processing of personal data, which will apply unless parties are bound by another international agreement or arrangement.

Safeguards for processing personal data (Article 14)						
Purpose and use	Quality and Integrity	Sensitive data	Retention Periods	Automated decisions	Data security and security incidents	Maintaining records
Same level of protection in onward sharing within parties	Prior authorization for onward transfer to non-parties	Transparency and notice	Access and rectification	Judicial and other remedies	Oversight	Consultation

Table 2 List of safeguards for processing personal data provided by Article 14 of the Second Additional Protocol to the Cybercrime Convention

### 3.2.2.3 Policy considerations

Time-consuming and complex nature of cross-border evidence sharing procedures and increased need for cross-border access to digital evidence led to the modernization of evidence sharing procedures. In fact, both law-making initiatives in the EU and Council of Europe described above have a potential to reduce time and costs required for exchange of evidence. However, there has been criticism for the abolishment of the requirement of first seeking the assistance of another state to access a broad range of data by foreign public authorities.

Legal scholars have questioned, particularly, whether the EU's proposed e-evidence legislation provide sufficient safeguards for fundamental rights, including personal data protection. Gless and Pfirter<sup>110</sup> (2021) finds three problematic issues that lead to imbalances. Firstly, responsibility and power to enforce cross border evidence are partially transferred to profit-oriented private companies who lacks the relevant legal expertise. The EU's proposed e-evidence legislation does not make it clear whether service providers can reject the direct access request coming from another country.<sup>111</sup> This creates uncertainty as to what will happen if the access request is in conflict with the laws of the respective countries and leave the solution of this conflict to the discretion of the service provider. Secondly, individuals whose data will be shared by the service provider are not granted adequate safeguards, such as the right to challenge the disclosure of data before a court in the enforcing state (where the service provider resides). This poses a risk of leaving the individual unprotected and unremedied in case a national authority directly obtain data, by acting against domestic laws and the right to data protection, the right to respect for

---

<sup>109</sup> Article 13, Additional Protocol.

<sup>110</sup> Sabine Gless and Pauline Pfirter, 'Cross-Border Access and Exchange of Digital Evidence: Cloud Computing Challenges to Human Rights and the Rule of Law', in *Surveillance and Privacy in the Digital Age: European, Transatlantic and Global Perspectives*, (eds) Valsamis Mitsilegas and Niovi Vavoula, Oxford: Hart Publishing, 2021.

<sup>111</sup> See 3.2.2.2.1 above.



private life or the right to freedom of expression. Thirdly, the defendant does not have equal opportunity to access digital evidence to use in his or her case, which can deprive him or her of the right to equality of arms, and fair trial<sup>112</sup>. **D2.2 Legal and Ethical Requirements** has noted that the future of this proposal seems precarious, and, in fact, the proposal has not been adopted until the date of writing of this D10.4.

Against this background, it is a welcome development that the Additional Protocol to the Cybercrime Convention makes a distinction between different data types and allow direct access only to certain data categories. This allows putting in place higher protection regime for data that is more intrusive. Subscriber information (e.g. subscriber identity, access number)<sup>113</sup>, for instance, will be available through direct access procedure, while traffic data will continue to be available through state-to-state cooperation. Nevertheless, it has been noted that traffic data may also qualify as subscriber information in certain cases.<sup>114</sup> The European Data Protection Supervisor (EDPS) observes that information needed for the purpose of identifying a subscriber of a service may include certain Internet Protocol (IP) address information – for example, the IP address used at the time when an account was created, the most recent log-on IP address or the log-on IP addresses used at a specific time, which would be considered as traffic data relating to the transmission of a communication under EU law.<sup>115</sup> Recent CJEU jurisprudence shows that traffic data can only be accessed by public bodies in very exceptional conditions.<sup>116</sup> Therefore, EU member states have a possibility to make a reservation to the relevant provision to prohibit direct access to IP addresses ('certain types of access numbers') from service providers. In this context, reservation means a unilateral statement to exclude the application of a certain provision of an international treaty or change its legal effect.<sup>117</sup> If EU member states make such a statement, direct access to IP addresses from service providers in these states will not be possible. It is recommended to make this reservation so that the competent authorities can assess the necessity and proportionality before the transfer takes place on a case-by-case basis.<sup>118</sup>

Furthermore, in accordance with Article 13 of the Additional Protocol, further safeguards should be established to ensure an adequate protection of human rights and liberties. These safeguards should address issues beyond the protection of personal data, such as right to fair trial<sup>119</sup>, equality of arms, right to adversarial proceedings, and right to access and contest (digital) evidence.

---

<sup>112</sup> *ibid.*, p. 18-19.

<sup>113</sup> Cybercrime Convention defines subscriber information as any information held by a service provider, relating to subscribers of its services other than traffic or content data. Subscriber information can include the subscriber's identity, access number and payment information.

<sup>114</sup> Explanatory Report to the 2<sup>nd</sup> Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence, available at <https://rm.coe.int/1680a49c9d>.

<sup>115</sup> European Data Protection Supervisor, Opinion 1/2022 on the two Proposals for Council Decisions authorising Member States to sign and ratify, in the interest of the European Union, the Second Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence, 20 January 2022, para 95.

<sup>116</sup> See 3.2 above.

<sup>117</sup> Article 2(d), Vienna Convention on the Law of Treaties.

<sup>118</sup> European Data Protection Supervisor, Opinion 1/2022 on the two Proposals for Council Decisions authorising Member States to sign and ratify, in the interest of the European Union, the Second Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence, 20 January 2022.

<sup>119</sup> For description of these rights, See D2.2 Legal and Ethical Requirements.

