UNIVERSITÀ DEGLI STUDI DI MACERATA

PHD PROGRAM IN
GLOBAL STUDIES. JUSTICE, RIGHTS, POLITICS

CICLO XXXVII

TITLE OF THE THESIS

Navigating Ethical Challenges in Cybersecurity: From Risk Assessment to Quantum-AI Applications

THESIS SUPERVISORS
Prof. Benedetta Giovanola
Prof. Marco Baldi

PHD CANDIDATE
Ludovica Ilari

COORDINATOR OF THE PHD PROGRAM
Prof. Benedetta Barbisan

YEAR 2023/2024

SDA
scuola di dottorato

Table of Contents

Tables of Figures

Tables of Tables

Acknowledgements

I would like to express my deepest gratitude to everyone who has supported me throughout this research journey, making this thesis possible.

First and foremost, I extend my heartfelt thanks to my advisors, Prof. Benedetta Giovanola and Prof. Marco Baldi, whose guidance, insight, and encouragement have been invaluable at every stage. Their expertise and constructive feedback have greatly contributed to shaping this work and expanding my understanding of the subject.

I am also immensely grateful to the faculty and staff of University of Macerata, who provided the resources, knowledge, and environment needed to undertake this research. Special thanks go to my colleagues and friends who offered constant support and valuable perspectives.

Finally, I would like to acknowledge the support received from BiMind, which has been instrumental in allowing me to dedicate myself fully to this research. Their commitment to fostering education and innovation has inspired me to pursue excellence in my studies.

To all who have contributed to this journey, directly or indirectly, I am deeply appreciative. This accomplishment reflects the support and encouragement I have received from each of you.

Thank you.

Abstract

This research, titled *"Navigating Ethical Challenges in Cybersecurity: From Risk Assessment to Quantum-AI Applications,"* addresses the ethical dimensions of cybersecurity and data protection within the context of advancing technologies.

Beginning with a foundational examination of ethical frameworks in cybersecurity, chapter 1 explores cyber risks, data protection, and the adaptive frameworks needed to address emerging challenges. It delves into specific ethical considerations for adversarial machine learning and predictive artificial intelligence, assessing how these technologies raise unique ethical risks.

Chapter 2 presents a structured approach to ethical risk assessment, with methodologies divided into defensive and offensive strategies. This includes both qualitative and quantitative models tailored for data protection, IT security, and AI, emphasizing their application within European regulatory frameworks. Additionally, the study addresses the ethical biases that can arise in AI-driven filtering systems, illustrating the complexities of ensuring fairness and transparency.

In Chapter 3, the study evaluates practical mitigation strategies, including blockchain solutions for ethical cybersecurity, with a focus on preventing DDoS and ransomware attacks. Business Intelligence (BI) tools, particularly Microsoft Power BI, are highlighted for their role in managing ethical risks in data protection, supported by a healthcare case study. This chapter also examines countermeasures for adversarial attacks in machine learning, showcasing adversarial de-biasing as an effective response.

Chapter 4 shifts the focus to the ethical and security implications of Quantum AI. This section explores hybrid quantum machine learning, identifying cybersecurity vulnerabilities and ethical considerations in this field. Discussions include post-quantum and quantum cryptographic approaches, ethical challenges in quantum cryptography, and defense strategies tailored for quantum machine learning applications. The chapter also considers data protection strategies within hybrid quantum systems.

The study concludes by underscoring the importance of adaptive ethical frameworks as technologies evolve, from AI to quantum applications. Through a comprehensive synthesis of cybersecurity ethics, the research emphasizes the need for vigilant risk assessment and robust ethical guidance in the face of emerging innovations.

Appendices provide supplementary code in R and DAX, supporting the methodologies and tools discussed throughout the study.

Introduction

Cybersecurity is a broad term encompassing the activities necessary to protect network and information systems, the users of these systems, and others potentially affected by cyber threats [1]. It is more closely aligned with system security, which aims to safeguard the integrity, availability, and confidentiality of technological infrastructures, rather than information security, which primarily deals with data protection (Figure 0-1). This distinction highlights that cybersecurity involves not only protecting data but also ensuring the functionality and resilience of the underlying systems that support critical digital operations.

Cybersecurity risks arise from the failure to protect systems at various levels—technical, organizational, and human—leading to a range of potential attacks such as malware, man-in-the-middle, denial of service, brute force, zero-day exploits, social engineering, or web application breaches. These vulnerabilities do not just threaten traditional systems but can also serve as entry points for targeting artificial intelligence (AI) systems. When AI is deployed in security applications, cybercriminals can exploit its weaknesses through adversarial attacks, where carefully manipulated inputs deceive the AI into making incorrect decisions, thus compromising the reliability and effectiveness of AI-based defenses.

As the world undergoes a digital transformation, characterized by the widespread adoption of big data, the Internet of Things (IoT), Industry 4.0, and AI, the surface area for potential cyber-attacks has expanded considerably. The interconnected nature of these technologies increases the opportunities for cybercriminals to exploit weaknesses, especially in AI systems that often lack robust defenses against sophisticated threats. This expansion of the attack surface underscores the need for comprehensive cybersecurity measures that go beyond traditional technical solutions.

In this evolving digital landscape, cybersecurity is more critical than ever for protecting sensitive information, ensuring business continuity, and maintaining national security. However, despite the growing importance of cybersecurity, there is still a significant lack of ethical frameworks to guide the implementation of security practices. This deficiency poses risks such as the erosion of privacy, the potential misuse of technology, and a lack of accountability in how cybersecurity measures are applied. In the absence of clear ethical guidelines, cybersecurity decisions may fail to consider their wider societal implications, potentially undermining public trust and jeopardizing individuals' rights. Moreover, the advent of the quantum computing era adds further complexity to the cybersecurity landscape. Quantum technologies threaten the effectiveness of current cryptographic methods, as

quantum computers could potentially break many of today's widely used encryption algorithms. This makes it crucial to develop new quantum-resistant encryption methods to ensure data remains secure in the future.

As these technological advancements continue to unfold, the lack of standardized cybersecurity certifications and shared ethical guidelines undermines confidence in technology, further emphasizing the need for a comprehensive approach that addresses both technical and ethical risks. To tackle these multifaceted challenges, it is essential to adopt a multidisciplinary approach to cybersecurity that considers ethical, legal, and technical dimensions. Such an approach should not only focus on developing ethical frameworks but also prioritize evaluating the ethical risks associated with the deployment of various technologies and implementing measures to mitigate these risks. These steps form the foundation of an ethical risk assessment, which comprises three key pillars: risk identification through frameworks, risk measurement using mathematical models, and risk mitigation with the help of innovative technological tools and mathematical methods. Incorporating these elements ensures that cybersecurity practices are not only effective in countering threats but also align with broader societal values and respect human rights.

The three chapters of the thesis address the three pillars of risk assessment: risk identification through frameworks (chapter 1), risk measurement through mathematical models (chapter 2), and risk mitigation through the utilization of both innovative technological tools and mathematical methods (chapter 3). While this structured approach facilitates a clear understanding of each phase's objectives, interdependencies between phases are also explored within individual chapters, reflecting the integrated nature of ethical risk assessment in practice.

The fourth chapter will deal with the future cyber risks of the quantum era and the related ethical aspects, with a particular focus on Quantum AI.

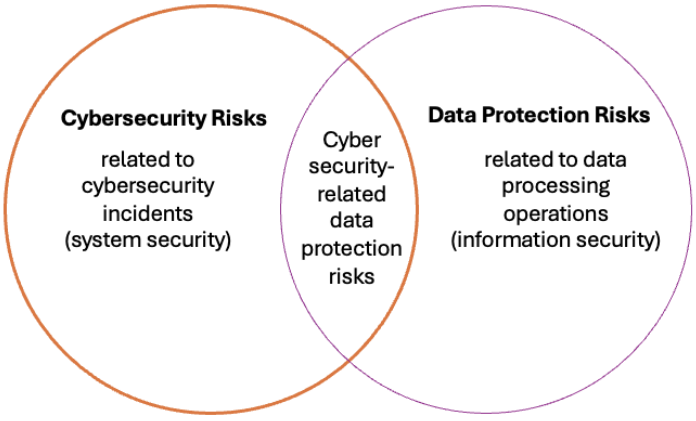

Figure 0-1. Relationship between cybersecurity and data protection [2]

Chapter 1: Ethics of cybersecurity

The ethics of cybersecurity encompass a broad spectrum of considerations, from the application of professional codes of ethics to the resolution of specific ethical dilemmas encountered in practice [3], [4]. While professional codes exist, they often lack the granularity needed to guide cybersecurity professionals through the complex ethical landscapes they navigate [3]. In fact, research indicates that while there is a growing body of work on the ethics of cybersecurity, there is a disconnect between theoretical research and practical application, with a need for more detailed guidance to address specific ethical challenges in various cybersecurity domains [3]. Three distinct approaches are employed to address ethical concerns in cybersecurity. The initial approach, known as "bottom up," examines ethical issues that arise from various case studies and categorizes them into groups such as privacy-related and security-related concerns. The second approach, termed "pragmatist," focuses on ethical practices currently utilized in cybersecurity, with emphasis on the confidentiality, integrity, and availability triad. The third approach, referred to as "top down," begins with broader ethical theories and applies them to the field of cybersecurity [5]. Each of these three approaches - bottom-up, pragmatist, and top-down–has demonstrated advantages and disadvantages. These are discussed in detail in Section 1.2.

Focusing exclusively on any single approach risks adopting a limited viewpoint that overlooks important considerations. By integrating these approaches into a unified framework, we can achieve a more comprehensive and nuanced understanding of the issues at hand, while also providing practitioners with a clearer path forward.

As emerged in the different approaches to cybersecurity, the integration of ethical considerations into cybersecurity is not only a theoretical exercise but also a practical necessity for the responsible advancement of the field [3]. This integration is crucial not only for the profession to develop cybersecurity practices that adhere to ethical standards but also for society to ensure that these practices positively influence society without compromising human dignity [6].

However, despite the recognition of the importance of ethics by society and the technology industry, many corporate leaders and cybersecurity professionals consider ethical questions to be often conflated with legal compliance or dismissed as relative, indicating a gap in ethical understanding within the field [12]. To address this issue, a comprehensive guide is required to enhance our understanding of ethics.

In summary, the ethics of cybersecurity is a multifaceted issue that requires further research and the development of detailed guidance to support ethical decision-making in practice, while also considering the different human factors that influence ethical priorities [6], [7]. A first step toward addressing these challenges is to develop a new ethical framework that can evaluate existing guidelines and encompass all ethical issues arising in various cybersecurity contexts. The ethical framework for cybersecurity is indeed a complex and widely debated topic, involving multiple principles and considerations, with various scholars proposing different frameworks that emphasize distinct ethical values.

These frameworks aim to address the ethical challenges and dilemmas faced in the cybersecurity domain by providing guidelines that professionals can apply across different contexts. For example, Formosa et al. (2021) introduce a principlist ethical framework specifically tailored for cybersecurity, drawing from the AI4People framework and encompassing principles such as beneficence, non-maleficence, autonomy, justice, and explicability. This framework is applied to various cybersecurity contexts (penetration testing, ransomware, denial of service and system administration) to demonstrate its utility in cultivating ethical expertise among professionals [4].

Fenech et al. (2024), meanwhile, investigate how individuals prioritize different ethical principles in cybersecurity decision-making, finding autonomy and justice to be of particular importance, and suggests that individuals feel they have agency in their decisions [6].

Interestingly, while Fenech et al. (2024) find autonomy to be the most valued ethical principle in cybersecurity scenarios, Sindiramutty et al. (2024) and Boopathi and Khang (2023) discuss the balance security and privacy, particularly in the context of drones and healthcare, respectively [8], [9]. These two studies explore how this balance manifests differently depending on the context.

In Sindiramutty et al. (2024), the balance between security and privacy in the context of drone technology involves addressing the tension between protecting drone systems from cyber threats and mitigating the privacy risks associated with surveillance. This requires securing the systems against cyber threats while adopting measures that limit surveillance risks and uphold individuals' privacy rights. This approach ensures that drones can perform their intended functions without compromising the privacy of those who may inadvertently be affected by their data collection capabilities. In Boopathi and Khang (2023), balancing security and privacy in healthcare involves not only protecting sensitive medical data from cyber threats but also addressing the ethical use of AI. This means ensuring that AI-driven healthcare practices are fair, unbiased, and transparent, with respect for patient autonomy and informed consent. Achieving this balance requires robust security

measures to protect data while also fostering trust through ethical data practices and the responsible use of AI in medical decision-making.

Hani et al. (2024) add to the ethical discourse by addressing privacy and bias in personality profiling within cybersecurity, advocating for informed consent and transparency [10].

Lottu et al. (2024) emphasize the need for a robust ethical framework for AI technologies within IT systems, suggesting that ethical considerations are integral throughout the AI development lifecycle [11]. This framework includes transparency, accountability, fairness, privacy, and security, which are essential for fostering trust and mitigating biases [11]. Special attention to the security principle highlighting that robust security protocols are essential for defending against malicious exploitation and adversarial attacks.

Tseng (2015) critiques existing ethical frameworks for their practical limitations and proposes a new framework integrating personal and professional values [12].

So, what emerges from these frameworks is a multiformity that arises from different approaches in development: in Formosa et al. (2021), an approach based on a preestablished system is then implemented in practical situations, that is, in different contexts; in Fenech et al. (2024) we find an approach that emerges from individual considerations while in Sindiramutty et al. (2024), Boopathi and Khang (2023), Hani et al. (2024), and Lottu et al. (2024) we find several principles that may emerge and be deemed more relevant based on cybersecurity contexts. In Tseng (2015), the impracticality of existing frameworks is criticized, giving an approach to the framework more related to the professional or community sphere.

In conclusion, the idea is that the variety of ethical frameworks presented in the studies reveals that no single approach is sufficient to address all the ethical challenges in cybersecurity. At the same time, all frameworks, regardless of the approach used, carry valuable ethical principles that can be grouped and categorized for a common framework. Formosa et al. (2021) seem to be the closest to this type of vision with an approach that, however, does not consider aspects and ethical issues that may arise in practicality, but that can be indirectly redirected to the suggested framework.

In conclusion, an effective ethical framework for cybersecurity must be adaptable encompassing a range of principles and considerations that reflect the dynamic nature of the field [4], [6], [8], [11], [12], [13]. This "adaptable, hybrid approach" is recommended because it would combine the benefits of different frameworks to create a more comprehensive and effective ethical guide. This ensures that ethical guidelines remain relevant, practical, and capable of guiding professionals through the diverse ethical dilemmas they may face.

This effective adaptable framework will ensure that the development, implementation, and maintenance of cybersecurity measures are aligned with societal values and individual rights [6], [8], [9], [10], [11].

1.1 Cybersecurity, cyber risks, and data protection

Building on the ethical frameworks previously discussed, it is important to recognize how these principles translate into the practical realm of cybersecurity. Cybersecurity itself involves a range of activities designed to protect information systems and networks from unauthorized access, damage, or other cyber threats. This protection goes beyond merely securing data; it also includes ensuring that systems can withstand various forms of cyber-attacks, thus addressing the ethical obligation to prevent harm and maintain trust. Several common cyber risks highlight the complexities and ethical challenges inherent in this field, as they often stem from failures in technical, organizational, or human safeguards, which in turn expose systems to vulnerabilities that can be exploited by attackers (Table 1). For instance, **brute force attacks** aim to gain access to a system by systematically guessing passwords. These attacks involve trying every possible password combination until the correct one is found, posing a significant risk to any system where passwords are the primary means of access control.

Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks aim to disrupt the normal functioning of a service by overwhelming it with excessive requests. In the case of a DDoS attack, this disruption is caused by a coordinated assault from multiple sources, often using a botnet of compromised devices. These attacks can cripple essential services, resulting in significant downtime and potentially severe financial losses.

Another significant threat is the **Man-in-the-Middle (MITM) attack**, in which an attacker secretly intercepts and possibly alters communication between two parties, usually without detection. This type of attack compromises the integrity and confidentiality of data transmitted over a network, potentially leading to the theft of sensitive information.

Malware, short for malicious software, encompasses any software designed to harm a computer system, steal data, or spy on the user. It spreads through various vectors, such as malicious email attachments, infected websites, or compromised devices. Once a system is infected, malware can have far-reaching effects, potentially compromising not just the targeted device but also others connected to the same network.

Social engineering attacks exploit human psychology rather than technical vulnerabilities, manipulating individuals into divulging confidential information or taking actions that compromise

security. Attackers may, for example, pose as legitimate contacts or use social manipulation tactics to extract passwords, banking details, or other private data.

A particularly dangerous type of attack is the **0-day exploit**, which takes advantage of a software vulnerability that is unknown to the software's developers. Because the vulnerability is not yet patched, attackers can exploit it to cause damage or steal data before a fix is available.

Attacks on **web applications** pose another major cybersecurity risk. Vulnerabilities in web services can be exploited to spread malware, steal or manipulate data, or even take control of corporate servers. These attacks can have serious repercussions, including financial loss, damage to reputation, and loss of trust from customers and partners.

In addition to these conventional risks, newer types of attacks are emerging in the realm of **machine learning and artificial intelligence**. **Poisoning attacks** [14] target the training process of machine learning models, either by corrupting the data used to train the model (**data poisoning** [14], [15]) or by manipulating the model's parameters (**model poisoning** [16]). This can lead to degraded performance or malicious behavior by the model.

Evasion attacks are another type of adversarial attack, where attackers subtly alter inputs to a machine learning system to cause incorrect predictions or classifications [17], [18], [19]. These inputs, known as adversarial examples, are crafted to look normal to humans but are designed to confuse the model.

Finally, **privacy attacks** in machine learning exploit vulnerabilities in models to infer information about the training data [20]. Examples include **membership inference attacks**, where attackers determine whether a particular data point was part of the model's training set, and **data reconstruction attacks**, which aim to recover sensitive information from the model's outputs.