Especially where service providers use automated systems to collect and process data, procedures should be established to ensure reliability of evidence (e.g. ensuring that data is not manipulated or altered). At national levels, procedural rules and protocols generally provide a means to ensure reliability<sup>120</sup>, for instance, by creating timestamps at the time when data is seized, however, it is not clear whether the same procedures will be used or adequate where data is directly collected from a service provider residing in another jurisdiction. Rules tailored to automated processing of data would be necessary to ensure effective protection of fundamental rights.<sup>121</sup>

#### **Policy recommendations**

- It is recommended to EU member states to make a reservation (a unilateral statement) to the Second Additional Protocol to the Cybercrime Convention to exclude the direct access to IP addresses ('certain types of access numbers') from service providers to make sure that traffic data can be accessed by component authorities only in necessary and proportionate circumstances in accordance with the European case-law.
- It is recommended to take legislative and policy measures to provide clear and adequate safeguards for the protection of human rights, including right to fair trial<sup>122</sup> and right to adversarial proceedings to ensure that the digital evidence involving a cross-border aspect – processed through automated or other means- is reliable.

### **3.3. Apprehensions between (cyber)security and freedom of expression: Policy considerations**

As mentioned in the Deliverable 2.2 on 'Legal and Ethical Requirements',<sup>1</sup> fundamental rights, namely freedom of expression and the right to privacy and data protection, are among the legal requirements relevant to the development and implementation of the CyberSANE system. So far, use cases of the CyberSANE system have been limited to transport, health and energy sectors.<sup>2</sup> However, the implementation of the CyberSANE system is not necessarily restricted to these sectors, as the purpose of CyberSANE is also to explore possibilities for successfully extending and applying CyberSANE results in other critical information infrastructures. Addressing the relationship between freedom of expression and cybersecurity technologies are necessary given the possibility of the wider use of the CyberSANE system. With that in mind, this section is concerned with the formulation of policy recommendations for public authorities that deal with the regulatory aspects of the fight against both cyber- attacks, risks and threats in critical information infrastructures (CIIs) from a freedom of expression perspective.

---

<sup>120</sup> Bart Custers and Lonneke Stevens, 'The Use of Data as Evidence in Dutch Criminal Courts', *European Journal of Crime, Criminal law and Criminal Justice* (2021) 29.

<sup>121</sup> *ibid.*

<sup>122</sup> For description of these rights, See D2.2 Legal and Ethical Requirements.

Part of CyberSANE's operations relies on surveillance and analysis of open source and encrypted data to proactively detect and prevent cyber-attacks. It collects and analyses communications and personal data at a mass scale targeted to find activity that indicates potential cyber-attacks. In doing so, it uses pre-determined keywords, URLs, graphics and the system updates itself to ensure the accuracy of input data. With these in mind, this section addresses surveillance related threats to freedom of expression.

The following section identifies legal considerations that stem from freedom of expression to the deployment of cyber incident handling systems across different types of critical information infrastructures (including infrastructures of different sizes and different business activities). The analysis and recommendations in this section builds on the explanations provided in Deliverable 2.2. The structure of the following section is as follows: first, it will first set the scene by establishing the tension between cybersecurity tool such as CyberSANE and freedom of expression concerns. Then, it discusses United Nations soft law documents to introduce pressing issues with regard to surveillance and human rights. Lastly, it selectively discusses European Court of Human Rights (ECtHR or the Court) judgements related to surveillance and freedom of expression.

### **3.3.1 Setting the scene: Cybersecurity and Freedom of Expression**

While cybersecurity solutions are deployed to ensure the integrity and safety of the infrastructure and for public security reasons, these initiatives may be in tension with fundamental rights and other policy areas in the highly converged digital environment. One point of tension between fundamental rights and cybersecurity relates to the confidentiality of electronic communications, including related privacy and data protection issues, and concerns about freedom of expression and other fundamental rights.<sup>123</sup> Indeed, as stated by the former United Nations (UN) Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression Frank La Rue, "privacy and freedom of expression are interlinked and mutually dependent; an infringement upon one can be both the cause and consequence of an infringement upon the other."<sup>124</sup>

David Kaye, the former UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, stated in his report that surveillance of individuals has been shown to lead to arbitrary detention of individuals exercising their freedom of expression such as journalists, activists, opposition figures.<sup>125</sup> It can also silence legitimate speech. Private

---

<sup>123</sup> This section only addresses freedom of expression concerns, but it should be kept in mind that other fundamental rights, such as the right to non-discrimination, fair trial, freedom of assembly, may also be implicated by cybersecurity solutions.

<sup>124</sup> United Nations General Assembly (UNGA), Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue, Human Rights Council Twenty-third session UN Doc A/HRC/23/40 (2013) available at [https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40\\_EN.pdf](https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf), para. 79.

<sup>125</sup> UNGA, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Human Rights Council 41st session UN Doc A/HRC/41/35 (2019) available at [https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session41/Documents/A\\_HRC\\_41\\_35.docx](https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session41/Documents/A_HRC_41_35.docx)

surveillance entities may unjustifiably give access to or share data with governments. Without doubt, compliance with the data protection and privacy laws will be relevant to safeguarding fundamental rights, including freedom of expression. This is why, in addition to the explanations below on fundamental rights protection, explanations above on sharing of and access to data and AI Act requirements should be taken into account for a full assessment of compliance with freedom of expression and other fundamental rights (see Section 3.2.2.).

As mentioned in Deliverable 2.2., freedom of expression and other fundamental rights considerations can flow from the application of Article 10 of the European Convention on Human Rights and/or the application of the Article 11 of the Charter of Fundamental Rights of the European Union when implementing the EU law. The former legal instrument principally binds contracting States of the European Convention on Human Rights. The CyberSANE system is a tool designed by private parties directed at the private sector. Horizontal application of fundamental rights can come into picture (application in disputes between private parties) as part of the positive obligations of the States. The principle of positive obligations under Articles 8<sup>126</sup> and 10<sup>127</sup> require States to take positive measures to ensure the effective exercise of freedom of expression. In addition, the CyberSANE system can also be used by public authorities in the future, in which case the State would be directly obligated to take necessary measures to ensure the compliance of the system with Article 8 and 10 of the European Court of Human Rights (ECHR). These principles must be kept in mind when considering the explanations and guidelines provided here.

### **3.3.2 Legal and Policy Framework**

#### **3.3.2.1 United Nations Soft Law Instruments: Reports of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression**

It is worth mentioning the UN Reports of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, as they set out in detail risks new technologies pose to fundamental rights, specifically, freedom of expression and the right to privacy, and provide recommendations to preserve fundamental rights in the digital age. Over the years, these principles developed by the Special Rapporteurs on the Promotion and Protection of the Right to Freedom of Opinion and Expression have been improved to a range of issues facing fundamental rights in the digital age. Below is a selection of them that relate to surveillance and freedom of expression.

---

<sup>126</sup> *Lozovyye v. Russia* (App no 4587/09) 24 April 2018.

<sup>127</sup> See, for example, *Dink v. Turkey* (App no 2668/07) 14 September 2010, para. 137.

The interrelation between freedom of expression and the right to privacy was laid out in detail in the report of the Special Rapporteur Frank La Rue,<sup>128</sup> where he analysed the implications of States' surveillance of communications on the exercise of the human rights to privacy and to freedom of opinion and expression. The report addresses national laws with inadequate fundamental rights protection in the digital age and defines the roles and responsibilities of the private sector.<sup>129</sup> The report emphasizes the role of the private sector in developing and deploying surveillance technologies, as voluntary measures deployed by private sector entities that collect and process massive amounts of data become massive repositories of personal information that are then accessible to States upon demand.<sup>130</sup> In that regard, the report recommends states to ensure that communications data collected by corporate actors in the provision of communications services meets the highest standards of data protection. Another recommendation is for the States to refrain from forcing the private sector to implement measures compromising the privacy, security and anonymity of communications services, including requiring the construction of interception capabilities for State surveillance purposes or prohibiting the use of encryption.<sup>131</sup> These recommendations preserve their validity today.

Principles laid out by Frank La Rue were further developed in specific contexts by subsequent Special Rapporteur mandates. In his thematic report "Surveillance and human rights",<sup>132</sup> the UN Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression the problem of targeted surveillance and proposes a legal and policy framework for regulation, accountability and transparency within the private surveillance industry. The report concerns itself with targeted surveillance, but the recommendations for the private sector and States to uphold human rights should be no less relevant for untargeted surveillance practices and the use of technology in that regard.<sup>133</sup> In this report, the Special Rapporteur advises

---

<sup>128</sup> UNGA Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue' Human Rights Council Twenty-third session UN Doc A/HRC/23/40 (2013) available at [https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40\\_EN.pdf](https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf)

<sup>129</sup> UNGA 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue' Human Rights Council Twenty-third session UN Doc A/HRC/23/40 (2013) available at [https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40\\_EN.pdf](https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf) para 19.

<sup>130</sup> UNGA 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue' Human Rights Council Twenty-third session UN Doc A/HRC/23/40 (2013) available at [https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40\\_EN.pdf](https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf) para 74.

<sup>131</sup> UNGA 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue' Human Rights Council Twenty-third session UN Doc A/HRC/23/40 (2013) available at [https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40\\_EN.pdf](https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf) paras 94-96.

<sup>132</sup> UNGA 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' Human Rights Council 41st session UN Doc A/HRC/41/35 (2019) available at [https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session41/Documents/A\\_HRC\\_41\\_35.docx](https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session41/Documents/A_HRC_41_35.docx)

<sup>133</sup> The European Court of Human Rights, for example, states that "While the safeguards already identified by the Court in the area of targeted interception regimes provided a useful framework, they had to be



private surveillance companies to publicly affirm their responsibility to respect freedom of expression, privacy and related human rights, integrate human rights due diligence processes into their product development and operations and establish transparency reporting.

On a final note, another report of the Special Rapporteur, acknowledging the abovementioned concerns about freedom of expression and surveillance, dedicated a full report the use of encryption and anonymity in digital communications. The report considers anonymity and encryption crucial for individuals and civil society to be protected against are subjected to interference and attack by State and non-State actors and reminds that States are obliged to protect privacy against unlawful and arbitrary interference and attacks.<sup>134</sup>

### 3.3.2.2 Council of Europe Human Rights Framework

Overtime, the ECtHR developed a vast body of principles dedicated to lawfulness of surveillance measures under Article 8.<sup>135</sup> Differently, the ECtHR case law on surveillance and freedom of expression under Article 10 is less developed than Article 8. An analysis of surveillance measures under Article 10 and freedom of expression is encountered in the context of the protection of the confidentiality of journalistic sources.<sup>136</sup> The ECtHR attaches a very strong level of protection to journalistic sources as it considers it one of the cornerstones of freedom of the press. It considers that without such protection, sources may be deterred from assisting the press in informing the public on matters of public interest.<sup>137</sup> Below is a selective discussion of some of the judgments of the ECtHR that relate to surveillance and Article 10.

In *Weber and Saravia v. Germany*,<sup>6</sup> the Court dealt with legislation allowing German Intelligence Services to conduct “strategic monitoring” of telecommunications that consisted of using “catchwords”<sup>138</sup> in order to identify and avert serious dangers facing the Federal Republic of

---

adapted to reflect the specific features of a bulk interception regime, the purpose of which was in principle preventive, rather than for the investigation of a specific target or an identifiable criminal offence.” *Centrum för rättvisa v. Sweden* [GC] (App no 35252/08) 25 May 2021.