The overarching theme is that cybersecurity involves more than just implementing technical defenses; it requires a nuanced understanding of the ethical implications of protecting data and systems. As the paragraph hints, the field is evolving to address not only traditional threats but also sophisticated new forms of attacks, highlighting the need for a proactive approach to both technological and ethical challenges in cybersecurity.

1.2 Ethical frameworks for cybersecurity

After introducing the various types of cyber-attacks, such as malware, man-in-the-middle, denial of service, brute force, zero-day exploits, social engineering, and web application attacks, it becomes evident that ethical considerations are central in the field of cybersecurity. These attacks target

systems and data directly connected to individuals, raising serious moral concerns about privacy, harm prevention, and the responsible use of technology. Therefore, the integration of ethical principles into cybersecurity practices, both in defensive measures and in the design of ethical attack simulations, is not just an option but a necessity. To achieve this integration in a systematic and meaningful way, it is important to review and evaluate existing ethical frameworks for cybersecurity, with the aim of identifying their strengths and limitations, and eventually working toward a unified, adaptable framework that can better guide ethical decision-making in this domain. The current landscape of cybersecurity ethics is fragmented, with different approaches emerging from distinct moral thories, research communities, and practical applications. There are three main approaches to cybersecurity ethics: the top-down, bottom-up, and pragmatic methods **[5]** (Figure 0-1).

Each has its own advantages and limitations, but none of them fully address the complexities of the evolving cybersecurity landscape when considered in isolation. The bottom-up approach, derived from case studies, lacks generalizability and often fails to account for new types of attacks or emerging security practices. For instance, the **Hernandez-Jaimes et al. (2023) framework** [21], which emphasizes practical measures to ensure confidentiality, integrity, and availability in healthcare systems, may not fully address new attack vectors like AI-generated malware. Its focus on specific threats such as DoS attacks or ransomware in healthcare environments limits its ability to adapt to rapidly evolving security challenges beyond its case-based examples.

The pragmatic approach, which focuses on technical principles like confidentiality, integrity, and availability, tends to be overly narrow, emphasizing technical solutions at the expense of broader ethical values. The **ENISA Framework [22]**, for example, prioritizes these technical principles to protect digital systems and prevent unauthorized access or data corruption. However, it may not thoroughly engage with ethical principles like fairness or justice, which are important when considering the societal impact of cybersecurity measures, such as how security protocols may disproportionately affect certain user groups. Similarly, the **ISO/IEC PDTR 13335-1 framework** addresses the importance of confidentiality and integrity through case studies, but often overlooks ethical considerations such as user autonomy or transparency.

Meanwhile, the top-down approach, based on philosophical theories such as deontology, utilitarianism, and the ethics of risk, offers a more abstract set of guidelines but may not always provide actionable measures in specific situations. For example, the **Formosa et al. (2021) framework** [23] draws on deontological principles like autonomy and non-maleficence, applying

these to cybersecurity practices such as penetration testing to ensure ethical standards are met. However, while it provides a strong ethical foundation, it may lack detailed guidance on implementing these principles in specific scenarios, such as deciding how much information about security measures should be disclosed to users to maintain transparency without compromising security. Similarly, the **Weber and Klein (2020) framework [24]** for healthcare and ICT focuses on ethical principles like beneficence and confidentiality, which are rooted in utilitarian considerations of patient well-being. Yet, translating these principles into specific, actionable cybersecurity policies can be challenging in rapidly changing healthcare environments.

Given these limitations, a systematic review of existing frameworks is essential. A systematic review allows for a critical evaluation of various ethical frameworks by categorizing their methodologies and assessing their relevance across different cybersecurity domains. Such a review can help identify patterns, strengths, and gaps within these frameworks, laying the foundation for a more comprehensive and adaptable hybrid model that incorporates elements from all three approaches. The top-down perspective can provide a normative basis grounded in ethical theory, offering overarching principles such as "do no harm" or "respect for privacy." The bottom-up approach can inform this normative framework by incorporating insights from practical case studies, ensuring that the framework remains relevant and adaptable to real-world situations. Finally, the pragmatic approach can add technical rigor, ensuring that the ethical guidelines are implementable in specific cybersecurity contexts.

A systematic review aiming to combine these approaches would not only categorize existing ethical frameworks but also seek to establish guiding principles that are both theoretically sound and practically relevant. The result could be a hybrid framework that adapts to new cybersecurity challenges while maintaining a commitment to core ethical values. This adaptable framework would be particularly valuable in high-stakes fields like healthcare, where the consequences of cyberattacks extend beyond financial losses to potentially life-threatening scenarios.

In conclusion, a systematic review of existing ethical frameworks in cybersecurity is critical for moving toward a unified, hybrid model that can be applied across different sectors, including healthcare. By systematically analyzing and integrating various approaches, it is possible to develop a comprehensive framework that guides ethical decision-making in cybersecurity in a way that is both principled and adaptable to changing threats. This process not only enhances the ethical standards of the field but also ensures a more holistic approach to cybersecurity that prioritizes the protection of individuals and society as a whole.

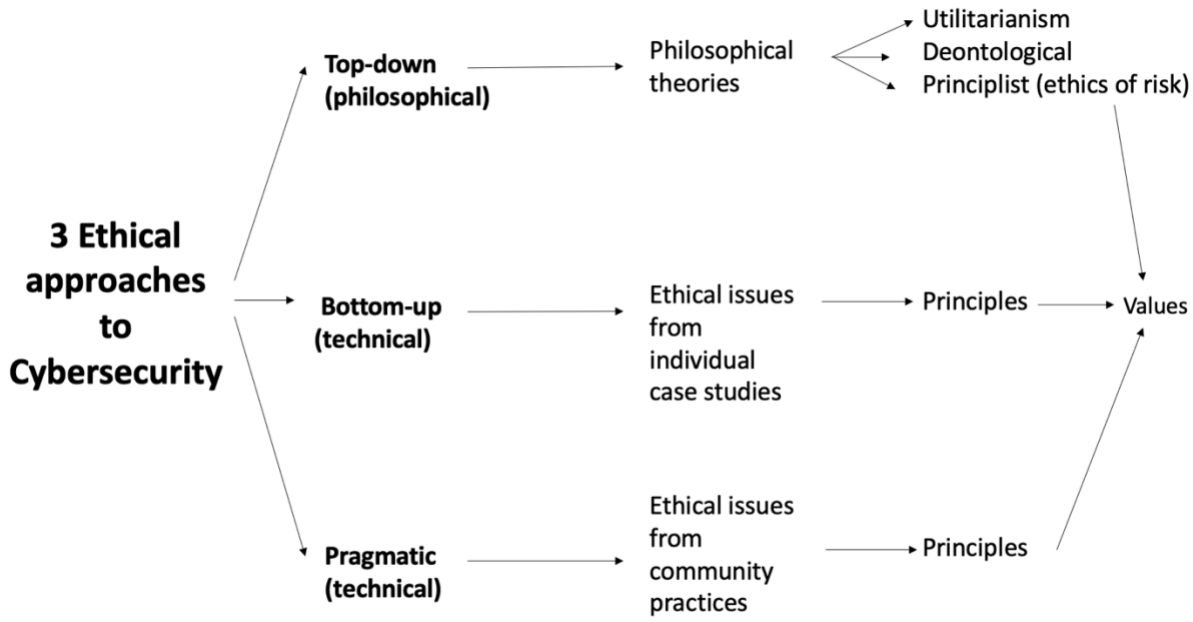

Figure 0-1. Ethical approaches to cybersecurity.

Related Work

In the field of cybersecurity, value conflicts have been identified through a comprehensive literature review that utilized a quantitative analysis of a substantial body of research and qualitative case studies [25]. Although in its initial stages, this study sheds light on the ethical challenges faced by cybersecurity in the business, health, and national sectors. The author of this study has highlighted ten key values: privacy, discrimination and prevention, fairness, equality, social justice, personal freedom, physical harm and prevention, and information harm and prevention. Although these values often conflict with one another, the concept of contextual integrity may serve as a helpful tool for understanding and mitigating such conflicts.

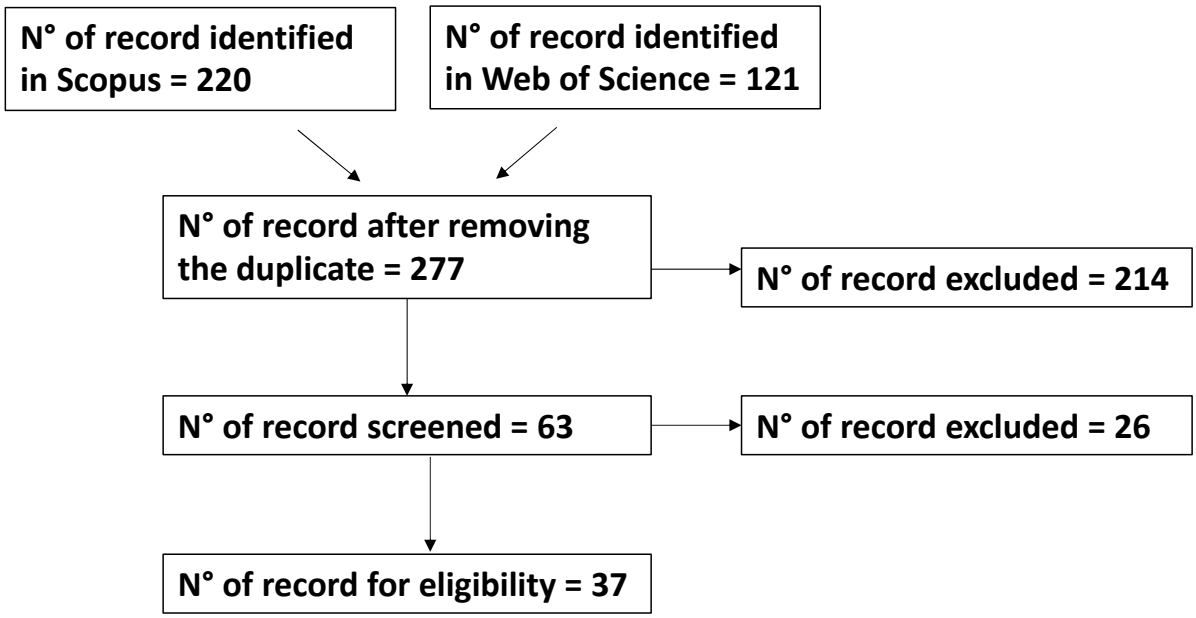

Figure 0-2. Systematic review procedure flowchart

1.2.1 Analysis and framework selection criteria

Eligibility criteria

The studies included for full-text eligibility were written in English. Studies dealing with non-ethical issues were excluded from full-text eligibility.

Search strategy

The studies were searched in the Scopus and Web of Science databases with the date of publication from January 2000 (2000-01-01) to November 2023 (2023-11-30); Therefore, studies published after this date were not included in the analysis. This represents a limitation that can be overcome in future research. The string chosen among all the possible ones is reported in Table 0-1for the N° of string = 5.

Exclusion criteria (2)

Papers are discarded based on the following criteria:

- They are not open access.
- The focus is solely on AI ethics and not on cybersecurity.

- The focus is on ethics in student education.
- The focus is on ethics in big data (not specific on cybersecurity).
- The focus is on cyberwarfare in general.

Table 0-1. Strings

Database	N° of String	String	Results	Rationale
Scopus	1	("cybersecurity" OR "information security" OR "computer security") AND ("ethics" OR "ethical considerations" OR "ethical frameworks" OR “ethical issues”) AND ("framework" OR "guidelines" OR "models")		• Generic terms
Web of science			121	
Web of science	2	("ethic*") AND ("issue*” OR “principle*” OR “value*”) AND ("cybersecurity") AND ("framework" OR "guideline*" OR "model*")	64	• Principle, value, issue • No "ethical considerations" OR "ethical frameworks" OR “ethical issues”
Scopus			101	
Scopus	3	("ethic*") AND ("cybersecurity") AND ("framework" OR "guideline*" OR "model*")	214	• No Principle, value, issue
Web of science				
Scopus	4	("ethic*") AND ("cybersecurity") AND ("framework" OR "guideline*" OR "model*" OR "methodolog**" OR "principl*" OR "valu*" OR "issu*")	387	• Adding All together "framework" OR "guideline*" OR "model*" OR "methodolog**" OR "principl*" OR "valu*" OR "issu*" • Adding "methodolog**"
Web of science				
Web of science	5	("cybersecurity") AND ("ethic*") AND ("issue*” OR “principle*” OR “value*” OR “threat” OR “opportunit*” OR “risk*”)	180	• NO "framework" OR "guideline*" OR "model*" OR "methodolog**" • Adding “threat” OR “opportunit*” OR “risk*”
Scopus			384	
Scopus	6	((cybersecurity AND ethic*) AND (framework OR principl* OR value*))	219	
Web of science			121	

Screening

The screening process for the studies was carried out based on their titles and abstracts, and the eligibility criteria were considered. This process is illustrated in Figure 0-2.

Data extraction

The information extracted from the studies included the date of publication (year), document type, source database, authors, title, origin of the principles, and ethical principles.

1.2.2 Summary of ethical frameworks

The date of publication (year), document type, source database, authors, title, origin of the principles, and ethical principles are listed in Table A-2 for each selected paper. In the next section, we provide the most relevant ethical frameworks extracted from the systematic literature review, considering the different approaches to cybersecurity ethics (Table A-3, Table A-4, Table A-5, Table A-6, Table A-7, Table A-8,Table A-9. Lorenzini et al. (2022)Table A-9 Table A-10, Table A-11, Table A-12, Table A-13, Table A-14).

Top-down (philosophical) framework

Table A-3, based on Kozhuharova et al. (2022), outlines several critical ethical and regulatory issues in the context of emerging technologies, particularly focusing on the Internet of Things (IoT), cloud computing, and new product testing.

1. Privacy and Consent

This principle highlights the evolving challenges in defining and securing privacy and consent in the digital age. The definition of privacy, though not explicitly provided, revolves around an individual's control over their personal information and the data processed by IoT devices. The principle of consent, which should be informed, specific, and voluntary, is not well-defined either, raising concerns. The table provides examples such as smartwatches and smart household appliances, which collect vast amounts of personal data, often with users unaware of how much data is collected or how it is used. The General Data Protection Regulation (GDPR) is mentioned as a mitigation strategy, ensuring compliance with consent and privacy standards. However, the

vagueness in defining these terms suggests the need for clearer guidelines and enhanced consumer awareness to protect individuals from unintentional breaches of privacy.

2. Security and Safety

Security, like privacy, is crucial in the IoT domain, but it is also "not well defined" in the context of the table. It refers to protecting IoT devices from various risks such as hacking, unauthorized access, and cyber intrusions. The case studies of hacked baby monitors and children's toys acting as surveillance devices illustrate the alarming security vulnerabilities in everyday technologies. The table points to collaboration between regulators and businesses as a solution, emphasizing the need for proactive rather than reactive measures. This issue is especially pertinent given the rapid proliferation of IoT devices, where security flaws can expose sensitive personal data or even pose physical dangers, as seen with compromised implantable cardiac devices.

3. Ownership of Data

Data ownership is another poorly defined area, according to the table, which outlines the difficulties in determining who controls and has rights over the data generated by users. With the rise of cloud computing platforms, questions of data control have become increasingly significant. The table mentions the need for regulation and informed decision-making, suggesting that users often do not fully understand the implications of storing data on cloud platforms. This issue is especially relevant as businesses and individuals alike rely on third-party platforms to store vast amounts of sensitive information, often without clear terms regarding ownership or usage.

4. Security in Cloud Computing

While security is not explicitly defined here either, the table refers to the protection of cloud-stored information against unauthorized access and data breaches. Cloud environments, which host enormous volumes of data, face constant cyber threats. The mitigation strategies mentioned—such as security checks, encryption, and cybersecurity expertise—are essential for maintaining the integrity of the data. However, there is an implicit understanding that these measures must be continuously updated to keep pace with evolving threats, underlining the dynamic and ever-present nature of security challenges in the digital space.

5. Testing of Future Products

This section addresses the potential risks associated with the testing of new technologies, where unforeseen dangers can arise. The example of Bluetooth earphones and their potential impact on the brain illustrates that these risks are not always immediately apparent, highlighting the ethical responsibility to anticipate possible harm. Moreover, the testing of autonomous vehicles, such as Uber's self-driving cars in urban areas, raises issues around consent. While test participants may agree to the trials, the bystanders (e.g., pedestrians in testing areas) have not consented, creating an ethical dilemma. The mitigation strategies, which include adopting Privacy Enhancing Technologies (PETs) and carrying out impact assessments, are necessary to ensure that these technologies are tested responsibly, with minimal risk to both direct and indirect participants.

6. Participant Consent

The concept of participant consent, particularly in new technology testing, is another area of concern. The table points out that consent is not always clearly given or obtained, especially when non-participants are indirectly involved in trials. In the example of self-driving cars, pedestrians in test environments did not explicitly consent to be part of the experiment. This scenario raises ethical questions about what constitutes informed and voluntary consent, and how it can be ensured when the lines between direct and indirect participation blur.

Table A-3 provides a comprehensive yet critical overview of the principles and issues surrounding privacy, security, and data ownership in emerging technologies. It highlights the challenges faced by regulators, businesses, and users in adapting to a rapidly changing technological landscape where the boundaries of privacy, security, and consent are continuously tested. The table emphasizes that more precise definitions and clearer guidelines are needed to mitigate the risks associated with IoT devices, cloud computing, and new technology testing. Additionally, the importance of cooperation between regulators and industry, alongside continuous updates to technical and organizational measures, is crucial for safeguarding the rights and safety of individuals in this evolving digital age.

Table A-4 from Formosa et al. (2021) presents a framework for evaluating cybersecurity technologies through the lens of ethical principles. Each principle offers a distinct dimension of ethical responsibility, helping guide the development and use of cybersecurity tools in ways that align with human values and societal needs. Below is a commentary on each of these principles,

illustrated by case studies such as penetration testing, denial of service (DoS) attacks, ransomware, and system administration.