<sup>134</sup> UNGA 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye' Human Rights Council 29th session UN Doc A/HRC/29/32 (2015) available at <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G15/095/85/PDF/G1509585.pdf?OpenElement> para 32.

<sup>135</sup> See Paul De Hert and Gianclaudio Malgieri, 'Article 8 ECHR Compliant and Foreseeable Surveillance: The ECtHR's Expanded Legality Requirement Copied by the CJEU. A Discussion of European Surveillance Case Law' Brussels Privacy Hub Working Paper (2020) 6(21) available at <https://ssrn.com/abstract=3544017>. See also European Court of Human Rights, 'Mass surveillance Fact Sheet' (2022) available at [https://www.echr.coe.int/documents/fs\\_mass\\_surveillance\\_eng.pdf](https://www.echr.coe.int/documents/fs_mass_surveillance_eng.pdf).

<sup>136</sup> See Section 3.3.3. Surveillance and chilling effects and journalism in Ronan Ó Fathaigh 'Article 10 and the chilling effect: a critical examination of how the European Court of Human Rights seeks to protect freedom of expression from the chilling effect' (2019) available at <https://biblio.ugent.be/publication/8620369>.

<sup>137</sup> *Goodwin v. the United Kingdom* (App. no. 17488/90) 27 March 1996 (Grand Chamber) para 39.

<sup>138</sup> The German law in question allowed both individual and strategic monitoring: the former is defined as the interception of telecommunications of specific persons, that serves to avert or investigate certain grave offences which the persons monitored are suspected of planning or having committed, whereas the latter aims at collecting information by intercepting telecommunications in order to identify and avert serious dangers, such as an armed attack on its territory or the commission of international terrorist attacks and certain other serious offences. *Weber and Saravia v. Germany* (App. no. 54934/00) 29 June 2006 (Admissibility decision) para 4. The main novelty of strategic monitoring is the use of catchwords. According

Germany. One of the applicants was a journalist and the scope of strategic monitoring included subjects about which the applicant journalist wanted to interview on.<sup>139</sup> There was a “danger” her telecommunications for journalistic purposes might be monitored and that her journalistic sources “might be either disclosed or deterred from calling or providing information by telephone”, implying chilling effects of the “strategic monitoring” on her exercise of freedom of expression.<sup>140</sup> Consequently, the Court decided that there was an interference with her right to freedom of expression “irrespective of any measures actually taken against her.”<sup>141</sup> The Court then found this interference to be prescribed by law,<sup>142</sup> and pursued the legitimate aim of protection of the interests of national security.<sup>143</sup>

The Court found the law in question was not in violation of Article 10. In the Court’s view, since the surveillance measures were not directed at uncovering journalistic sources the interference with freedom of expression by means of strategic monitoring could not be characterised as particularly serious.<sup>144</sup> The strategic monitoring was carried out in order to prevent certain offences, and was not aimed at monitoring journalists and generally the authorities would know only when examining the intercepted telecommunications, if at all, that a journalist’s conversation had been monitored.<sup>145</sup> Despite lacking special rules safeguarding the protection of freedom of the press and, in particular, the non-disclosure of sources, contained numerous safeguards to keep the interference with the secrecy of telecommunications – and therefore with the freedom of the press – within the limits of what was necessary to achieve the legitimate aims pursued.<sup>146</sup>

The aim of CyberSANE is not to target certain individuals or journalists. Nevertheless, its DarkNet component collects data and analyses articles from news sites, social media and the World Wide Web in order to raise awareness about published articles, topics discussed in forums related with cyber-security incidents.<sup>147</sup> Considering the aforementioned principles developed in *Weber and Saravia v. Germany*, the use of a cybersecurity solution such as the CyberSANE system by public authorities, even if it is not directed at the press, may constitute an interference with Article 10. But its use can be legitimized if adequate safeguards exist to ensure interference with the confidentiality of communications necessary to the aim pursued. These following criteria were determinant to evaluating the necessity of the interference with confidentiality of communications: the nature of offences which may give rise to an interception order; a definition of the categories of people liable to have their communications intercepted; a limit on the duration of interception; the procedure to be followed for examining, using and storing the data obtained; the precautions

---

to the new G10 Act such catchwords cannot contain distinguishing features that allowed the interception of specific telecommunications and had to be listed in the monitoring order *Weber and Saravia v. Germany* (App. no. 54934/00) 29 June 2006 (Admissibility decision) para 40.

<sup>139</sup> *Weber and Saravia v. Germany* (App. no. 54934/00) 29 June 2006 (Admissibility decision), para. 145.

<sup>140</sup> *ibid.*, para. 146.

<sup>141</sup> *ibid.*, para. 144.

<sup>142</sup> *ibid.*, para. 147.

<sup>143</sup> *ibid.*, para. 149.

<sup>144</sup> *ibid.*, para. 151.

<sup>145</sup> *ibid.*

<sup>146</sup> *ibid.*, para. 152.

<sup>147</sup> CyberSANE\_D4.3\_Specification of the Deep and Dark Web mining and intelligence\_v1.0



to be taken when communicating the data to other parties; and the circumstances in which intercepted data may or must be erased or destroyed.<sup>148</sup>

The case of *Big Brother Watch and Others v. UK*<sup>149</sup> is the first decision on bulk surveillance in post-Snowden era, the Court also analysed whether the surveillance regimes for in question provided sufficient safeguards to protect the confidentiality of journalistic sources under Article 10 of the ECHR. The surveillance regime at hand in the case of Big Brother Watch consisted of three surveillance practices: international data sharing practices of the UK secret services, collection of data amongst service providers and bulk data surveillance.

Importantly, *Big Brother Watch and Others v. the United Kingdom* do not declare a bulk interception regime in and of itself violated the Convention.<sup>150</sup> Interception regimes are a “valuable technological capacity to identify new threats in the digital domain”,<sup>151</sup> according to the Court. They may be necessary in the investigation of national security threats and serious crime, including in the context of global terrorism, cyber-attacks, counter-espionage, election interferences, drug trafficking, and child pornography, adding that “hostile State and non-State actors to disrupt digital infrastructure and even the proper functioning of democratic processes through the use of cyberattacks, a serious threat to national security which by definition exists only in the digital domain and as such can only be detected and investigated.”<sup>152</sup> The Court further assess the compatibility of the surveillance regime as to the existence of safeguards against arbitrariness and abuse, without requiring full transparency on how bulk interception regime operates “on the basis of limited information”.<sup>153</sup>

However, such a regime had to be subject to “end-to-end safeguards”,<sup>154</sup> meaning that, at the domestic level, an assessment should be made at each stage of the process of the necessity and proportionality of the measures being taken; that bulk interception should be subject to independent authorization at the outset, when the object and scope of the operation were being defined; and that the operation should be subject to supervision and independent ex post facto review.

The Court found the bulk interception of communications regime to be in violation of Article 10 considering the lack of safeguards against abuse of power. The UK law governing the bulk interception of communications had contained no requirement that the use of selectors or search terms known to be connected to a journalist be authorized by a judge or other independent and

---

<sup>148</sup> Weber and Saravia v. Germany (App. no. 54934/00) 29 June 2006 (Admissibility decision), para. 95

<sup>149</sup> *Big Brother Watch and Others v the United Kingdom* [GC] (App nos 58170/13 62322/14 24960/15) 25 May 2021.

<sup>150</sup> *Big Brother Watch and Others v the United Kingdom* [GC] (App nos 58170/13 62322/14 24960/15) 25 May 2021.

<sup>151</sup> *Big Brother Watch and Others v the United Kingdom* [GC] (App nos 58170/13 62322/14 24960/15) 25 May 2021 para. 323.

<sup>152</sup> *Big Brother Watch and Others v the United Kingdom* [GC] (App nos 58170/13 62322/14 24960/15) 25 May 2021 para. 323.

<sup>153</sup> *Big Brother Watch and Others v the United Kingdom* [GC] (App nos 58170/13 62322/14 24960/15) 25 May 2021 para. 323.

<sup>154</sup> *Big Brother Watch and Others v the United Kingdom* [GC] (App nos 58170/13 62322/14 24960/15) 25 May 2021 para. 350.

impartial decision-making body.<sup>155</sup> Under the UK bulk interception regime, confidential journalist material could have been accessed by the intelligence services intentionally, through the deliberate use of selectors or search terms connected to a journalist or news organization. At the same time, confidential journalist material could be accessed unintentionally, as a “bycatch” of the bulk interception operation; in such case, the degree of interference with journalistic communications and/or sources could not be predicted at the outset. In such instances, there had been no safeguards to ensure that it could only continue to be stored and examined by an analyst if authorised by a judge or another independent decision-making body with the power to determine whether continued storage and examination was “justified by an overriding requirement in the public interest”.<sup>156</sup> The Court did not find the regime for receiving intercept material from foreign governments and/or communications service providers did not breach Article 10 of the Convention.

### 3.3.2.3 Policy considerations

There is a wide range of regulatory, policy and judicial interventions directed at aligning surveillance practices, which may also include cybersecurity solutions, with human rights protections, including freedom of expression and the right to privacy. This section provided a selective insight into these interventions, based on the reports of the UN Special Rapporteurs and the case law of the ECtHR. All cybersecurity initiatives at the development stage and in its deployment by public or private entities and laws that enable their use must comply with international human rights standards some of which outlined above.

As mentioned above, cybersecurity solutions that entail surveillance of communications by public authorities is not *per se* prohibited under the European Convention on Human Rights (ECHR). This means that cybersecurity solutions may be deployed for legitimate aims such as prevention of crime or national security. Still, a mere possibility that confidentiality of journalistic communications may be impaired may be sufficient to find an interference with the right to freedom of expression under Article 10. Because there may be a threat of breach of confidentiality of journalistic material, the press may be deterred from exercising its function as the public watchdog in a democratic society. Cybersecurity solutions must be designed in a way to accommodate procedural safeguards to ensure that interference with confidentiality of communications is strictly necessary to the aim pursued.

Owing to technological developments, surveillance which was not targeted directly at individuals can have the capacity to have a very wide reach. National laws conferring powers to authorities to deploy cybersecurity solutions must contain robust safeguards regarding the storage, examination, use, onward transmission and destruction of such confidential material assessed in an ongoing manner. These may also include specific safeguards to ensure the confidentiality of journalistic sources in particular, the lack of which was an important factor in the finding the

---

<sup>155</sup> Big Brother Watch and Others v the United Kingdom [GC] (App nos 58170/13 62322/14 24960/15) 25 May 2021.

<sup>156</sup> Big Brother Watch and Others v the United Kingdom [GC] (App nos 58170/13 62322/14 24960/15) 25 May 2021 para. 351.

surveillance regime in *Big Brother Watch and Others v. the United Kingdom* in violation of Article 10.

As it can be observed from the Court's decision in *Big Brother Watch and Others v. the United Kingdom*, the compatibility of mass surveillance with Article 10 is defined by the existence of procedural safeguards but perhaps their proportionality, functionality and effectiveness are not subject to sufficient scrutiny. This approach can be criticized, as the Court's emphasis on procedural safeguards lies on top of its finding that bulk surveillance does not in itself violate the ECHR. Instead, bulk surveillance through technology conducted for, among others, national security reasons can be made subject to stricter conditions. The Court's assessment of surveillance practices can recognize the societal harms that are associated with subjecting populations in bulk to surveillance (e.g. chilling effects on speech).