1. Autonomy

Autonomy refers to the respect for individuals' ability to make informed decisions about how technology impacts their lives. In the realm of cybersecurity, this means that individuals should have control over how security systems interact with their data and online activities. For instance, in penetration testing and ransomware attacks, respecting autonomy would involve informing users about the risks and securing their consent to protect their information. By contrast, a failure to respect autonomy occurs when security decisions are made without user input or awareness, such as in situations where cybersecurity systems automatically handle personal data without providing transparency. The case study of ransomware is particularly relevant here, as it demonstrates how breaches of autonomy (e.g., encrypting personal files without permission) can lead to severe consequences for individuals and organizations.

2. Explicability

Explicability addresses the need for transparency in how cybersecurity technologies operate. This principle ensures that the processes behind security systems are intelligible, and it’s clear who is responsible for them. For cybersecurity technologies to meet the principle of explicability, they must be designed in ways that allow users to understand how they work and how they might affect their digital lives. In penetration testing or cyber defense strategies, it’s crucial that those subjected to security measures (e.g., employees in a company undergoing testing) understand what’s being done and why. Explicability also ensures accountability, particularly in cases of failure, such as DoS attacks or ransomware breaches. Transparency about how these threats are being managed is essential for users' trust and confidence in cybersecurity practices.

3. Non-Maleficence

The principle of non-maleficence insists that cybersecurity technologies should not cause harm. In practice, this means ensuring that technologies are not used maliciously to exploit vulnerabilities, create disruptions, or cause any undue harm to individuals or societies. The examples in the table (DoS attacks, ransomware, etc.) represent scenarios where this principle is often violated, particularly by malicious actors seeking to exploit or harm others. However, even in the context of

penetration testing or cybersecurity system administration, the principle still applies—security professionals must avoid causing unnecessary disruptions or accessing sensitive information that could harm users. Proper ethical guidelines in non-maleficence help mitigate risks and ensure that cybersecurity efforts do not inadvertently make individuals' digital environments more hostile.

4. Justice

Justice in cybersecurity relates to fairness, equality, and impartiality in the application of technology. Technologies should not disproportionately affect certain groups, nor should they undermine social solidarity. In cybersecurity, the principle of justice ensures that protections are applied equally across different groups of people and that security technologies do not create digital divides or unfair barriers to access. For instance, ransomware attacks often disproportionately affect vulnerable organizations, such as hospitals or smaller businesses, highlighting the importance of justice in ensuring that these entities have equal access to protective measures. Additionally, cybersecurity frameworks need to ensure that defensive tools like firewalls or encryption technologies are not reserved only for wealthy individuals or corporations but are accessible to all, promoting digital equity.

5. Beneficence

Finally, the principle of beneficence asserts that cybersecurity technologies should contribute positively to human well-being. This principle underscores the role of cybersecurity in promoting overall societal benefits, ensuring that systems are in place to enhance rather than detract from the quality of life. In terms of penetration testing or defending against ransomware, beneficence is reflected in how these technologies safeguard not just individual users but also the broader community. By preventing attacks and securing information systems, cybersecurity professionals contribute to a safer, more resilient digital environment. Furthermore, beneficence requires ongoing efforts to improve cybersecurity, ensuring that new technologies do not just solve immediate problems but also anticipate future threats and continue to promote human flourishing.

The ethical principles of autonomy, explicability, non-maleficence, justice, and beneficence provide a robust framework for evaluating the role of cybersecurity technologies in modern society. Through the case studies mentioned—penetration testing, ransomware, DoS attacks, and cybersecurity system administration—these principles highlight the necessity of designing and using technologies

that are not only effective in defending against threats but also aligned with broader ethical considerations. In practice, balancing these principles can be challenging, as the complexity of cyber threats often forces difficult decisions. However, by adhering to these guidelines, cybersecurity professionals can help ensure that their actions promote trust, security, and fairness in the digital world.

Table 6 from Weber and Klein (2020) outlines various ethical principles in the context of healthcare and information technology (ICT) systems, particularly as they relate to issues like medical devices, privacy, and patient safety. This detailed framework examines how these principles are applied, as well as the challenges and mitigation strategies in real-world scenarios, such as the use of cardiac pacemakers and electronic health cards (eHC) in healthcare systems like Germany's. Below is a more in-depth discussion of the key ethical principles addressed in Table A-5.

1. Autonomy

The principle of autonomy emphasizes respecting individuals' rights to make their own informed decisions, particularly in the context of medical treatments. This principle is fundamental to patient-centered care and is exemplified by cases involving cardiac pacemakers and other implantable medical devices. For instance, a patient must be given the freedom to choose whether to undergo a life-altering medical procedure, such as receiving an implantable device. This also extends to technologies like the eHC, which could significantly influence patient access to healthcare services. The implementation of these technologies must respect individuals' right to control their healthcare information and decisions.

2. Non-Maleficence

This principle refers to the ethical mandate to avoid causing harm, which is especially crucial in healthcare. Non-maleficence ensures that new medical technologies, procedures, and treatments are not harmful to patients. When introducing new systems such as ICT in healthcare, special attention must be given to minimizing potential risks, whether technical (such as data breaches) or medical (such as malfunctioning devices). Ensuring non-maleficence involves careful planning and rigorous testing of new treatments to avoid unintended negative outcomes for patients.

3. Justice

The concept of justice in healthcare involves ensuring fair treatment for all patients, preventing discrimination, and making efficient use of scarce resources. In healthcare systems, justice calls for equal access to services, particularly as ICT systems are introduced. Justice is aligned with the idea that no patient should be unfairly disadvantaged by the adoption of new healthcare technologies. This principle also ties into the efficient allocation of healthcare resources—whether human, technological, or financial—to ensure that everyone receives equitable care. Shared responsibility among all stakeholders (patients, healthcare providers, and policymakers) is crucial for upholding this principle.

4. Beneficence

Beneficence in healthcare means actively promoting the well-being of others, such as improving patient health outcomes and contributing to public welfare. This principle underlies the adoption of many new healthcare technologies that aim to enhance quality of care, increase patient satisfaction, and improve overall health outcomes. For example, the use of advanced ICT systems, which streamline administrative processes and make healthcare more efficient, ultimately benefits patients by reducing waiting times and facilitating quicker access to medical information.

5. Privacy of Information and Confidentiality

The ethical principles of privacy and confidentiality are critical in healthcare, particularly in the digital age where sensitive health data is often stored and shared electronically. Safeguarding patient information is vital to maintaining trust in the healthcare system. The rise of big data and its application in healthcare necessitates robust privacy protections to prevent unauthorized access to patient records, which could result in harm to individuals if misused.

6. Efficiency and Quality of Services

Efficiency and quality are intertwined with justice and beneficence. In healthcare, efficiency refers to optimizing resources to achieve the best possible outcomes, while quality refers to the standard of care provided. ICT systems, when effectively implemented, can greatly enhance both. For example, by reducing administrative burdens, such systems can allow healthcare professionals to focus more on patient care, thus improving the overall quality of services.

7. Usability of Services

Usability is a key technical principle that ensures healthcare technologies are accessible and user-friendly for all stakeholders, including patients, medical staff, and administrators. This aligns with the principles of non-maleficence, justice, quality, and efficiency. If a system is difficult to use, it can create barriers to care, harm patients through errors, or introduce inefficiencies that undermine the healthcare system's overall effectiveness.

8. Safety

Safety is another critical technical consideration, closely related to non-maleficence and beneficence. The implementation of healthcare technologies should reduce health risks, ensuring that they do not pose additional dangers to patients. For example, ensuring that medical devices like pacemakers function correctly is paramount to patient safety. Additionally, ICT systems must be robust enough to protect against cyber threats, which could compromise patient safety if health records are manipulated or lost.

9. Fairness and Equality

Equitable access to healthcare services is essential for ensuring fairness and equality. This means that everyone, regardless of their background or socioeconomic status, should have access to the same quality of care. In the context of ICT systems, it is important that such systems do not create new inequalities by favoring certain groups over others, whether through accessibility barriers or discriminatory practices.

10. Privacy and Trust

Trust in the healthcare system is foundational to its success. Patients need to trust that their personal information will be protected and that their health data will not be misused. Privacy protections must be robust enough to prevent breaches, as any loss of personal data can erode trust in the system. The challenge lies in balancing the use of big data to improve healthcare outcomes with the need to protect individual privacy rights.

11. Freedom and Consent

Freedom and consent are integral to the principle of autonomy. Patients should have the freedom to choose whether to adopt new healthcare technologies and should be fully informed about the

implications of their decisions. Informed consent is essential for ensuring that patients can make autonomous decisions about their care, particularly when it involves complex new technologies.

12. Dignity and Solidarity

Lastly, the principles of dignity and solidarity emphasize the inherent worth of every individual and the universal right to healthcare. Beneficence plays a key role here, as healthcare professionals must work to uphold the dignity of their patients. This includes treating patients with respect and ensuring that new technologies do not dehumanize the care process but instead enhance the respect and dignity afforded to each person.

Overall, the table from Weber and Klein (2020) provides a comprehensive ethical framework for evaluating the impact of new healthcare technologies on patient care. These principles—ranging from autonomy and non-maleficence to privacy and trust—serve as guidelines to ensure that technological advancements benefit patients while minimizing harm and promoting fairness and dignity.

Table A-6 from Domingo-Ferrer (2019) outlines four key principles: Autonomy, Security, Privacy, and Fairness. Each principle is defined and briefly explained. Here's a commentary on each principle, analyzing their interrelationships and implications, with particular reference to the definitions provided.

1. Autonomy

Autonomy is defined as the ability to make uncoerced decisions, representing self-determination and freedom from external control. It is a foundational concept in ethics, often linked with individual rights and personal freedom. In the context of information security and privacy, autonomy underscores the importance of allowing individuals to control their own data and make informed decisions about its use. The emphasis on autonomy aligns with data protection regulations such as the General Data Protection Regulation (GDPR), which grants individuals control over their personal data, including consent and access rights.

The table suggests that autonomy supports other principles such as privacy and security, creating a cascade effect. For instance, when individuals have the power to control their personal data, this autonomy directly bolsters privacy by preventing unauthorized access. Moreover, this respect for

privacy enhances overall security, as data protection mechanisms reduce the risk of breaches. The reciprocal relationship implies that strengthening one principle could reinforce others, creating a more resilient and ethical framework for data governance.

2. Security

Security is described as the state of being free from danger or threat. In the context of data management, it involves measures to protect data against unauthorized access, corruption, or theft. Security serves as the backbone for upholding autonomy, privacy, and fairness, as it provides the necessary infrastructure for these principles to be operationalized. Without robust security, individuals cannot exercise their autonomy, as their data could be compromised or misused without their consent.

Moreover, security's role in supporting privacy is evident. When security measures are strong, personal data remains protected from unauthorized observation or interference, thus upholding the principle of privacy. In turn, maintaining data privacy helps preserve fairness by preventing discriminatory practices that could arise from the misuse of sensitive information. For example, if personal data were compromised and exploited to make biased decisions, it would directly violate the principle of fairness.

3. Privacy

Privacy is articulated through various definitions, emphasizing the right to be left alone and the individual's control over personal information. Historically, privacy has been a central concern in discussions around data protection, reflecting not just a legal right but also an ethical obligation to respect personal boundaries. Privacy underpins autonomy by allowing individuals to control the flow of information about themselves, thereby enabling informed and voluntary decision-making.

The interconnectedness between privacy and the other principles is evident in the table. Privacy not only supports autonomy by enabling self-determination but also enhances security by setting boundaries on who can access personal data. The idea of privacy contributing to fairness arises from its role in protecting individuals from profiling, discrimination, or exploitation based on personal data. When privacy is compromised, fairness is at risk, as personal information could be used to perpetuate biases.

4. Fairness

Fairness is defined as the impartial treatment of individuals without favoritism or discrimination. It is essential for ethical data management practices, ensuring that all individuals are treated equitably. The principle of fairness is particularly relevant in the era of big data and algorithmic decision-making, where biases can inadvertently creep into systems, leading to unjust outcomes.

The table suggests that the other principles—autonomy, security, and privacy—contribute to fairness. For instance, when individuals have autonomy over their data, they can protect themselves from unfair treatment based on unauthorized data usage. Similarly, strong security measures can prevent data manipulation that might lead to biased decisions, and robust privacy protections can shield individuals from discriminatory profiling. Hence, fairness is not an isolated principle but one that is supported by and reinforces the other principles in a cyclical manner.

The interdependencies described in Table A-6 illustrate a holistic approach to ethical data management, where strengthening one principle naturally supports the others. The cyclical nature of these relationships indicates that failing to uphold one principle could undermine the others. For example, a breach in security could compromise privacy, limiting individuals' autonomy over their data and leading to unfair consequences.

Bottom-up (technical) framework.

Table A-7 from Hernandez-Jaimes ML et al. (2023) provides a comprehensive overview of various cyberattacks targeting Internet of Medical Things (IoMT) systems, highlighting the importance of maintaining data security and service availability in healthcare environments. Each attack type is analyzed according to its principle—confidentiality, integrity, and availability—and its mitigation strategies, while also referencing real-world case studies where applicable.

1. Denial of Service (DoS)

DoS attacks focus on compromising the availability of medical services by targeting the application, transport, and network layers of healthcare systems. Availability is a critical aspect here, as the continuous operation of medical devices and patient records is crucial for emergency responses. For instance, a healthcare system must ensure that devices like pacemakers or insulin pumps, which rely on timely data, are operational without interruptions. Confidentiality and integrity are also indirectly threatened, as system overloads could expose or corrupt patient data. Mitigation for DoS

attacks involves the implementation of new standards and laws that govern the handling of IoMT data. Furthermore, Intrusion Detection and Prevention Systems (IDPS) are recommended to monitor traffic and detect unusual patterns indicative of an attack. These measures help in maintaining system robustness and ensuring that healthcare services remain operational even under duress.

2. Distributed Denial of Service (DDoS)

DDoS attacks extend the principles of DoS attacks by amplifying the scale, making it harder to protect systems. Similar to DoS, the availability of healthcare services is the primary target, with attacks often overwhelming servers or networks, rendering services like patient records or telemedicine unavailable. This type of attack also threatens the integrity of healthcare services by creating delays that can impact patient care. Although the table does not specify case studies for DDoS attacks, these are prevalent in real-world scenarios, often targeting large healthcare institutions. Mitigations typically involve network-level defenses such as load balancers, firewalls, and rate-limiting strategies to reduce the attack surface.

3. Man-in-the-Middle (MitM)

MitM attacks exploit vulnerabilities in the network and transport layers, aiming to compromise both the confidentiality and integrity of IoMT data. In these scenarios, attackers intercept communications between medical devices and servers, potentially altering or stealing sensitive data, such as health records or device telemetry. The table emphasizes confidentiality and integrity as the key concerns in MitM attacks, but it does not list specific case studies. Nevertheless, the impact is significant, as any tampered data could lead to incorrect medical decisions, thereby endangering patients. Encryption protocols, secure communication channels, and network monitoring are critical in preventing these attacks.

4. Ransomware

Ransomware is one of the most prominent threats to the integrity and availability of healthcare data, particularly at the application layer. In this attack, malicious software encrypts patient data, locking it behind a paywall until a ransom is paid. This not only compromises the integrity of the system (since data is held hostage) but also the availability of services, as systems become unusable. The table lists several high-profile case studies, including the University of Vermont Medical Center (2020) and the infamous WannaCry attack on the National Health Service (NHS) in Britain (2017).

These events caused widespread disruption in healthcare services, delaying treatments and surgeries. Mitigation strategies include regular data backups, robust anti-malware software, and employee training to avoid phishing schemes that often serve as entry points for ransomware.

5. Information Gathering

Information gathering attacks, aimed at breaching confidentiality, occur in the network layer. Attackers may seek to collect sensitive patient information, such as health records, financial data, or even private conversations between patients and doctors. While these attacks may not directly disrupt services, the breach of confidentiality can have long-lasting effects on patient trust and lead to significant legal consequences for healthcare providers. Mitigating these attacks involves encryption, access controls, and regular security audits to ensure that vulnerabilities are addressed before they can be exploited.

6. Malware

Malware attacks, particularly at the application layer, focus on compromising both the integrity and availability of medical systems. Once malware is introduced, it can disrupt services by corrupting files, deleting data, or causing system crashes. This can prevent doctors from accessing critical patient information in real-time, jeopardizing patient care. Malware can be introduced through various means, such as infected USB drives or phishing emails, making awareness and proactive defenses, like anti-virus software and regular system updates, essential to mitigation.

7. Injection Attacks

Injection attacks, often seen in the application layer, affect confidentiality, integrity, and availability by inserting malicious code into vulnerable applications. For instance, an SQL injection could allow attackers to gain unauthorized access to patient databases, altering or stealing critical data. The confidentiality of the patient's information is directly compromised, and the availability of services may be disrupted if systems are taken offline for remediation. The mitigation for injection attacks typically involves secure coding practices, input validation, and regular security testing to identify vulnerabilities early in the development cycle.

8. Password Attacks

Password attacks threaten all three principles: confidentiality, integrity, and availability, particularly at the application layer. Weak or stolen passwords can grant unauthorized users access to medical

systems, allowing them to manipulate patient records, shut down services, or expose sensitive data. Given the interconnected nature of IoMT systems, a single compromised password could have cascading effects across an entire network. Strong password policies, two-factor authentication (2FA), and regular password rotations are standard methods to prevent such attacks.

Table A-7 provides a clear and structured overview of the different attack vectors targeting IoMT systems and their consequences on confidentiality, integrity, and availability. Through real-world case studies and mitigation strategies, it emphasizes the critical need for robust security measures in healthcare environments. Implementing a multi-layered defense, including encryption, intrusion detection systems, and continuous employee training, is essential for protecting sensitive medical data and ensuring the seamless operation of healthcare services.

Table A-8 from Yeng et Wright (2023) addresses phishing simulation, a common cybersecurity training and research method used to assess an individual's susceptibility to phishing attacks. In these simulations, participants receive simulated phishing emails, and their responses—whether they click on the malicious link or report the email—are monitored to gauge their vulnerability to real-world phishing attempts.
The table highlights several ethical issues associated with conducting phishing simulations:

1. Deception vs. Informed Consent

Phishing simulations often involve a degree of deception, where participants are not fully informed about the nature of the study to avoid biasing their responses. This creates an ethical dilemma, as research ethics guidelines typically require participants to be informed about the nature of the study.
Deception might be justified in this context to ensure that the results accurately reflect participants' natural behavior when faced with a phishing attempt. If participants are aware they are being tested, they may behave differently, compromising the validity of the study. However, balancing this with the need to obtain informed consent remains a challenge.

2. Potential Harm to Participants

There is a risk that participants may experience psychological distress, embarrassment, or a sense of betrayal if they discover they were deceived. For instance, those who fall for the phishing simulation may feel embarrassed or anxious about their susceptibility.

The table suggests mitigating this harm by avoiding the collection of personally identifiable information (PII), thereby reducing the impact on participants' privacy. Additionally, providing debriefings after the study, where participants are informed about the deception and the purpose of the research, can help address any potential negative feelings.

3. Infringement upon Privacy

Even in simulated environments, phishing simulations can infringe on participants' privacy if personal data is collected or exposed. Privacy concerns are heightened if data collection is done without fully informing participants of what is being gathered.

The use of encryption methods, adhering to data protection regulations such as GDPR, is recommended to protect any sensitive information collected during the study. This approach ensures that even if data is inadvertently disclosed, it remains secure.

4. Violation of Participants' Autonomy

The element of deception may also impact the autonomy of participants, as they are not fully aware of their involvement in a study on phishing. This raises questions about their ability to make an informed choice regarding participation.

The table does not explicitly suggest specific mitigation strategies for this issue but underscores the ethical tension between conducting realistic phishing simulations and respecting participants' autonomy.

5. Limited Control of Participants

Phishing simulations often do not allow participants to exercise full control over their engagement with the study, as they are not aware that their actions are being observed. This lack of control can raise ethical concerns about manipulating participants' behavior without their consent.
This issue ties back to the overall debate about the ethics of deception in research. While realistic simulations are essential for studying genuine reactions, the ethical cost of limiting participants' control must be considered.

6. Justification for Withholding Information

The decision to withhold certain information from participants must be justified by the research's value. If withholding information significantly enhances the study's validity, it may be ethically permissible, provided appropriate safeguards are in place.
Researchers should demonstrate that the benefits of gaining insights into phishing susceptibility outweigh the potential ethical risks associated with deception.

7. Phishing Susceptibility (Click/Not Click)

The primary metric in phishing simulations is whether participants click on the simulated phishing link. The study aims to understand the decision-making process behind clicking or not clicking, which provides insights into user behavior under potential cyber threats.
Conducting multiple phishing simulations or varying the types of phishing scenarios presented to participants may provide a more comprehensive understanding of factors influencing susceptibility.

Overall, Table A-8 presents a nuanced view of the ethical dilemmas inherent in phishing simulation research. While the use of deception is often necessary to maintain the validity of the study, careful consideration must be given to the ethical implications, particularly concerning informed consent, potential harm, and privacy. Researchers should employ mitigation strategies, such as anonymizing data and conducting thorough debriefings, to minimize ethical risks while still obtaining valuable insights into phishing behavior.

Table A-9 from Lorenzini et al. (2022), discusses various principles and issues associated with penetration testing in the context of cybersecurity. It highlights different aspects, such as privacy,

freedom, and the ethical considerations in penetration testing, and connects these principles to practical challenges and mitigation strategies.

1. Privacy

The principle of privacy in penetration testing emphasizes the importance of protecting sensitive data. Penetration testers (pen-testers) must ensure that any data encountered during testing is kept confidential and is not disseminated or leaked. This aligns with ethical standards in cybersecurity where safeguarding personal and organizational data is paramount. The challenge here lies in balancing the need for thorough testing with the imperative to protect sensitive information. Without clear guidelines, there could be an inconsistency in how privacy is maintained during these tests.

2. Freedom

The concept of freedom in this context refers to open access to information, which can empower individuals to educate themselves. However, when applying this principle to penetration testing, the challenge is ensuring that the freedom to access information does not compromise security. For example, while open-source tools are freely available and used for penetration testing, they can also be exploited by malicious actors. The balance here involves providing enough information and resources for pen-testers while safeguarding against misuse.

3. Penetration testing is not commonly practiced in healthcare facilities

This is due to several factors, including:

- Fears and Prejudices. The perception of penetration testing as akin to hacking creates reluctance in adopting it within healthcare settings. There's a misconception that penetration testing might disrupt operations or compromise patient data rather than securing it. This highlights the need for better education about the benefits and ethical practices of penetration testing.
- Lack of an Official Code of Ethics. The absence of a universally recognized code of ethics for penetration testers leads to uncertainty about what constitutes acceptable practices.

This can deter organizations from engaging in penetration testing due to the potential risks and lack of clearly defined standards.

- Limited Financial Resources. Healthcare facilities often operate on constrained budgets, limiting their capacity to invest in comprehensive cybersecurity measures like penetration testing. This underscores the need for cost-effective solutions and perhaps government incentives to improve cybersecurity in critical sectors like healthcare.

4. Mitigation Strategies

1. Hiring Penetration Testing Companies:
    - One suggested mitigation is to hire specialized companies that provide penetration testing services. This can help healthcare facilities access expert services without needing to maintain in-house capabilities. These companies would likely follow established best practices and ethical guidelines, even in the absence of a universal code of ethics.
2. Drafting an International Code of Ethics:
    - Establishing an international code of ethics for penetration testing is proposed as a solution to address the ambiguity around moral principles and best practices. Such a code would help standardize the expectations for pen-testers, ensuring that they adhere to defined ethical standards, thus reducing the hesitancy from organizations to engage in these services.

Table A-9 underscores the complex interplay between ethical principles, practical challenges, and the application of penetration testing in cybersecurity. It highlights the need for enhanced awareness and education, financial considerations, and standardization of ethical guidelines.
The table effectively connects these principles with practical solutions, advocating for a structured approach to integrating penetration testing into organizational cybersecurity strategies.

Pragmatic (technical) framework.

Table A-10 from ENISA outlines key cybersecurity principles and their definitions, providing insight into fundamental aspects of information security. Each principle serves as a cornerstone for protecting digital assets and ensuring trust in digital communication. Here's an analysis of these principles.

1. Confidentiality

Confidentiality ensures that communications or stored data are protected against unauthorized access. It aims to keep sensitive information out of the hands of individuals who do not have permission to access it. In cybersecurity, confidentiality is crucial for protecting personal and organizational information. It involves implementing measures such as encryption, access control, and secure communication protocols to prevent data breaches. Effective confidentiality practices reduce the risk of information exposure, which can lead to identity theft, financial losses, or even legal consequences for companies. Additionally, it plays a vital role in maintaining privacy, especially in sectors like healthcare and finance, where data sensitivity is paramount.

2. Integrity

Integrity is the assurance that data remains accurate and unaltered during transmission, storage, or processing. Data integrity is fundamental for ensuring the accuracy and reliability of information. It prevents unauthorized users from making alterations that could compromise data quality, lead to misinformation, or cause system malfunctions. Techniques to preserve integrity include hashing, and digital signatures, which help detect and prevent unauthorized modifications. In sectors such as e-commerce and government, where data authenticity is critical, breaches in integrity could undermine operations, resulting in significant operational and reputational damage.

3. Availability

Availability refers to the ability of data and systems to be accessible and operational when needed. Ensuring availability is crucial for organizations that depend on digital services and information for daily operations. Denial of service attacks (DoS) and natural disasters can threaten availability, leading to disruptions. Mitigation strategies like redundancy, failover systems, and regular maintenance are essential to minimize downtime and ensure continuous service delivery. For instance, availability is particularly vital for critical infrastructures such as healthcare systems, where downtime could have life-threatening consequences.

4. Authenticity

Authenticity ensures that an entity is what it claims to be. Authenticity is important in verifying the legitimacy of users, devices, and communications in digital environments. It helps prevent spoofing attacks, where malicious actors pose as trusted entities to gain unauthorized access. Techniques such as two-factor authentication (2FA), digital certificates, and biometrics enhance authenticity by providing multiple layers of verification. In online banking, for instance, verifying the authenticity of both the user and the service provider is critical for preventing fraud.

5. Non-repudiation

While the table does not provide a detailed definition for non-repudiation, it is generally understood as ensuring that a party in a communication cannot deny the authenticity of their signature on a document or the sending of a message.
Non-repudiation is vital for legal and financial transactions, where parties must be accountable for their actions. It involves using mechanisms like digital signatures and audit logs to provide proof of the origin and integrity of data. Non-repudiation measures are crucial for e-commerce, legal agreements, and any context where it is important to verify that communications and transactions have occurred as claimed.

The principles outlined in Table A-10 form a comprehensive foundation for understanding and implementing effective cybersecurity measures. Each principle addresses a specific aspect of data security and works in conjunction to ensure a holistic approach to protecting information.

Table A-11 from ISO/IEC PDTR 13335-1 outlines several key principles of information security, focusing on confidentiality and integrity. Below, we present a detailed analysis of these principles in relation to information security practices.

1. Confidentiality

Confidentiality refers to the practice of ensuring that sensitive information is accessible only to those authorized to view it. In an organizational context, this involves protecting data from

unauthorized access to prevent data breaches and safeguard personal or proprietary information. Implementing confidentiality measures might include encryption, access control mechanisms (like passwords and multi-factor authentication), and network security protocols to prevent interception by unauthorized parties.

Ensuring confidentiality is essential for maintaining trust between an organization and its stakeholders. For instance, in healthcare, protecting patient records is not only a legal requirement but also crucial for upholding the integrity of the medical profession. In cases where confidentiality is breached, the consequences can be severe, leading to legal penalties, reputational damage, and loss of business.

2. Integrity

Integrity ensures that data remains accurate, consistent, and unaltered unless modified by authorized users. It protects data from being tampered with or corrupted, either accidentally or maliciously. Ensuring data integrity is critical in environments where decisions are made based on data, such as financial institutions, healthcare, and government services.

Maintaining data integrity involves using mechanisms like cryptographic hash functions and audit trails that track changes made to data. These mechanisms help detect any unauthorized modifications and can trigger alerts for further investigation. For example, in a financial institution, maintaining data integrity ensures that transaction records are accurate and free from unauthorized alterations.

3. Case Study and Mitigation Measures

While the table mentions these principles broadly, a case study approach can be instrumental in providing a clearer picture of the principles' applications and challenges. For instance, a case study in a banking sector could illustrate how data confidentiality and integrity breaches have been managed and mitigated through robust encryption protocols and regular security audits. Moreover, it would be beneficial to explore mitigation strategies such as implementing least privilege access controls and continuous monitoring to ensure the effectiveness of the security measures.

Overall, the principles outlined in Table A-11 are foundational to information security. Implementing these principles in a comprehensive security framework helps protect sensitive information and maintain the trustworthiness of an organization's operations.

1.2.3 Ethical implications and future directions

The need for ethical review processes arises in cybersecurity research and practice owing to the inherent ethical risks present, such as privacy violations and system vulnerabilities [26]. Additionally, the necessity for a comprehensive systematic framework to identify, evaluate, and address ethical concerns in the field of cybersecurity arises from intrinsic value conflicts [27]. This is also more important in the context of digital healthcare, where the ethical considerations of cybersecurity are particularly salient as they intersect with biomedical ethics, care-based ethics, and technical aims. Moreover, given the sensitive nature of healthcare data and the potential consequences of a breach, it is crucial for healthcare organizations to prioritize cybersecurity and take proactive measures to protect patient information. Furthermore, the necessity for a systematic framework arises from the new ethical issues resulting from the rapid growth of digital technologies.

First, this study analyzes the current frameworks in the field of cybersecurity ethics. Second, this study developed an adaptive, comprehensive, and structured framework to identify, evaluate, and address ethical principles in cybersecurity practices. This framework guarantees adherence to the essential values and principles.

Top-down approaches, primarily inspired by bioethics, emphasize principles like autonomy, non-maleficence, justice, and beneficence, which are traditionally applied in medical ethics but increasingly relevant to cybersecurity. For example, the AI4PEOPLE framework, analyzed in Formosa et al., incorporates principles such as explicability alongside autonomy and justice, suggesting their critical role in complex scenarios like ransomware and denial-of-service (DoS) attacks. The inclusion of explicability highlights the need for transparency in cybersecurity practices, ensuring users understand how their data is protected and what risks they may face.

Furthermore, the top-down approach identifies overlaps and hidden interdependencies between principles. For instance, beneficence and non-maleficence often encompass elements of security and privacy, as noted by Weber et al. and Kuznezova et al. These overlaps demonstrate the multi-layered nature of ethical decision-making in cybersecurity, where protecting user data aligns with preventing harm and promoting well-being. The mapping between ethical and technical principles

by Weber et al. adds a valuable dimension, linking confidentiality to autonomy and fairness to justice, thus extending the ethical framework's applicability to technical fields.

Top-down frameworks also emphasize the significance of conducting a self-assessment ethics checklist to identify and evaluate ethical risks as a mitigation measure. While this checklist provides a structured approach for ethical evaluation, its qualitative nature introduces subjectivity into the assessment process. Addressing this limitation will require the development of a quantitative assessment model in the future, which would allow for more objective and consistent evaluations.

Conversely, bottom-up approaches derive ethical principles from practical case studies, focusing on specific cybersecurity challenges. For example, Hernandez-Jaimes et al. identify how principles like availability are affected in DoS attacks, while confidentiality and integrity are compromised in man-in-the-middle scenarios. These case-based frameworks effectively highlight the real-world implications of ethical principles in action. However, while bottom-up frameworks can identify specific risks and propose mitigation strategies, they may lack the overarching coherence offered by top-down ethical theories.

In integrating bottom-up findings, it becomes evident that many technical issues—such as confidentiality, availability, and integrity—can be aligned with broader ethical principles like autonomy, beneficence, and justice. The analysis shows that principles derived from practical scenarios, such as explainability and privacy, can fit within the ethical dimensions outlined in bioethics or AI ethics frameworks. This alignment suggests that bottom-up findings can complement top-down approaches by grounding abstract ethical principles in concrete cybersecurity practices.

In the context of mitigation, bottom-up frameworks that focus on the health sector emphasize the importance of implementing a universal code of conduct specifically tailored to cybersecurity. This code of conduct should address common scenarios such as phishing attacks, penetration testing, and data breaches, offering standardized guidelines for mitigating these risks effectively. This approach helps bridge the gap between theoretical ethics and practical cybersecurity measures by providing actionable steps to address sector-specific threats.

Pragmatic approaches, including those from ENISA and ISO/IEC, tend to focus on classical cybersecurity principles like confidentiality, integrity, availability, and authenticity. These principles are indispensable in ensuring the trustworthiness and resilience of digital systems. For instance, ENISA’s framework explicitly addresses the importance of authenticity and non-repudiation,

underscoring the need to verify user identities and secure data integrity. Although these technical principles are more narrowly defined, they serve as the foundation upon which higher-level ethical concerns—such as privacy, fairness, and safety—can be built.

The integration of technical principles within ethical frameworks offers a pathway for a comprehensive approach. Combining the pragmatic focus on data security with ethical principles from top-down and bottom-up perspectives could lead to more holistic strategies for managing cybersecurity risks. For example, aligning confidentiality measures with ethical requirements for patient autonomy and data ownership would strengthen both ethical compliance and technical robustness in healthcare settings.

Regarding mitigation strategies, pragmatic frameworks like ENISA propose the implementation of a cybersecurity certification scheme. This scheme aims to certify products, services, and procedures as compliant with established cybersecurity standards, enhancing the reliability and security of digital systems. The certification acts as a quality assurance mechanism, ensuring that technological solutions adhere to best practices and regulatory requirements.

The analysis reveals that while top-down approaches offer a philosophical grounding for ethical considerations, they may sometimes lack practical specificity. In contrast, bottom-up frameworks can provide detailed guidance for particular cybersecurity scenarios but may miss the broader ethical implications. Pragmatic approaches fill a critical role by establishing the fundamental principles required for data protection but may not address the ethical conflicts arising in complex situations.

To bridge these gaps, the development of an adaptive, multi-layered ethical framework that incorporates elements from all three approaches is recommended. Such a framework should:

1. Integrate Ethical and Technical Principles: establish clear mappings between ethical principles (e.g., autonomy, justice) and technical principles (e.g., confidentiality, integrity), allowing for seamless application in various scenarios.
2. Adapt to Context-Specific Requirements: tailor ethical guidelines to specific fields, such as healthcare, where the consequences of cybersecurity breaches are particularly severe.
3. Promote Continuous Ethical Review: encourage ongoing assessment and adaptation of ethical guidelines to accommodate new technologies and emerging ethical dilemmas.

Ultimately, a well-rounded framework that combines ethical theories with practical considerations and technical principles can more effectively address the multifaceted challenges of cybersecurity. This approach not only ensures compliance with ethical standards but also enhances the resilience of digital systems by incorporating diverse perspectives on risk mitigation and ethical responsibility.

1.2.4 Adaptive Ethical Cybersecurity Framework

An Adaptive Ethical Cybersecurity Framework that synthesizes insights from the top-down, bottom-up, and pragmatic approaches can create a holistic model for addressing the ethical dimensions of cybersecurity. This model aims to be dynamic and responsive to evolving threats, while also rooted in well-established ethical principles. To achieve this, the framework must integrate core values such as autonomy, justice, non-maleficence, beneficence, confidentiality, integrity, availability, proportionality, transparency, and inclusivity. Each of these principles plays a distinct role in guiding the ethical practices within cybersecurity, while collectively contributing to a balanced and comprehensive approach to digital security.