At the same time, with increasing surveillance practices to tackle cyberthreats, technological developments and the use of AI solutions in doing so, it is necessary to assess procedural safeguards in the light of the state-of-the-art technology and potential risks they may pose to fundamental rights. In addition to the state-of-the-art technology, laws that concern cybercrime, privacy, data protection and information sharing with the law enforcement and other public authorities in the light of freedom of expression to ensure the highest level of protection of fundamental rights.

#### **Policy Recommendations**

- It is recommended that laws that incentivize and foresee the use of cybersecurity solutions must implement procedural safeguards to ensure that surveillance measures are used only to that extent that it is strictly necessary. Appropriate safeguards must be provided in laws to ensure the confidentiality of journalistic sources at all times, even where a measure is not directed at surveillance of journalists.
- It is recommended that the private sector development and deployment of cybersecurity solutions must accommodate procedural safeguards to ensure the protection of freedom of expression.
- It is recommended that cybersecurity solutions that entail surveillance of communications must comply with privacy and data protection laws. Any interference with the right to privacy under Article 8 of the ECHR must be necessary and pursue a legitimate aim. It is recommended to evaluate all legislative framework that concern cybercrime, privacy, data protection and information sharing with the law enforcement and other public authorities in the light of freedom of expression.

### **3.4 Lessons learned derived by the CyberSANE System**

The CyberSANE provides a platform that collects, compile and analyse a large scale of data coming from both structured and unstructured sources. It integrates data mining and machine

learning techniques such as deep learning (the so-called artificial intelligence) to make predictions about anomalies and cyber incidents. One of the lessons learned from the CyberSANE project is that the existence of clear rules that technological solutions should comply with facilitates their development and use. An analysis of the legal framework and the platform reveal uncertainties about the rules applicable to the AI-based components of the platform. More specifically, this deliverable discussed that because of the restrictive wording of the proposed AI legislation, the platform or some of its components may or may not qualify as high risk AI depending on the sector in which it will be deployed and the interpretation of their function (if and when this legislation comes into force).<sup>157</sup> In the future, this could create legal uncertainty and implementation costs for the providers to determine whether they should comply with the AI requirements. As a result, such uncertainty could be a challenge for the replicability and wider use of AI-based systems in different sectors. It is therefore recommended to avoid these uncertainties by providing clear guidelines and standards for the developers and service providers.

Another lessons-learned emanating from the CyberSANE research is that the structured and harmonized knowledge sharing mechanisms can increase the resilience of critical information infrastructures. Lack of cooperation and coordination in cross-border and cross-sectoral incidents can create obstacle for building a fully resilient European cyber-shield.<sup>158</sup> Further research can help to tackle this obstacle on local and international level. It is therefore recommended to incentivize further (interdisciplinary) research on the creation of a harmonized and automatized information sharing mechanisms in critical infrastructures.

Last but not least, CyberSANE project has put substantial efforts to produce training materials on cybersecurity. A related lesson-learned is that education and training in the development and use of the system facilitate the legal and ethical compliance. Legally and ethically compliant future use of the platform will depend on the involvement of actors with necessary skills and expertise who can assess the ethical use of (advanced) technologies.<sup>159</sup> In order to sustain the project's impact on the long run, it is recommended to increase and incentivize education and training on different parts of the society (business, academia, society at large).

---

<sup>157</sup> For a detailed discussion on this, see 3.2.2 above.

<sup>158</sup> For a detailed discussion on this, see 3.3 above.

<sup>159</sup> For a detailed discussion on this, see 3.2.1.1 above.

## 4 Conclusion

This Deliverable provided an assessment of the CyberSANE platform from a legal and ethical perspective. It demonstrated how CyberSANE has integrated the legal and ethical requirements identified in earlier stages of the project. In order to ensure sustainable use of new technologies, it has been crucial to identify any potential risks and mitigate them as early as possible in the development stage. Having this in mind, the CyberSANE project has adopted a privacy and data protection by design approach and an ethics-by-design approach and implemented technical and organizational measures.

Furthermore, this deliverable provided an assessment of the legal and ethical framework from a policy perspective in the area of the regulation of critical infrastructure AI, information and evidence sharing with public and private entities, and freedom of expression. It formulated policy recommendations for public authorities dealing with regulatory aspects of critical infrastructure protection, as well as best practices for businesses that will likely deploy the CyberSANE platform or any other automated means for network security. A list of these best practices and policy recommendations are further provided below. This deliverable highlights that it is of outmost importance to provide clear and harmonized rules for emerging technologies designed to be used in critical infrastructures. Legal certainty and the consistent application of rules and safeguards across sectors will eliminate regulatory burden for both businesses and supervisory authorities and ensure effective protection for individuals.

### **Best practices**

- Adherence to international standards to put in place adequate technical means or measures.
- Establishment of privacy-friendly policies for the day-to-day organization's operation
- Provision of training to employees on privacy, data protection and security aspects, as well as ethical aspects concerning the new technologies such artificial intelligence.
- Inclusion of a trustworthiness assessment for the use AI systems in the organization's operations.
- Establishment of strong access control systems and provision of restricted access to authorized persons.
- Retaining and transferring personal data to component authorities (such as law enforcement) only if there is a law in the country in which it resides, which allows such transfers, and which provide safeguards (e.g. independent review, time limits).
- Engagement in ad-hoc information sharing.

### **Policy recommendations**

- It is recommended to amend the Assessment List for Trustworthy AI to eliminate the so-called 'yes or no questions' and include questions that ask an explanation or justification to the answer provided.