The principle of **autonomy** serves as a foundational value, emphasizing the right of individuals to exercise control over their personal information and make informed choices about its use. In the digital realm, where data collection and surveillance are pervasive, respecting autonomy means more than simply obtaining user consent. It requires meaningful engagement with users, offering them clear information about how their data will be handled, what risks are involved, and giving them the power to opt-in or opt-out of data-sharing practices. Ensuring autonomy is respected also implies that cybersecurity practices should not covertly compromise user choices, such as implementing hidden data tracking mechanisms without proper disclosure. In this context, autonomy is not just about individual rights but also about empowering users to participate actively in shaping their digital environments.

Justice and fairness are equally critical, ensuring that cybersecurity measures are applied equitably and do not perpetuate existing social inequalities. In a world where access to technology and digital literacy can vary significantly across populations, it is crucial that security protections are designed to be inclusive. For example, while well-resourced organizations may have the means to implement advanced cybersecurity defenses, smaller businesses or individuals from marginalized communities might lack similar capabilities. The ethical framework must therefore strive to create policies that

prevent these disparities from translating into unequal levels of protection, ensuring that everyone benefits from robust security standards. Additionally, justice in cybersecurity extends to addressing digital harms, such as online harassment or data breaches, in a way that does not discriminate against particular groups. A just approach to cybersecurity requires mechanisms for redress and accountability when harm occurs, offering protection and support to those who are most vulnerable.

The principle of **non-maleficence**, the commitment to avoid causing harm, requires cybersecurity professionals to consider not just direct threats like data breaches or malware, but also the indirect consequences of their security measures. For instance, a highly invasive monitoring tool designed to detect insider threats may reduce the risk of data theft, but it could also infringe on employees' privacy, create a culture of distrust, and reduce morale. An ethical cybersecurity framework should weigh such trade-offs carefully, ensuring that the benefits of a security measure significantly outweigh the potential harms. This principle encourages a risk-based approach that not only addresses technical vulnerabilities but also takes into account the social and psychological impact of security interventions. Implementing non-maleficence in cybersecurity involves a commitment to continuous monitoring of outcomes to detect any unintended negative effects and promptly address them.

Complementing non-maleficence is the principle of **beneficence**, which calls for actively enhancing the well-being of individuals and society through cybersecurity efforts. Beneficence goes beyond preventing harm to promote positive outcomes, such as fostering user trust, protecting critical infrastructure, and ensuring safe and secure access to digital services. For instance, by implementing advanced encryption and secure communication channels, organizations can help protect users' sensitive information, thereby reducing the risk of identity theft or fraud. Furthermore, beneficence in cybersecurity entails not just responding to known threats but anticipating future risks and proactively building defenses to protect users from emerging dangers. This forward-looking approach reflects a commitment to continuous improvement in security practices, aiming to create a safer and more resilient digital ecosystem.

Confidentiality remains a fundamental ethical principle in cybersecurity, centered on protecting sensitive information from unauthorized access. While the technical aspects of confidentiality—such as encryption, access controls, and secure data storage—are well established, the ethical

dimension involves ensuring that data protection measures are consistently aligned with user expectations and legal requirements. For example, data sharing agreements between organizations must be transparent and uphold the promises made to users about the confidentiality of their information. This means not only implementing strong technical safeguards but also creating a culture of data ethics where privacy is respected as a core value. The ethical framework should guide organizations in balancing the need for data access with the rights of individuals to keep their personal information private, especially in contexts like healthcare, finance, or social media, where breaches could have severe consequences.

The principle of **integrity** is about ensuring the accuracy, reliability, and trustworthiness of data. In cybersecurity, this involves protecting information from being tampered with or corrupted, either accidentally or through malicious activity. Ensuring data integrity is especially crucial in sectors like healthcare, where incorrect information can lead to life-threatening decisions, or in finance, where inaccurate records can result in significant financial losses. The ethical commitment to integrity extends beyond technical solutions to include practices like implementing audit trails, conducting regular data quality checks, and educating users on the importance of data hygiene. By upholding data integrity, cybersecurity measures can maintain the credibility of digital systems, which is essential for user trust.

Availability ensures that digital services and data remain accessible and usable when needed, which is a fundamental requirement for operational continuity in critical systems. The ethical implications of availability are particularly significant in sectors like healthcare, emergency services, and public safety, where service disruptions can have life-or-death consequences. Ethical cybersecurity practices should therefore prioritize resilience, incorporating measures such as redundancy, backup systems, and robust incident response plans to mitigate the impact of potential disruptions. The principle of availability also involves being prepared for a wide range of threats, from cyberattacks like denial-of-service to natural disasters, ensuring that systems can recover quickly and resume normal operations.

The concept of **proportionality** addresses the need for security measures to be appropriate to the level of risk they are designed to mitigate. This principle encourages a balanced approach that avoids excessive or unnecessarily intrusive security practices. For instance, while biometric authentication can enhance security, using it in contexts where the risks do not justify such an invasive measure

may not be ethically appropriate. Proportionality ensures that security efforts do not disproportionately infringe on individual rights or freedoms, maintaining a balance between effective protection and respecting personal autonomy. This approach requires a nuanced understanding of the context, recognizing that what is appropriate for securing critical infrastructure may not be suitable for everyday digital services.

Transparency and explicability are crucial for fostering trust in cybersecurity. Users need to understand the nature of the security measures in place, why they are necessary, and what implications they might have. This involves communicating policies in a clear, accessible manner and being open about the data practices and security decisions that affect users. Transparency is not only about information disclosure but also about accountability, where organizations must be willing to explain and justify their cybersecurity practices. When security measures are transparent and their rationale is clearly articulated, it helps users feel more in control and confident in the digital systems they interact with.

Finally, **inclusivity** ensures that cybersecurity practices are designed to be accessible and beneficial to all users, regardless of their technical skills, socio-economic background, or physical abilities. This principle requires that security measures do not create barriers or exclude certain groups from accessing digital services safely. Inclusivity also involves engaging a diverse range of stakeholders in the development of security policies, incorporating perspectives from different communities to ensure that security solutions reflect the needs and values of a broad user base. By prioritizing inclusivity, cybersecurity practices can promote digital equity, ensuring that everyone has the opportunity to participate safely in the digital world.

Incorporating these ethical principles into an Adaptive Ethical Cybersecurity Framework provides a comprehensive approach to managing the complex challenges of digital security. The integration of autonomy, justice, non-maleficence, beneficence, confidentiality, integrity, availability, proportionality, transparency, and inclusivity ensures that cybersecurity measures are not only technically robust but ethically sound. This approach supports the development of flexible and adaptive security practices that can respond to new ethical and technological challenges, balancing the need to protect digital assets with the imperative to uphold human values. The result is a dynamic framework that can guide organizations in navigating the evolving landscape of cybersecurity while maintaining a commitment to ethical principles.

Mitigation strategies

To integrate ethical principles into cybersecurity practices effectively, it is essential to adopt mitigation measures that not only address technical requirements but also respect the moral and social implications of digital security efforts. These measures should align with key ethical principles, ensuring that the practices not only protect data and systems but also support broader societal goals, promote fairness, and safeguard individual rights. A more in-depth discussion of these mitigation strategies illustrates how they can be employed to uphold the principles of autonomy, justice, non-maleficence, beneficence, confidentiality, integrity, availability, proportionality, transparency, and inclusivity.

In the realm of **autonomy**, respecting individuals' rights to control their data and make informed choices is a fundamental ethical obligation. To uphold autonomy, cybersecurity practices must ensure that users are provided with meaningful and informed consent options, allowing them to understand exactly what data they are sharing and for what purposes. This means moving beyond vague or generalized consent forms and instead offering detailed, comprehensible explanations that empower users to make truly informed decisions. The implementation of consent dashboards, where users can easily manage their data-sharing preferences, is crucial. These tools should allow individuals to modify their settings, withdraw consent, or delete their personal information, thus reinforcing the ongoing nature of user control. Ensuring autonomy also extends to providing transparency regarding data processing practices, such as explaining how algorithms make decisions that affect users and giving them the opportunity to challenge those decisions if necessary.

Justice and fairness demand that cybersecurity measures are applied equitably and do not disproportionately impact certain groups, particularly vulnerable or marginalized populations. This principle requires a deliberate focus on designing security policies that account for diverse user needs and potential inequalities. For instance, cybersecurity solutions must be accessible to individuals who may lack advanced technical skills, by including user-friendly features and providing guidance tailored to non-experts. Ensuring that security tools are affordable and widely available is also a critical aspect of promoting justice, as it prevents digital security from becoming a privilege only accessible to well-resourced organizations or individuals. Furthermore, organizations should proactively identify any potential discriminatory effects of their cybersecurity policies through equity risk assessments, ensuring that protective measures are distributed fairly and do not exacerbate existing social disparities.

The principle of **non-maleficence**, or avoiding harm, extends beyond technical defenses against cyber threats to include the social and psychological impacts of cybersecurity measures. Organizations must conduct thorough risk-benefit analyses before implementing security protocols to evaluate any potential adverse effects on users. For instance, while tools for monitoring employee behavior might enhance security, they could also infringe on privacy and contribute to a culture of surveillance, thus harming employee morale. To mitigate such risks, it is essential to use the least invasive measures that are effective for the given security objective. This might involve anonymizing user data or implementing automated alerts that do not directly track individuals' activities. A commitment to non-maleficence also involves establishing ongoing feedback mechanisms to monitor the impacts of security measures, allowing for timely adjustments if unintended harms are detected.

Cybersecurity practices should not only avoid harm but also actively contribute to the **beneficence** principle, which calls for promoting the well-being of individuals and society. Measures that go beyond merely defensive postures and aim to improve digital safety proactively align with this principle. For example, organizations can implement proactive threat detection systems that use artificial intelligence to identify emerging threats before they cause significant damage, thus protecting users more effectively. Providing comprehensive security awareness programs is another critical aspect of beneficence; educating users on how to recognize phishing scams, secure their accounts, or safely share information empowers them to take an active role in their digital safety. Additionally, cybersecurity initiatives can extend their positive impact by collaborating with public institutions and non-profits to strengthen the security of critical infrastructures, such as healthcare facilities or utilities, which are essential for public safety.

Maintaining **confidentiality** is a core aspect of data protection, requiring measures that safeguard sensitive information from unauthorized access. While traditional methods such as encryption and access controls are standard practices, it is essential to recognize the ethical dimension of these measures, ensuring they are not just technically robust but also aligned with the expectations of users. Confidentiality protocols should include safeguards against insider threats, such as limiting data access based on the principle of least privilege, where only those who need specific information to perform their job duties have access to it. Moreover, organizations must be transparent with users about any data sharing with third parties, ensuring that they are informed about who has access to their information and why. Protecting confidentiality also involves implementing strong

policies for data retention and deletion, so that sensitive information is not kept longer than necessary, thereby minimizing the risk of data exposure.

The principle of **integrity** focuses on ensuring the accuracy and trustworthiness of data, which is especially crucial in fields where decisions based on data can have significant consequences. For instance, in healthcare, ensuring the integrity of patient records is vital for accurate diagnoses and treatments, while in finance, data integrity prevents fraudulent transactions. Mitigation measures to support data integrity include using cryptographic techniques such as digital signatures, which can detect any unauthorized alterations to data. Organizations should also perform regular data audits and establish version control practices, which track changes and ensure that any modifications are properly authorized and documented. Maintaining robust backup and recovery solutions is another key strategy, allowing organizations to restore original data in the event of corruption or cyberattacks, thus minimizing the potential impact on users.

Ensuring **availability** involves more than just keeping systems operational; it also reflects the ethical responsibility to guarantee that essential services remain accessible during crises or cyber incidents. Cybersecurity measures should therefore include redundancy strategies, such as deploying multiple servers and backup power supplies, to ensure that service disruptions do not occur even if one part of the system fails. Protecting against Distributed Denial of Service (DDoS) attacks, which can overwhelm networks and render services inaccessible, requires the use of specialized anti-DDoS solutions and network traffic filtering. Additionally, performing regular system maintenance and updates helps prevent vulnerabilities that could cause unexpected downtime. In sectors like healthcare or public safety, where service interruptions can have life-threatening consequences, the ethical imperative to ensure availability is particularly strong.

The principle of **proportionality** necessitates that cybersecurity measures are appropriate to the level of risk involved. Organizations must avoid using overly restrictive or invasive measures in low-risk situations, which could unnecessarily infringe on user privacy or freedoms. For instance, while biometric authentication can provide strong security, it may not be justified for securing everyday accounts where the potential impact of unauthorized access is minimal. By conducting risk assessments, organizations can better understand the threat landscape and apply security measures that match the level of risk. Proportionality also involves conducting privacy impact assessments before implementing new security technologies, ensuring that the benefits outweigh any potential drawbacks.

Transparency and explicability in cybersecurity are essential for building user trust. When organizations are open about their data practices and the security measures in place, users can make more informed choices and feel more confident in the digital systems they use. This transparency should include providing detailed explanations about how data is collected, processed, and stored, along with clear communication about users' rights regarding their data. In the event of a security incident, organizations should promptly inform affected users, explaining what happened, what is being done to mitigate the problem, and how similar incidents will be prevented in the future. Documenting the rationale behind security policies and decisions not only supports transparency but also enables stakeholders to understand and evaluate the ethical considerations taken into account.

Lastly, the principle of **inclusivity** ensures that cybersecurity practices do not exclude or disadvantage certain user groups. It is important to involve a diverse range of stakeholders in the design of security measures to account for different needs and perspectives, particularly those of underrepresented communities. For instance, designing interfaces that are accessible to people with disabilities, providing content in multiple languages, and offering customizable security settings can help ensure that security tools are usable by everyone. Organizations should also adhere to accessibility standards, such as the Web Content Accessibility Guidelines (WCAG), to ensure that cybersecurity tools are inclusive and do not create barriers for individuals with disabilities.

By thoroughly implementing these mitigation measures, organizations can integrate ethical principles into their cybersecurity practices, creating a robust framework that goes beyond mere compliance with technical standards. This approach helps address the complexities of digital security, recognizing the interconnectedness of technical measures and ethical considerations. It ensures that cybersecurity strategies are not only effective in defending against threats but also in respecting the rights and dignity of individuals, promoting trust, fairness, and the overall well-being of society.

1.3 Ethical frameworks for data protection

In the realm of cybersecurity, data protection emerges as a crucial area with its own distinct ethical considerations. While cybersecurity focuses on protecting systems and networks against external and internal threats, data protection centers more specifically on individuals' rights and the responsible handling of personal data. This encompasses not only securing the data but also upholding fundamental rights such as privacy and informational self-determination. Although there

is some overlap between cybersecurity and data protection within the broader field of information security, differentiating them enables a more tailored approach to the unique issues and ethical requirements of each field. To gain a deeper understanding of these issues, we now turn our attention to the ethical frameworks governing data protection for addressing the ethical implications of managing personal information in an increasingly digital world.

The ethics of data protection are closely tied to principles like autonomy, human dignity, and respect for individuals, especially when personal data is compromised or lost, as often occurs in cybersecurity incidents [28], [29]. Furthermore, data protection is recognized as a fundamental human right within the EU Charter of Fundamental Rights and the Treaty on the Functioning of the European Union, and is rigorously enforced through the General Data Protection Regulation (GDPR) [28], [30], [31], [32], [33]. The GDPR not only protects individuals' privacy rights but also grants them control over every phase of their personal data's lifecycle—from collection and processing to storage, transfer, and deletion—thus establishing comprehensive standards for data handling, security, and accountability across the EU.

1.3.1 Fundamental ethical frameworks

In this section, three frameworks are included as a result of an assessment of the most relevant frameworks in the data protection landscape with regard to ethics. The three frameworks have been developed by different actors (Table A-12, Table A-13, Table A-14) : IBM, Independent Data Protection Supervisory Authorities, and the World Health Organization (WHO). These frameworks have been included to make some of the following subparagraphs more comprehensible, which concern a quantitative assessment of ethical risks related to data protection and the development of useful tools for monitoring risks. This tool will then be used in the context of business intelligence. Table A-12 from IBM outlines essential principles, issues, and risks related to data protection. It emphasizes core values such as ownership, transparency, privacy, and responsible use of AI. Each principle is paired with case studies or examples of how these challenges are addressed and offers potential mitigations to ensure data protection integrity.

1. Ownership (Upholding Data Integrity)

The concept of ownership is fundamental in data protection. Just because a user provides data doesn't imply that the organization has ownership of that data. This distinction is crucial, as it

touches on the ethical use of data and the responsibility that comes with handling personal or sensitive information. Organizations must acknowledge that the data remains the property of the user, and they are merely custodians of it. Failure to respect this principle can result in significant breaches of trust and compliance issues. Mitigation strategies here involve ensuring that companies uphold data integrity through consent-based practices that respect user rights and adhere to data protection laws.

2. Transparency (Mistrust)

Transparency plays a critical role in maintaining trust between a company and its users. The table highlights the importance of being explicit about how customer data is used. Mistrust can arise when organizations fail to clearly communicate the purposes and methods of data usage, particularly in an era where big data is increasingly leveraged for commercial gain. To counteract this, organizations are encouraged to empower users with a comprehensive understanding of the data lifecycle, from collection to deletion, ensuring they are well informed and comfortable with the processes.

3. Privacy

Privacy is perhaps one of the most discussed issues in the modern data landscape. The table emphasizes that companies must only use, store, share, or retain information for the original purposes for which it was obtained. Expanding on this, organizations must implement robust privacy practices that not only comply with regulations like GDPR but also exceed them to ensure ethical handling of data. The potential misuse of data beyond the agreed-upon purposes is a significant risk, and companies must put in place data ethics policies and security mechanisms to guard against this. Mitigation includes clear guidelines on how data should be managed through its lifecycle, as well as how it should be disposed of when no longer needed.

4. Intention (Responsible AI; Trust in Technology)

The table touches on the responsible use of AI, a crucial area as businesses increasingly rely on AI-driven analytics and decision-making. Being clear about the purpose of AI in handling user data

fosters trust. Users must know that their data is used ethically and that the technology serving them is trustworthy. IBM, as noted in the table, is developing tools such as explainer toolkits and AI governance solutions to increase transparency and trust in AI systems. These measures are designed to make AI more understandable, reducing the risk of its misuse and ensuring it aligns with ethical standards.

5. Prevention (Data Breaches, Ransomware Attacks)

Prevention is centered around minimizing risks such as data breaches, ransomware attacks, and other cybersecurity threats. According to the table, IBM's research reveals that companies integrating AI and automation into their security frameworks save significantly on data breach costs, highlighting the financial benefits of proactive security measures. AI and automation enable faster detection and response to threats, which is critical in preventing attacks from escalating. Risk management strategies, bolstered by AI, are essential in ensuring that companies remain resilient in the face of increasingly sophisticated cyber threats.

Table A-12 from IBM's data protection framework offers a comprehensive look at the key principles organizations must adhere to in today's complex data landscape. It underscores the importance of ownership, transparency, privacy, responsible AI, and preventive measures in ensuring data integrity. Mitigation strategies range from enforcing robust security frameworks to developing user-friendly tools that enhance transparency and build trust. As data becomes an ever more valuable asset, adherence to these principles is critical not only for regulatory compliance but also for maintaining strong, trust-based relationships with users.

Table A-13 from the Independent Data Protection Supervisory Authorities outlines a comprehensive standard model for data protection, detailing various principles and risks alongside corresponding mitigation strategies. Each section ties in closely with GDPR requirements, offering both theoretical underpinnings and practical examples of how to safeguard data.

1. Availability

Availability refers to the need for continuous and immediate access to personal data and its processing, ensuring that no delays impact the proper functioning of data. This principle is critical in the context of system resilience, recoverability, and mitigation strategies as stipulated by the GDPR (e.g., B1.18 Availability, B1.19 Resilience, B1.20 Recoverability). To ensure availability, organizations are encouraged to implement redundant systems, establish detailed backup protocols, and prepare contingency plans. These measures mitigate the risks of data loss or system failure, especially in cases of external threats like malware or natural disasters.

2.Integrity

Data integrity is bifurcated into two main concepts: the ongoing compliance of IT systems with predefined specifications and the requirement that data remains intact, complete, and correct. According to GDPR articles such as B1.4 (Correctness) and B1.6 (Integrity), the integrity of data is safeguarded by restricting write and modification permissions, using cryptographic methods, and regularly testing systems for security gaps. Maintaining data integrity ensures that the information remains unaltered from its original state, which is crucial for both organizational decision-making and legal compliance.

3.Confidentiality

Confidentiality ensures that only authorized individuals can access personal data. This principle is vital for maintaining trust between data controllers and subjects, and it's explicitly required under GDPR (B1.7). To ensure confidentiality, organizations must implement secure authentication processes and limit access to personnel who are thoroughly vetted. Encryption and secure communication channels are key strategies for preventing unauthorized data access, particularly in the face of growing cybersecurity threats. External protections, such as guarding against hacking and espionage, are also integral components of this principle.

4.Unlinkability

Unlinkability focuses on preventing personal data from being combined or linked across systems or processes. This principle, highlighted under GDPR (B1.2), is crucial for maintaining privacy, especially when dealing with large data sets. Organizations should use pseudonymization, anonymization, and strict separation of roles to enforce unlinkability. These measures are essential for minimizing the risk of creating detailed personal profiles or re-identifying individuals from anonymized data, thereby protecting privacy rights.

5.Transparency

Transparency requires that data subjects, system operators, and supervisory authorities are all able to clearly see how personal data is collected, processed, and used. GDPR emphasizes this with regulations on documentation and accountability (e.g., B1.8). Organizations must maintain clear records of all data processing activities, including business processes, data flows, and system interactions. This ensures accountability and allows data subjects to be informed about how their personal data is being handled. By fostering transparency, organizations not only comply with legal obligations but also build trust with the public and stakeholders.

6.Intervenability

Intervenability guarantees that data subjects can exercise their rights, such as the right to rectification, erasure, and objection, in a timely and effective manner. This principle emphasizes the need for robust mechanisms that allow individuals to challenge the processing of their data, particularly in the context of automated decisions. GDPR requirements (e.g., B1.10, B1.11, B1.12) are clear in mandating processes for correcting or erasing incorrect data, handling objections, and ensuring the proper use of consent management. These measures empower data subjects, ensuring that their rights are not merely theoretical but practically enforceable.

7.Data Minimization

Data minimization is one of the most critical tenets of data protection law, requiring that the collection and processing of personal data be limited to what is necessary. As mandated by the GDPR (B1.3), this principle focuses on reducing the amount of data collected and limiting its use to specific, essential purposes. By default, systems should be designed to collect only the minimum necessary data, and anonymization or pseudonymization techniques should be employed wherever possible. This approach not only reduces the risk of data breaches but also ensures compliance with data protection standards by limiting the exposure of personal information.

This model represents a highly structured approach to managing data protection risks by aligning specific principles with concrete, actionable mitigation strategies. The integration of GDPR requirements ensures that all dimensions—ranging from availability to data minimization—are comprehensively covered, allowing organizations to address the multifaceted nature of data protection in a holistic manner. Implementing this model fosters not only legal compliance but also operational excellence in handling sensitive personal data.

Table A-14 from the World Health Organization (WHO) focus on the protection of personal data in health information systems (HIS). It covers various key aspects, principles, and risks associated with data management in health systems, along with case studies and mitigation strategies. In particular, the table outlines fundamental principles that HIS must follow to ensure the ethical and legal processing of personal data. These principles are designed to protect the privacy of individuals while allowing the use of their data for public health purposes.

1. Fair, Lawful, and Transparent Processing

This principle emphasizes the need for fairness, legality, and transparency in data processing. Organizations handling personal data must do so only if it is permitted by law or with the explicit consent of the data subject. This ensures that individuals are aware of how their data is being used, which fosters trust in health information systems. For example, in a healthcare context, patients need to be informed about the purposes for which their medical data is being collected and how it will be protected.

To ensure compliance, institutions are advised to establish a comprehensive risk management system for data protection. This should cover risks across various domains, including financial and reputational risks. A strong recommendation is to align with internationally recognized standards like ISO 27001, which provides a systematic framework for managing information security. Regular audits, both internal and external, help institutions maintain accountability. Additionally, conducting Data Protection Impact Assessments (DPIAs) for high-risk activities is crucial for proactively identifying and addressing potential privacy issues.

2. Purpose Limitation

Personal data should only be collected for specific, lawful purposes. This principle ensures that data isn’t reused in ways that are incompatible with its original purpose. For instance, if a patient’s health data is collected to provide medical treatment, it should not later be used for unrelated research without obtaining additional consent.
The lack of a case study or mitigation suggestions here might indicate a broad agreement on this point or an assumption that the measures mentioned for "Fair Lawful and Transparent" processing will also address this concern. However, setting strict internal policies and training employees about the purpose of data collection can ensure this principle is respected.

3. Accuracy

Ensuring that personal data remains accurate and up to date is critical, particularly in health information systems, where incorrect data could lead to misdiagnosis or improper treatment. Health institutions must implement mechanisms to verify and update patient information regularly.

4. Data Minimization

Data minimization restricts the collection of personal data to only what is necessary for a specific purpose. In healthcare, this principle supports the idea that patients' sensitive data should not be collected or stored unless absolutely required. This reduces the risk of exposure and misuse of unnecessary data.

Health institutions must evaluate which data is essential for treatment or research purposes and avoid over-collection. Minimization is a proactive step toward reducing the risks of a data breach, as less data is exposed to potential threats.

5. Storage Limitation

Personal data should not be retained for longer than necessary. This principle encourages organizations to develop clear retention policies, especially for sensitive health data, and to delete or anonymize data once it is no longer needed. Long-term storage of personal data increases the risk of unauthorized access or data breaches, so limiting the storage period is a key mitigation strategy.

6. Rights of Data Subjects

Under applicable data protection laws, individuals (data subjects) have certain rights regarding their personal data. These rights may include the right to access, correct, delete, or restrict the processing of their data. Health organizations must respect these rights and establish procedures to allow individuals to exercise them.
For instance, patients should be able to easily request access to their medical records, and institutions should have clear processes in place to comply with these requests in a timely manner.

7. Integrity and Confidentiality

Ensuring the integrity and confidentiality of personal data involves implementing appropriate technical, legal, and organizational measures to prevent unauthorized access or accidental loss. Health data is particularly sensitive, and any breaches could have severe consequences for both patients and institutions.
This principle likely refers to the implementation of strong encryption, regular security audits, and robust access control mechanisms. Staff training on data confidentiality is another critical measure, ensuring that all employees understand their role in maintaining data security.

8. International Transfer of Personal Data

The final principle addresses the risks associated with transferring personal data across borders. When data is transferred to another country or international organization, it must be ensured that the receiving entity provides an adequate level of protection in line with international standards. This is particularly important for global health initiatives or research collaborations involving data sharing across different jurisdictions.

For healthcare providers and research institutions that work internationally, this principle requires careful scrutiny of data-sharing agreements and compliance with varying data protection laws across different countries.

Table A-14 offers a comprehensive framework for managing the privacy and protection of personal data within health information systems, based on WHO principles. By adhering to these principles, institutions can not only comply with legal obligations but also maintain the trust of the public, which is essential in the health sector. Each principle is underpinned by clear mitigation strategies that range from risk management systems to adherence to international standards. The emphasis on audits, impact assessments, and compliance with standards such as ISO 27001 indicates a structured and proactive approach to data protection.

The principles in the table stress the importance of a holistic and systematic approach to data protection, with particular emphasis on leadership and accountability at the highest levels of the organization. Data protection is not just a technical issue but one that requires a cultural shift within organizations to prioritize privacy at every level.

1.3.2 Observations and synthesis of the frameworks

An analysis of various data protection frameworks highlights two primary observations: the existence of a common terminology and conceptual overlap; the integration of the various models into the SDM framework.

Many principles across different frameworks share common terminology, such as "availability" and "integrity." These terms often recur in data protection frameworks because they represent foundational aspects of information security, ensuring that data remains accessible and unaltered. Other principles, while distinct in name, may function as subsets of broader concepts. For example, the principle of "storage limitation" under GDPR (General Data Protection Regulation) can be seen

as a specific aspect of "data minimization," as both aim to limit the unnecessary retention of personal data. This overlap in terminology and scope reflects a shared understanding of core data protection values, though different frameworks may emphasize certain aspects depending on their regulatory or operational focus.

The SDM framework integrates principles from other models, notably incorporating the core concepts of GDPR. The SDM is built on the GDPR's principles, such as lawfulness, fairness, transparency, data minimization, and purpose limitation, among others. By organizing these principles systematically, the SDM framework provides a coherent and structured approach to data protection that addresses gaps and aligns overlapping concepts. The result is a more unified framework that not only maintains compliance with GDPR but also provides a comprehensive guide for managing data protection practices consistently across different contexts.

Therefore, the unified framework proposed by the SDM is considered more holistic and systematic. It consolidates the principles from the GDPR and other models, organizing them in a way that reduces redundancy and aligns the various data protection requirements under a single coherent structure. This integration helps organizations adopt a more consistent and efficient approach to data protection while remaining compliant with relevant regulations.

1.4 Ethical framework for adversarial machine learning

Adversarial machine learning (ADM) attacks can be seen as failures in cybersecurity because they exploit vulnerabilities in artificial intelligence systems, resulting from inadequate implementation or protection of these systems. If a machine learning model is vulnerable to adversarial attacks, it indicates that its design, training, or operational environment is not robust enough to ensure the system's overall security. These vulnerabilities extend traditional cybersecurity challenges but have a unique characteristic: instead of directly targeting computer systems or networks, the attack manipulates the characteristics of the system, causing the system to make mistakes.

Separating the ethical frameworks allows for a more precise approach to these specific ethical challenges and the development of solutions that consider not only information and network security but also the robustness and integrity of machine learning algorithms.

However, including these distinct frameworks within the broader chapter on cybersecurity acknowledges that, ultimately, adversarial machine learning attacks represent a failure to ensure comprehensive cybersecurity, broadening the definition of what constitutes a security threat in the age of artificial intelligence.

In summary, keeping the frameworks separate enables a targeted approach to the specific vulnerabilities and defenses associated with machine learning, while the connection to traditional cybersecurity highlights that weaknesses in AI models are, fundamentally, manifestations of broader failures in protecting information systems.

Continuing the approach from previous sections, we will examine the ethical considerations outlined in a NIST-developed framework concerning AML attacks. AML is a field that studies the security aspects of ML algorithms when under attack by adversaries. It encompasses both the development of attacks aimed at deceiving ML models and the creation of defenses to protect against such attacks [34], [35]. The field has gained significant attention due to the widespread application of ML across various domains and the corresponding increase in cybersecurity threats. One interesting domain in which ML systems are spreading is the healthcare sector, which represents a critical infrastructure whose data can be highly sensitive. As an example, ML algorithms are used to predict mild cognitive impairment (Figure 0-1) and related dementia states**.** This is an example of predictive artificial intelligence (Predictive AI) that I worked on for a national project.

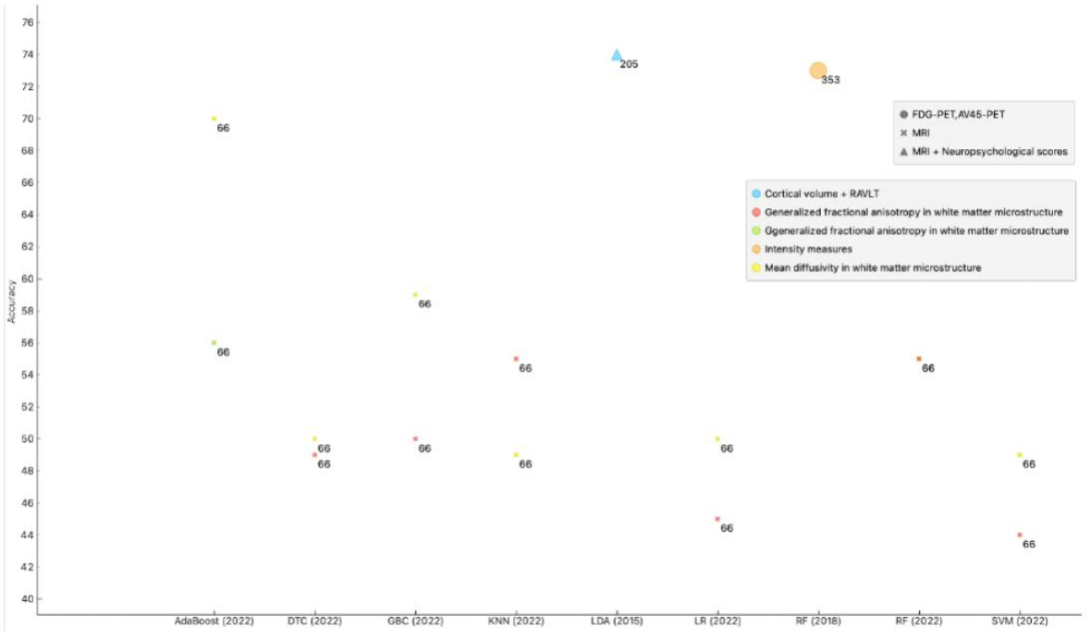

Figure 0-1. Accuracy of the classification methods used in each study (year) for the early detection of Mild Cognitive Impairment. Features (colors), neuroimaging technique (shape) and sample size (number) are represented. From Orange3 Software.

1.4.1 A framework for predictive artificial intelligence

Predictive AI refers to a branch of AI focused on using historical data to make predictions about future events or behaviors. Predictive AI models are trained on past data to identify patterns and relationships that can be used to forecast outcomes or make informed predictions based on new input data. In this section, we examine the various phases at which an adversary can compromise the system, the potential attack methods during these phases (Figure 0-2), and the specific targets of such attacks. Subsequently, we explore the ethical risks identified by NIST [36].

Poisoning attacks are adversarial attacks against ML that occur during training. They can be categorized as follows:

- **Data poisoning**: poisoning attacks in which a part of the training data is under the control of the adversary [36]
- **Model poisoning**: poisoning attacks in which the model parameters are under the control of the adversary [36]

Evasion attacks are adversarial attacks against ML that occur during the deployment phase. The attacker's goal is to generate adversarial examples, which are test samples that can be minimally altered to cause the system to misclassify them into any category chosen by the attacker at the time of deployment [19].

Privacy attacks are adversarial attacks against ML that occur during the deployment phase. Different privacy attacks exist [36], starting with **data reconstruction attacks**, where an adversary attempts to reverse engineer private information from aggregate data; **membership inference attacks**, where an adversary determines whether a specific record is used in a dataset for training a ML model; **model extraction attacks**, where an adversary extracts details about an ML model, such as its architecture or parameters; and **property inference attacks**, where an adversary aims to reveal the global properties of a training dataset, such as sensitive attribute distributions. Privacy attacks can be categorized in:

- **Data privacy**: attacks against ML models to extract sensitive information about training data [36]
- **Model privacy**: attacks against ML models to extract sensitive information about the model [36]

1.4.2 Ethical risks for predictive AI

Depending on the type of the attack, we will find a specific ethical risk, which can be availability breakdown, integrity violations, privacy compromise.

In data poisoning, the ethical risk lies in the availability breakdown, whereas in model poisoning, the ethical risks involve both the availability breakdown and integrity violations. In evasive attacks, the ethical risk concerns integrity violations. In attacks targeting data and model privacy, the ethical risk is related to privacy compromise (Figure 0-3). Below, we'll look at each risk specifically.

Availability

The ethical risk of availability refers to the threat posed by malicious actors who aim to disrupt the performance and reliability of ML models during their deployment. These attacks, known as availability attacks, target the system's ability to function properly and deliver accurate results. There are several ways an attacker can compromise availability:

1. **Data Poisoning**. The attacker introduces malicious or misleading data into the training dataset, compromising the model's ability to learn correctly. This can result in poor performance or even complete failure during deployment. This type of attack has been demonstrated in various ML models, including support vector machines, linear regression, and neural networks.
2. **Model Poisoning**. Here, the attacker manipulates the model's parameters directly, either during training or in scenarios like federated learning, where the model is trained across multiple devices. The aim is to degrade the model's accuracy or functionality.