- It is recommended to provide further guidance or clarification to those who will be involved in the assessment process, by providing examples or giving further information about the next steps.
- It is recommended to put in place policy initiatives to support the training and education of involved stakeholders.
- It is recommended to eliminate the sectoral inconsistencies to ensure a consistent application of AI-related requirements across all critical infrastructures.
- It is recommended to expand the list of high-risk AI systems in the proposed AI Act or through delegated acts that can be adopted after the entry into force of the proposed AI Act to cover all AI systems with significant fundamental rights impacts.
- It is recommended to take any other legislative or policy measures to ensure that all fundamental rights risks are taken into account in the development and use of AI systems.
- It is recommended to provide further guidance on how to establish and implement objective criteria for the purposes of limiting access and subsequent use of personal data, especially in the context of the automated network security tools.
- It is recommended to EU member states to make a reservation (a unilateral statement) to the Second Additional Protocol to the Cybercrime Convention to exclude the direct access to IP addresses ('certain types of access numbers') from service providers to make sure that traffic data can be accessed by component authorities only in necessary and proportionate circumstances in accordance with the European case-law.
- It is recommended to take legislative and policy measures to provide clear and adequate safeguards for the protection of human rights, including right to fair trial and right to adversarial proceedings to ensure that the digital evidence involving a cross-border aspect – processed through automated or other means- is reliable.
- It is recommended to facilitate sustained cooperation and collaboration between different stakeholders and sectoral, national and European bodies.
- It is recommended to focus on creating incentives for cross-border and cross-sectoral information sharing of cyber-incidents through law and policy making. These incentives can focus on aligning the economic incentives for information sharing and incident reporting, providing a harmonised structure for knowledge sharing such as a common reporting tool, improving the quality of information.
- It is recommended to put in place an integrated incident handling process that serves both cybersecurity obligations and others that stem from other laws such as data protection.
- It is recommended to establish a single joint reporting body for all notifications required under the NISD and other Union law, in particular, the GDPR to mitigate the complications that arise from the different supervision styles embodied in different legislative frameworks.
- It is recommended to enhance cross-sector collaboration on supervision, and exchange good practices about incident reporting, for instance by harmonizing and aligning incident reporting formats across sectors and different pieces of legislation.



- It is recommended that laws that incentivize and foresee the use of cybersecurity solutions must implement procedural safeguards to ensure that surveillance measures are used only to that extent that it is strictly necessary. Appropriate safeguards must be provided in laws to ensure the confidentiality of journalistic sources at all times, even where a measure is not directed at surveillance of journalists.
- It is recommended that the private sector development and deployment of cybersecurity solutions must accommodate procedural safeguards to ensure the protection of freedom of expression.
- It is recommended that cybersecurity solutions that entail surveillance of communications must comply with privacy and data protection laws. Any interference with the right to privacy under Article 8 of the ECHR must be necessary and pursue a legitimate aim.
- It is recommended to evaluate all legislative framework that concern cybercrime, privacy, data protection and information sharing with the law enforcement and other public authorities in the light of freedom of expression.



## 4 List of Abbreviations

Abbreviation	Translation
AI	Artificial Intelligence
HLEG AI	High-Level Expert Group on Artificial Intelligence
ALTAI	Assessment List for Trustworthy Artificial Intelligence
API	Application Programming Interfaces
CI(s)	Critical Infrastructure(s)
CII(s)	Critical Information Infrastructure(s)
CJEU	Court of Justice of the EU
CG	Cooperation Group
CSIRT(s)	Computer Security Incident Response Team(s)
DSP	Digital Service Providers
EBA	European Banking Authority
ECB	European Central Bank
ECHR	European Convention on Human Rights
ECtHR	European Court of Human Rights
EDPS	European Data Protection Supervisor

D10.4 - Evaluation and Benchmarking Methodology Best Practices and Policy Development Guidelines for Replicability and Wider Use

EEEC	European Electronic Communications Code
e-evidence	Electronic evidence
e-ID	Electronic Identification
eIDAS Regulation	The regulation for electronic identification and trust services
ENISA	European Union Agency for Network and Information Security
EU	European Union
ePrivacy	Electronic privacy
GDPR	General Data Protection Regulation
IP	Internet Protocol
ISO	International Organization for Standardization
NCA	National Competent Authority
MDR	Medical Device Relation
NIS	Network and Information Systems
NIS 2.0	Proposal for a Directive of the European Parliament and of the Council on measure for a high common level of cybersecurity across the Union, repealing Directive (EU) 2016/1148 (COM(2020) 823 final).
NISD	NIS Directive
NRA	National Telecom Regulatory Authorities
OES	Operators of Essential Services

D10.4 - Evaluation and Benchmarking Methodology Best Practices and Policy Development Guidelines for Replicability and Wider Use

p.	page
para.	paragraph
PSD2 Directive	Directive (EU) of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC [2015] OJ L 337/35
UN	United Nations
UNGA	United Nations General Assembly
WP29	Article 29 Data Protection Working Party

## 5 Bibliography

Biasin E. and Kamenjasevic E., 'Cybersecurity of medical devices: new challenges arising from the AI Act and NIS 2 Directive proposals', forthcoming (2022)

Custers B. and Stevens L., 'The Use of Data as Evidence in Dutch Criminal Courts', *European Journal of Crime, Criminal law and Criminal Justice* (2021) 29

De Hert P. and Malgieri G., 'Article 8 ECHR Compliant and Foreseeable Surveillance: The ECtHR's Expanded Legality Requirement Copied by the CJEU. A Discussion of European Surveillance Case Law', *Brussels Privacy Hub Working Paper* (2020) 6(21)