Ethically, these risks can undermine trust in ML systems, lead to financial and operational damage, and even pose safety concerns if critical systems are affected, such as those used in healthcare or autonomous driving.

Integrity

The ethical risk of integrity refers to the danger posed by attacks aimed at compromising the accuracy and trustworthiness of ML model outputs. These integrity violations result in incorrect

predictions by the model, which can have serious implications depending on the application. There are several types of integrity attacks:

1. **Evasion Attacks**. These occur at deployment time, where the attacker modifies the input samples to create adversarial examples that cause the model to misclassify them. The modifications are subtle and often imperceptible to humans, allowing the attack to go undetected while still manipulating the model's output.
2. **Poisoning Attacks**. These occur during the training phase and can be classified into different types:
    - **Targeted Poisoning Attacks**. The attacker introduces malicious data into the training set, aiming to affect the model's performance on specific, targeted samples. This requires control over some portion of the training data.
    - **Backdoor Poisoning Attacks**. The attacker embeds a "backdoor pattern" in the training data that is also present in the testing phase. When the pattern appears, the model is manipulated to produce a specific incorrect output. These attacks necessitate control over both training and testing data.
    - **Model Poisoning**. The attacker modifies the model's parameters, potentially through federated or centralized learning processes, to introduce vulnerabilities that can lead to either targeted or backdoor integrity attacks.

The ethical risks associated with integrity attacks include potential harm to users who rely on ML models for decision-making, especially in critical domains like healthcare, finance, or autonomous systems. Compromising model integrity can erode trust in AI systems, lead to biased or harmful outcomes, and enable malicious entities to exploit the system for personal or financial gain.

Privacy

The ethical risk of privacy involves the threat of attackers compromising the confidentiality of either the training data used to build ML models or the models themselves. Privacy attacks can undermine trust in AI systems and expose sensitive information. There are two main types of privacy compromises:

1. **Data Privacy Attacks**. These focus on learning information about the training data, with several possible objectives:
    - **Data Reconstruction**. The attacker attempts to infer the content or features of the original training data, potentially revealing sensitive information about individuals or entities.
    - **Membership-Inference Attacks**. The attacker seeks to determine whether a particular data point was part of the training set, which could reveal personal or confidential information.
    - **Data Extraction**. Involves extracting pieces of the training data from a predictive model, potentially exposing sensitive data that the model was trained on.
    - **Property Inference**. The attacker aims to infer statistical properties or characteristics of the training data distribution, which may expose private attributes about the dataset as a whole.
2. **Model Privacy Attacks.** These target the ML model itself, aiming to extract detailed information about the model's structure, parameters, or training processes. Model Extraction allows attackers to replicate or steal the intellectual property of the model, or even use it to conduct further privacy attacks.

The ethical risks associated with privacy compromises are significant. They can lead to unauthorized exposure of sensitive information, violating individual privacy rights, and breaching data protection regulations. Moreover, successful privacy attacks may erode public trust in AI and ML technologies, hinder innovation, and even lead to legal and financial repercussions for organizations that fail to safeguard their data and models adequately.

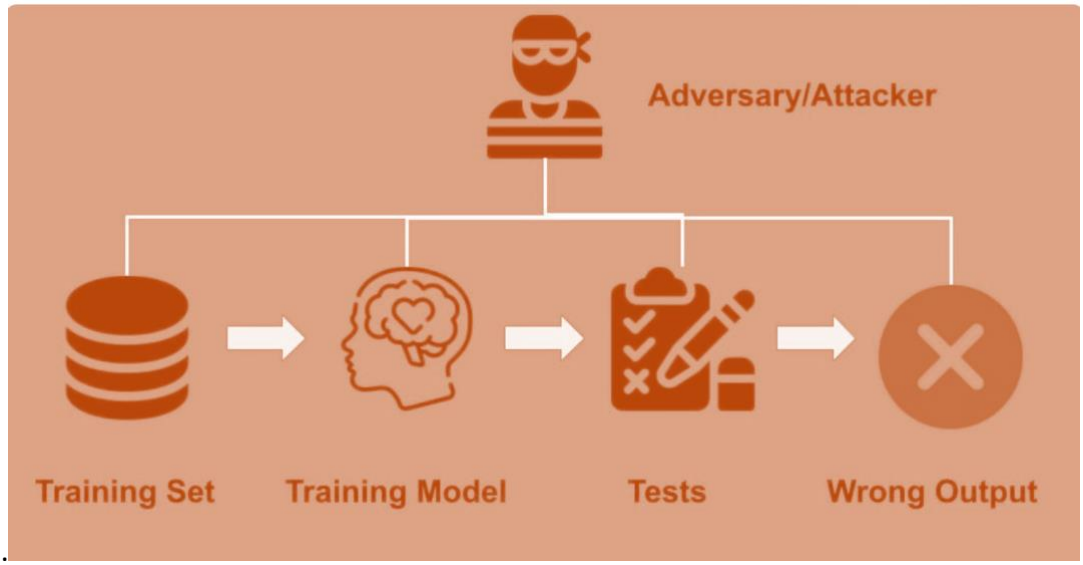

Figure 0-2. Phases of a machine learning system that can be attacked by an attacker or adversary.

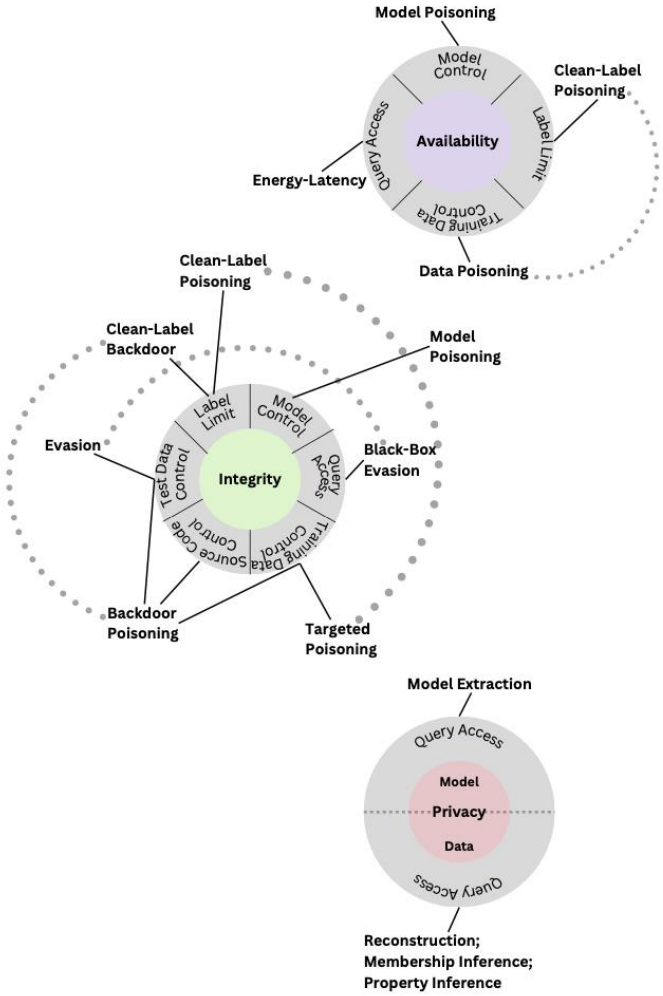

Figure 0-3. Attacks and ethical risks identified in Predictive AI systems [37].

Chapter 2: Models and methodologies for ethical risk assessment

In this chapter, we delve into various models and methodologies for assessing ethical risk, exploring both defensive and offensive strategies.

Defensive strategies typically focus on minimizing harm, mitigating risks, and safeguarding against ethical breaches before they occur. These approaches are designed to prevent negative impacts by identifying potential issues early and implementing safeguards to limit their effects.

On the other hand, offensive strategies take a more proactive stance. They often aim to enhance ethical performance and drive positive change, going beyond mere compliance to actively promote ethical values.

Through this chapter, we will gain insights into how these diverse models and methodologies can be applied across various contexts, offering a comprehensive view of ethical risk assessment.

2.1 Defensive strategies

We discuss various models used in defensive methodologies for ethical risk assessment, particularly focusing on evaluation techniques that range across qualitative, semi-quantitative, and quantitative methods.

1. Qualitative Model: this model relies on non-numeric categories to classify risk levels, making it easy to interpret and quick to apply. However, its major limitation is the subjective nature of its results, which can vary greatly based on the evaluator's perspective. This subjectivity leads to challenges in reproducing results and comparing them across different contexts.
2. Semi-Quantitative Model: this approach bridges the gap between qualitative and quantitative evaluations by incorporating numeric categories. While it introduces some objectivity, it still retains a level of simplicity in implementation compared to fully quantitative models.
3. Quantitative Model: this is the most data-driven approach, using fully numerical categories to quantify risk. Quantitative models offer robust, comparable, and reproducible outcomes, providing clear metrics for decision-making. However, the trade-off is its high cost in terms of time and resources. Additionally, quantifying complex elements of risk—such as the probability of an event and its potential severity—poses significant challenges.

These models can be applied to evaluate ethics in specific domains such as information technology, data protection, and machine learning, demonstrating the broad applicability of these evaluation tools in modern technological contexts.

2.1.1 Qualitative model for assessing ethics risks for data protection

The document released by the European Commission "Ethics and data protection" gives rise to the qualitative model for evaluating ethical risk in data protection according to a decision-making tree (Figure A-1).

Figure A-2 and Figure A-3 show this decision tree for ethical assessment in data protection structured around a series of yes/no questions. This setup is an example of a qualitative risk assessment method, where risks are assessed not by numerical scoring but by guiding the user through a sequence of simple, descriptive choices.

In this decision tree, each yes/no question is designed to help identify whether a specific ethical risk is present at each stage of data handling. If the answer is “yes,” it means that a potential ethical issue exists, prompting the user to take further steps to manage or mitigate that risk. A “no” answer, on the other hand, suggests that no further action is needed in that particular area, allowing the user to move on without unnecessary checks.

For instance, the tree might ask, “Does the research involve processing of personal data?” If the answer is “yes,” it leads to additional questions about how that data will be managed ethically. If the answer is “no,” the researcher can skip related steps, streamlining the process and focusing only on areas where risks are actually present.

One of the strengths of this yes/no approach is its step-by-step structure, which simplifies risk management. Rather than confronting the user with a complex set of requirements all at once, it breaks down decisions into manageable steps. Each "yes" answer leads to specific follow-up actions or additional questions, progressively building a path that adapts to the unique needs of the project. For example, if the answer to “Does the research involve privacy-intrusive techniques?” is “yes,” the researcher is instructed to provide a detailed description of these techniques and the safeguards in place. But if the answer is “no,” they can move on to the next question without needing to document any specific protections for those techniques, saving time and focusing resources where they’re actually needed.

Unlike quantitative methods that use numbers to assess risk levels, this approach focuses on qualitative judgments. A “yes” answer doesn’t quantify risk but signals that attention and action

are required in that area. This makes the process accessible to people who may not be familiar with statistical or numerical risk evaluation but who still need to ensure ethical standards are met.

For example, when asked “Is the research processing sensitive categories of data (like genetic or health data)?” a “yes” answer doesn’t assign a numerical risk level. Instead, it prompts the researcher to document why this data is necessary and to implement specific safeguards, recognizing that sensitive data requires more careful handling.

This decision tree not only helps identify risks but also ensures compliance with legal and ethical standards through specific documentation steps. Each “yes” response in the tree leads to actionable requirements. If a particular aspect of the research poses an ethical risk, the decision tree provides clear instructions on what the researcher should document or implement to manage that risk.

For instance, if data is being transferred outside the EU to a country without recognized data protections, a “yes” answer in the tree directs the researcher to document the legal basis for this transfer and explain the protections in place for the data. This helps ensure that the research aligns with GDPR and other relevant data protection regulations.

The cumulative structure of this decision tree also reflects the escalation of ethical risk: the more “yes” answers, the more sensitive the project becomes, and the more attention is required to ensure ethical compliance. Each layer of questions represents an additional qualitative measure of risk, requiring increasingly detailed actions or reviews.

For instance, an initial “yes” to processing personal data might lead to further questions about whether sensitive data is involved, whether privacy-intrusive methods are used, and if cross-border data transfers are planned. Each “yes” answer adds a layer of ethical scrutiny, guiding the researcher towards a comprehensive assessment as the potential risk level rises.

In summary, this yes/no decision tree is a great example of a qualitative method for ethical risk assessment. It simplifies complex ethical decisions by categorizing them into clear, actionable responses. Researchers don’t need to calculate precise risk scores; instead, they are led through a thoughtful process that adapts based on the specific characteristics of their project. This approach is particularly useful in areas where ethical considerations are crucial but where stakeholders may not be familiar with quantitative risk assessments. Through structured guidance, this method ensures that ethical risks are identified, documented, and mitigated in a way that is accessible, efficient, and compliant with legal requirements

Practical example

The decision tree for ethics and data protection (Figure A-2, Figure A-3), aimed at guiding researchers and organizations through the process of ethically handling personal data. This type of flowchart assists in identifying potential ethical risks, ensuring GDPR compliance, and applying "ethics by design" principles in data processing activities.

Key Components and Flow

1. Initial Question:
    a. The decision tree begins by asking whether the research involves the processing of personal data. This initial question is essential as it determines the relevance of the entire ethical and data protection evaluation.
2. Alternative Solutions and Minimization:
    a. If personal data is involved, the next set of questions explores whether research objectives could be met using anonymized data or pseudonymized data, which are less sensitive and pose lower ethical risks.
    b. There is also a prompt to justify why personal data is necessary if the use of alternative data is not possible.
3. Risk Mitigation and Justification:
    a. For projects that require personal data, the tree guides users to assess the specific risks and decide on methods to mitigate them. This step is essential to demonstrate compliance with GDPR's accountability principle, where each use of data should have a documented and justified purpose.
4. Data Minimization:
    a. A key step asks whether the researcher has conducted a data minimization review—a process of ensuring only the necessary amount of data is collected. This follows GDPR principles, which emphasize limiting data to what is strictly necessary for the intended purpose.
5. Informed Consent:
    a. The decision tree addresses the importance of informed consent, with questions focusing on whether individuals involved in the project have been adequately informed and have provided consent for data use.
6. Special Categories of Data:

a. Additional questions deal with processing data concerning vulnerable groups, such as children. Special protections are required here, as this data category poses higher ethical risks.

7. Technical and Organizational Measures:
    a. The tree prompts users to consider technical and organizational safeguards to protect data subjects' rights, such as encryption or anonymization.
8. Data Protection Officer (DPO) Involvement:
    a. The tree checks whether the institution has a Data Protection Officer (DPO), who can provide oversight and guidance, ensuring that data protection requirements are met throughout the project.
9. Special Categories of Data
    a. The tree asks if the research involves processing special categories of data, which include genetic, biometric, health-related data, or information on political and religious beliefs. If the answer is “yes,” specific requirements are outlined:
        i. Justification: researchers must justify the need for using these data types.
        ii. Safeguards for Rights and Freedoms: measures should be detailed in the proposal to protect individuals’ rights and freedoms.
        iii. Ethical Approval: clearances or ethical approvals are often required due to the sensitive nature of this data.

10. Privacy-Intrusive Techniques
    a. Next, the decision tree addresses whether the research employs privacy-intrusive techniques, such as surveillance or tracking:
    b. If it does, researchers are required to provide detailed descriptions of these techniques, their objectives, and the safeguards in place to protect individuals' rights.
    c. This includes involving a Data Protection Officer (DPO) if applicable, to ensure compliance with relevant privacy regulations.

11. Re-Use of Previously Collected Data
    a. If the research involves further processing of previously collected data, the tree outlines requirements such as:
    b. Providing documentation on the legal basis for the re-use.

c. Clarifying the purpose of the additional data use and any additional safeguards implemented.

12. Publicly Available Data
    a. For projects that use publicly available data, the decision tree indicates fewer specific requirements, assuming that data is available within the bounds of the original purpose of publication. However, it emphasizes ensuring that the data usage complies with the terms of its publication.
13. Cross-Border Data Transfers
    a. The tree also addresses the complex issue of cross-border data transfers:
        i. Data Export from the EU to Non-EEA Countries: if data is being transferred outside the EU/EEA to countries not recognized as having adequate data protection, specific requirements must be met to protect the data.
        ii. Data Import into the EU from Non-EU Countries: when bringing data into the EU from non-EU countries, the type of data and origin must be documented to ensure compliance with GDPR regulations.
14. High Ethics Risks
    a. Finally, the tree assesses whether the data processing presents high ethical risks to participants:
    b. If high ethical risks are involved, researchers are required to evaluate these risks in-depth, implement detailed mitigation measures, and consult with a DPO or ethics board as necessary.
    c. This step ensures that research involving potential ethical concerns is thoroughly reviewed and that appropriate safeguards are put in place.

This section of the decision tree highlights specific requirements for handling complex or high-risk data scenarios. Each step emphasizes GDPR compliance, the necessity of robust documentation, and consultation with relevant authorities to ensure that data processing activities uphold ethical standards and legal obligations. This approach helps organizations navigate intricate regulatory requirements and ethically sensitive situations in research and data management.

2.1.2 Quantitative model for assessing security risks in data processing operation

Still in the area of data protection, we now turn to a quantitative model developed by the Bimind team (https://www.bimind.it). Before showing the quantitative model, it is necessary to know when this model can be used in a Data Protection Impact Assessment (DPIA).

DPIA is a critical process required under the GDPR for assessing and mitigating risks associated with the processing of personal data, particularly in projects likely to impact the rights and freedoms of individuals. DPIAs are designed to identify potential data protection issues early on, evaluate the necessity and proportionality of data processing activities, and implement appropriate safeguards.

The DPIA process includes describing the nature and purpose of data processing, assessing any risks posed to data subjects, and documenting measures to protect personal data and ensure regulatory compliance. By systematically addressing these aspects, a DPIA helps organizations not only comply with legal requirements but also build trust and transparency with stakeholders.