ENISA, *Telecom Security Incidents 2020 Annual Report*, 26 July 2021, available at <https://www.enisa.europa.eu/publications/telecom-annual-incident-reporting-2020>

ENISA, *Trust Services Security Incidents 2020 - Annual Report*, 26 July 2021, available at <https://www.enisa.europa.eu/publications/trust-services-security-incident-2020-annual-report>

European Banking Authority, *Guidelines on major incidents reporting under PSD2*, 10 June 2021, available at [https://www.eba.europa.eu/sites/default/documents/files/document\\_library/Publications/Guidelines/2021/Guidelines%20on%20major%20incident%20reporting%20under%20PSD2%20EBA-GL-2021-03/1014562/Final%20revised%20Guidelines%20on%20major%20incident%20reporting%20under%20PSD2.pdf](https://www.eba.europa.eu/sites/default/documents/files/document_library/Publications/Guidelines/2021/Guidelines%20on%20major%20incident%20reporting%20under%20PSD2%20EBA-GL-2021-03/1014562/Final%20revised%20Guidelines%20on%20major%20incident%20reporting%20under%20PSD2.pdf).

European Commission, *Artificial Intelligence for Europe* {SWD(2018) 137 final}, 25 April 2018 COM(2018) 237 final, available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN>

European Commission, *Ethics By Design and Ethics of Use Approaches for Artificial Intelligence*, 25 November 2021, available at [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence\\_he\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf)

European Commission, *Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence*, 21 April 2021, available at [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_21\\_1682](https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682)

European Commission, *Joint Communication to the European Parliament and the Council, The EU's Cybersecurity Strategy for the Digital Decade*, 16 December 2020, JOIN(2020) 18 final

European Commission, EU Electronic Communications Code: Commission refers 10 Member States to the Court of Justice of the EU, 6 April 2022, available at <[https://ec.europa.eu/commission/presscorner/detail/en/IP\\_22\\_1975](https://ec.europa.eu/commission/presscorner/detail/en/IP_22_1975)

European Council, Cover Note from General Secretariat of the Council to Delegations, EUCO 14/17, 19 October 2017, available at <https://data.consilium.europa.eu/doc/document/ST-14-2017-INIT/en/pdf>

European Data Protection Board, 'Guidelines 05/2020 on consent under Regulation 2016/679', 4 May 2020

European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default', 20 October 2020

European Data Protection Supervisor, Opinion 1/2022 on the two Proposals for Council Decisions authorising Member States to sign and ratify, in the interest of the European Union, the Second Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence, 20 January 2022

European Data Protection Supervisor, Opinion 8/2022 on the Proposal for a Regulation laying down measures for a high common level of cybersecurity at the institutions, bodies, offices and agencies of the Union, 17 May 2022, available at [https://edps.europa.eu/system/files/2022-05/2022-05-17\\_opinion\\_cybersecurity\\_regulation\\_en.pdf](https://edps.europa.eu/system/files/2022-05/2022-05-17_opinion_cybersecurity_regulation_en.pdf).

Fathaigh R. O., 'Article 10 and the chilling effect : a critical examination of how the European Court of Human Rights seeks to protect freedom of expression from the chilling effect' (2019) available at <<https://biblio.ugent.be/publication/8620369>

Gless S. and Pfirter P., 'Cross-Border Access and Exchange of Digital Evidence: Cloud Computing Challenges to Human Rights and the Rule of Law', in *Surveillance and Privacy in the Digital Age: European, Transatlantic and Global Perspectives*, (eds) Valsamis Mitsilegas and Niovi Vavoula, Oxford: Hart Publishing, 2021

HLEG AI, Ethics Guidelines for Trustworthy AI, 8 April 2019, available at <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

HLEG AI, The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment, 17 July 2020, available at <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence>

Kop M., 'EU Artificial Intelligence Act: The European Approach to AI', Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University 2(2021)

NIS Cooperation Group, Synergies in Cybersecurity Incident Reporting (CG Publication 04/2020), available at [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=72147](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=72147)

Roadmap NIS-Review (Position Paper, 2020), 5, available at <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12475-Cybersecurity-review-of-EU-rules-on-the-security-of-network-and-information-systems/F542104>

UNGA, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, Human Rights Council 29th session UN Doc A/HRC/29/32 (2015) available at <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G15/095/85/PDF/G1509585.pdf?OpenElement>

UNGA, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue, Human Rights Council Twenty-third session UN Doc A/HRC/23/40 (2013) available at [https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40\\_EN.pdf](https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf)

UNGA, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Human Rights Council 41st session UN Doc A/HRC/41/35 (2019) available at [https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session41/Documents/A\\_HRC\\_41\\_35.docx](https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session41/Documents/A_HRC_41_35.docx)

Schmitz-Berndt S. and Anheier F., 'Synergies in Cybersecurity Incident Reporting – The NIS Cooperation Group Publication 04/20 in Context', European Data Protection Law Review (2021) 7(1)

Smuha N., Ahmed-Rengers E., Harkens A., Li W., MacLaren J., Pisellif R. and Yeung K., 'How the EU can achieve legally trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence', LEADs Lab, University Birmingham, 5 August 2021

Smuha N., Towards a Practical Assessment Tool for Trustworthy AI, Presentation at the European AI Week 2022, 15 March 2022, available at <https://www.youtube.com/watch?v=tb47bUIKPec&t=858s>

Smuha N., 'Towards the EU Harmonization of Access to Cross-Border E-Evidence: Challenges for Fundamental Rights & Consistency', European Criminal Law Review (2018) 8(1)

Vogiatzoglou P., 'Mass Surveillance, Predictive Policing and the Implementation of the CJEU and ECtHR Requirement of Objectivity', European Journal of Law and Technology (2019) 10(1)