To enhance the effectiveness of a DPIA, a **quantitative model** can be integrated, offering a structured and objective approach to assess the security risks tied to data processing activities, especially around confidentiality, integrity, and availability.

The GDPR mandates DPIAs for processing activities that pose high risks to individuals' rights and freedoms (Table A-16, Table A-17), and a quantitative model enables organizations to meet these regulatory requirements by providing measurable and replicable insights into these risks.

Introducing a quantitative model for risk assessment in data protection involves establishing a structured approach that calculates security risks associated with data processing activities. Unlike qualitative models that rely on descriptive categories, a quantitative model assigns numerical values to various risk factors, such as the probability and severity of loss of confidentiality (R), integrity (I) and availability (D) of the data.

$$Risk = Severity * Probability$$

This approach enables a more precise, objective, and comparable assessment of risks, which is essential for making data-driven decisions in data protection.

The evaluation is performed considering that the severity of the loss of R has the same severity in case of loss of I and D after which on the basis of the purpose of processing it is possible to modulate. In fact, the purpose operates a mediation to reduce the severity in case of loss of I and D in case of

an objective of interest of the data subject or an objective more related to the data controller [25]. Thus, the intrinsic severity is calculated as:

$$intrinsic\ severity.R\ =\ R\ *\ 1$$
$$intrinsic\ severity.I\ =\ R*\ I$$
$$intrinsic\ severity.D\ =\ R\ *\ D$$

where R is the value assigned in the category of personal data (Table A-18); I and D are the values corresponding to the purpose of the processing. After evaluating the intrinsic severity in case of loss of R, I and D, the severity is calculated as follows:

$$severity.R\ =\ intrinsic\ severity.R*amplification\ factors$$
$$severity.I\ =\ intrinsic\ severity.I\ *\ amplification\ factors$$
$$severity.D\ =\ intrinsic\ severity.D*\ amplification\ factors$$

where amplification factors are calculated as:

$$amplification\ factors\ =\ A\ *\ B\ *\ C\ *\ D\ *\ E\ *\ F\ *\ G$$

A = category of data subjects; B = age group of data subjects; C = number of data subjects; D = quantity of data subjects; E = category of data processing; F = frequencies of data processing; G = category of recipients. These factors are quantified for each specific situation as indicated in Table A-18, Table A-19, Table A-20, Table A-21, Table A-22, and Table A-23 .

The calculation of probability is different whether we are in the presence of a paper or digital process. In the case of processing on paper, the probability is calculated by means of the probability matrix (Table 0-1) given by the level of structuring of document management policies and procedures vs paper archives. The value corresponding to the result of the matrix $Probability\ (processing\ on\ paper)$ is multiplied by the probability amplification factor (probability of threat competition). Whereby:

$$Probability\ R=\ Probability\ (processing\ on\ paper)*\ \ amplification\ factors\ R$$

$$Probability\ I = \ Probability\ (processing\ on\ paper) * \ \ amplification\ factors\ I$$

$$Probability\ D = \ Probability\ (processing\ on\ paper) * \ \ amplification\ factors\ D$$

In the case of computerized processing, the probability is calculated by means of the probability matrix (Table 0-1) given by the level of structuring by the software systems vs. the technology chain $Probabilità\ (digital\ processing)$, the value corresponding to the structuring of the backup & restore system (b) and the probability amplification factors (probability of threat competition). Note: The maturity of backup and restore systems only affects the probability of occurrence of threats related to I and D.

$$Probability\ R = \ Probability\ (digital\ processing) * \ \ 1 * \ \ amplification\ factors\ R$$

$$Probability\ I$$
$$= \ Probability\ (trattamento\ «digitale») * b_I * \ \ amplification\ factors\ I$$

$$Probability\ D$$
$$= \ Probability\ (trattamento\ «digitale») * b_D * \ \ amplification\ factors\ D$$

Table 0-1. Probability matrix

	Unstructured	Semi-structured	Structured	Highly structured
Unstructured	0,9	0,8	0,7	0,5
Semi-structured	0,8	0,7	0,5	0,3
Structured	0,7	0,5	0,3	0,2
Highly structured	0,5	0,3	0,2	0,1

2.1.3 Quantitative model for assessing ethics risks in information technology

In this section, we examine two quantitative models for calculating the ethics of technology according to constitutive and circumstantial ethics [38]. This inspiration comes from the model in [39].

Specifically, constitutive ethics quantifies the level of compliance with one or more of the frameworks. For each ethical principle, we consider a set of controls.

For each control, the possible answers (along with their corresponding scores) could be, for example:

- Yes, No, or NA: if the control in question is satisfied, not satisfied, or not applicable (with associated scores of 1, 0, or NA);
- High, Medium, Low, or NA: if the control in question can be implemented to a certain degree (with associated scores of 1, 0.5, 0, or NA).

Additionally, each control can be assigned a weight based on its significance within the technology under consideration. To determine a constitutive ethics score for the technology in question, a weighted average is then computed. Therefore, the constitutive ethics score (E') is given by:

$$E' = \frac{\sum_{i=1}^{n} s_i \times w_i}{\sum_{i=1}^{n} w_i}$$

Where n is the number of controls, $0 \leq s_i \leq 1$ and $w_i \geq 0$ are the score and the weight of the i-th control, respectively.

Finally, the constitutive ethics score is defined as a value ranging from 0 to 10. To calculate the new score, we simply multiply the result by 10.

$$E = E' \times 10$$

It's important to note that the evaluation process can be conducted by either a single evaluator or multiple evaluators. In cases where multiple evaluators are involved, the ultimate score E can be determined by simply taking the average of the scores provided by each individual evaluator.

In the model, it is also important to consider the type of data processed by the technology (sensitive, personal, non-personal), as follows:

$$EI_{const} = \alpha^{(1-E')} \times E$$

where α is a number associated with the sensitivity of the processed data. This parameter can take on different values, such as 0.3 for sensitive data, 0.6 for personal data, and 0.9 for non-personal data.

Instead, circumstantial ethics quantifies the robustness of the technological infrastructure (for example, the complexity of the technological infrastructure on which the considered technology is built, if some cybersecurity controls are implemented, the probability of having cyber incidents, etc.), as well as the constitutive ethics of which it is part.

For this quantification, we created a set of controls we grouped them into three categories: Technical Robustness and Safety, Identification and Online Services.

- The controls in Technical Robustness and Safety may include - some or all - the cybersecurity controls of the frameworks in the NIST (National Institute of Standards and Technology) Cybersecurity Framework [2], or the CIS (Center for Internet Security) [40], the user's awareness, how the system is made relevant, secure, and resilient over time, etc.
- The controls in the Identification property of a system technology could deal with who can access data and how, which type of data can be seen or processed, how critical data are protected, and so on.
- The controls in the Online Services category could aim to assess how the facilitation of communication provided by online services makes the technical part of the system more difficult, complex, and vulnerable.

The evaluation of the technological infrastructure complexity is similar to that of the constitutive ethics compliance score described above. Specifically, we assign a score for each answer (e.g., 0 for Minimal, 1 for Low, 2 for Moderate, 3 for High).

A weight can be assigned to each control according to its relevance. Finally, the technological infrastructure complexity index is defined as a value ranging from 0 to 10

Considering the constitutive ethics score (E) as a variable x, we can assume that the probability of success P(x) of an adverse event follows a generalized logistic function:

$$P(x) = A + \frac{K - A}{1 + e^{-\beta (x - x_0)}}$$

where A and K are the lower and the upper horizontal asymptote, respectively, B is the growth rate, and x0 is the midpoint, i.e., the value of x corresponding to half the maximum level; we associate the complexity of technological infrastructure to x0.

Furthermore, considering that different organizations process different types of data, the final value of the probability of success of an adverse event can be weighted according to some parameter:

$$p(x) = \beta P(x)$$

with $0 < \beta \leq 1$.

Possibles values for the parameters of the logistic function and β have potential values outlined in [16]. The probability that an event has success after i – 1 failing attempts (likelihood of occurrence), can be calculated as:

$$L(i) = (1 - p)^{i-1} \times p$$

We then validate our model by applying it to recent health-related blockchain frameworks.

2.1.4 Qualitative and semi-quantitative models for assessing ethics risks in artificial intelligence

In 2019, The Ethics Guidelines for Trustworthy Artificial Intelligence are published by the High-Level Expert Group on Artificial Intelligence (AI HLEG), set up by the European Commission.

Within these guidelines, Chapter 3 includes an Assessment List for Trustworthy Artificial Intelligence (ALTAI) to evaluate if an AI system—whether in development, deployment, procurement, or use—complies with the seven essential requirements for Trustworthy AI as outlined in the guidelines (Figure A-4):

1. Human Agency and Oversight.
2. Technical Robustness and Safety (Figure A-5, Figure A-6).
3. Privacy and Data Governance (Figure A-7, Figure A-8).
4. Transparency.
5. Diversity, Non-discrimination and Fairness.
6. Societal and Environmental Well-being.
7. Accountability.

From the offline document of ALTAI, an online tool is developed**.** Figure A-9 shows a segment of the **ALTAI** online tool, which focuses on evaluating the **Technical Robustness and Safety** of AI systems through a qualitative assessment framework (the corresponding ALTAI offline version is in Figure A-5, Figure A-6);. This approach relies on descriptive categories and yes/no responses to help assess the resilience, reliability, and safety of an AI system without requiring complex numerical data or precise scoring.

In this tool, the assessment begins with straightforward yes/no questions, asking evaluators to confirm basic details about the system's security, such as whether it meets certain cybersecurity standards or is vulnerable to cyber-attacks. This binary structure is a hallmark of qualitative assessments, enabling a quick identification of potential risks based on simple, predefined criteria. If the answer is "yes" to a question about, say, exposure to cyber threats, it signals a potential area of concern that may need further attention. This format allows the evaluator to make broad judgments without needing specific measurements, focusing instead on the presence or absence of risks.

As the assessment continues, evaluators are prompted to consider the level of risk associated with each area—such as the probability of a cybersecurity breach or the severity of a safety failure—using descriptive categories like "Low," "Moderate," "Significant," or "High." These qualitative ratings help to categorize the perceived risk in a structured but flexible way. By relying on subjective assessments rather than exact data points, this approach provides a general understanding of the potential impact without requiring precise quantification, which can be especially helpful when data is limited or when the assessment focuses on complex ethical issues.

Following each risk identification, the ALTAI tool asks the evaluator to rate the adequacy of the measures that have been implemented to address these risks. The options for responses, including "Not existent", "Completely Inadequate,", "Almost Adequate", "Adequate," and "Fully Adequate," encourage the evaluator to reflect on the effectiveness of current safeguards in place. Again, this qualitative assessment relies on descriptive categories rather than numerical scores, allowing for nuanced judgment on the robustness of the AI system's protections.

One unique feature of this tool is its emphasis on considering potential negative consequences stemming from the AI system's "objective functions," or the goals it is programmed to pursue. Here, the evaluator is prompted to think about whether these objectives could unintentionally lead to harmful outcomes, such as reinforcing biases or making unethical decisions to achieve its goals. This part of the assessment does not quantify risk but instead relies on the evaluator's understanding of

possible ethical concerns, highlighting areas where the AI's design might conflict with responsible or trustworthy outcomes.

The assessment also addresses the AI system's reliability, fall-back plans, and reproducibility. Questions about whether the system has fail-safe mechanisms or backup plans in case of critical failures are answered through qualitative ratings. Evaluators are asked to consider, for example, if the system has the necessary fall-back measures to maintain functionality during unexpected disruptions, using descriptive categories to indicate the adequacy of these contingency plans. By focusing on general stability rather than precise metrics, this part of the assessment builds a broad picture of the AI system's dependability under various conditions.

After this qualitative assessment, the tool calculates a score. Figure A-10 represents a shift in the ALTAI system from a purely qualitative model to a semi-quantitative approach, allowing for a more structured and measurable assessment of AI compliance across various criteria. Previously, the ALTAI tool relied on yes/no responses and broad categories, which gave a basic, high-level understanding of an AI system's performance. However, in this semi-quantitative version, the assessment results are visualized in a radar chart, or spider graph, which provides a more nuanced, measurable depiction of how well the AI system meets each requirement.

In this semi-quantitative model, each criterion—such as Technical Robustness, Privacy, Transparency, and Accountability—is plotted on a scale, with scores ranging from 0.0 to 1.0, reflecting the level of compliance. Rather than a simple binary "met" or "not met," this approach enables evaluators to capture degrees of fulfillment. For example, a score of 0.6 might suggest partial compliance with a criterion, while a score closer to 1.0 indicates a high degree of adherence. This structured scale provides a clear visual representation of strengths and weaknesses, enabling evaluators to quickly identify areas needing improvement, like "Technical Robustness and Safety" or "Privacy," if those segments score lower than others. This visual format presents a comprehensive view of the AI system's overall trustworthiness, making it easier to interpret performance at a glance.

The semi-quantitative approach also introduces a layer of granularity that the qualitative model lacked. By scoring each criterion on a continuum, the assessment can provide more detailed feedback. Rather than just indicating that a certain area is "adequate" or "inadequate," the semi-quantitative scores allow for a finer analysis, highlighting aspects of the AI system that are partly fulfilled but still require enhancements. This nuanced feedback is invaluable, particularly for complex AI systems where certain criteria may be met to varying degrees.

Additionally, this scoring system allows for more effective benchmarking and progress tracking. Organizations can compare their performance over time or against industry standards, something not possible in a purely qualitative model, where responses are often too subjective or isolated. In this semi-quantitative model, the ability to monitor changes in each criterion's score means organizations can set specific targets, make informed decisions on areas for improvement, and demonstrate their commitment to meeting ethical and regulatory standards over time.

In essence, the semi-quantitative model in this ALTAI assessment bridges the gap between a basic qualitative approach and a fully quantitative one. By combining the clarity of a visual chart with the precision of scaled scoring, it allows for a more detailed, actionable assessment of AI compliance. This model provides a balanced method for evaluating AI systems, enhancing the trustworthiness of AI and the ability to take targeted actions based on measurable insights.

2.1.5 Usefulness of methods for assessing ethical issues in European projects

The utility of defensive strategies and assessment methods in data protection, AI trustworthiness, and information technologies is pivotal, especially within the framework of European projects. These methodologies—spanning qualitative, semi-quantitative, and quantitative approaches—are essential tools for ensuring that technological systems are robust, secure, and aligned with ethical standards.

Qualitative and quantitative methods in data protection offer a comprehensive assessment of risks associated with data management, ensuring compliance with regulations such as the GDPR. Qualitative approaches allow for flexibility and adaptability, accommodating the complexity of ethical concerns and providing nuanced insights. Meanwhile, quantitative methods lend objectivity and standardization, which are crucial for cross-border projects in the EU where comparability of results is necessary. This dual approach helps European initiatives to maintain strict standards of data security while fostering trust among stakeholders.

Similarly, ensuring the trustworthiness of AI systems, particularly in terms of security, is achieved through qualitative and semi-quantitative assessment methods. These approaches allow evaluators to address the inherently complex and evolving nature of AI technologies. Qualitative methods capture the ethical and social nuances in AI applications, offering flexibility in risk assessment, while semi-quantitative approaches introduce numerical elements that support risk prioritization without compromising adaptability. This balanced approach is highly relevant for European projects, where AI systems are expected to meet high security standards across various sectors.

In the realm of information technology, quantitative assessment methods are indispensable for measuring risks with precision and clarity, enabling stakeholders to make data-driven decisions. The objectivity provided by quantitative assessments aligns well with the rigorous requirements of EU projects, where standardized metrics ensure consistency and facilitate cooperation between member states. Moreover, quantitative approaches support the scalability and integration of IT solutions across different industries, reinforcing a pan-European technological framework that values both innovation and security.

In sum, these assessment methods—tailored to specific domains and methodologies—serve as a foundation for European projects aiming to foster safe, reliable, and ethical technological advancements. By employing a mix of qualitative, semi-quantitative, and quantitative approaches, EU initiatives can achieve a balanced and effective ethical risk management strategy that is crucial for harmonizing innovation with regulatory compliance across member states.

In Table A-24 and Table A-25, we present the assessment of ethics issues in the european project EDIH4MARCHE, which we presented together with the team in the proposal phase of the project. The project was then among the winners of the selection.

In Table A-24, we find the qualitative assessment of ethical issues related to data protection and AI, respectively, while Table A-25, shows the results of the assessment using the data protection decision tree and ALTAI tools.

2.1 Offensive strategies

On the other hand, in offensive methodologies, we will have to try to attack the system as ethical attackers to discover possible vulnerabilities in the system and improve the defense capacity. We will examine a case study conducted on Google "Perspective API".

2.1.1 Ethical biases in machine learning based filtering of internet communication

In this section, we examine the strategy of attacking a machine learning system to assess its security [41]. The study uses quantitative analysis.

Background

The Perspective API was developed by Google – Jigsaw. It evaluates communications through a series of emotional concepts called attributes and uses ML (multilingual models based on BERT and monolingual convolutional neural networks) to identify "toxic" comments (Figure 0-1).

Related Searches

Perspective API was found to be vulnerable to bias and discrimination-based attacks [42], [43]. These studies have identified biases in different categories by qualitatively and/or quantitatively analyzing attributes such as toxicity. From these studies, language (i.e., adjectives or word arrangements) was found to be a defining feature of bias in the categories of gender, race, ethnicity, and topical content [42], [43]. Vulnerabilities are reduced with subsequent Perspective updates.

Attribute name	Description
TOXICITY	A rude, disrespectful, or unreasonable comment that is likely to make people leave a discussion.
SEVERE TOXICITY	A very hateful, aggressive, disrespectful comment or otherwise very likely to make a user leave a discussion or give up on sharing their perspective. This attribute is much less sensitive to more mild forms of toxicity, such as comments that include positive uses of curse words.
IDENTITY ATTACK	Negative or hateful comments targeting someone because of their identity.
INSULT	Insulting, inflammatory, or negative comment towards a person or a group of people.
PROFANITY	Swear words, curse words, or other obscene or profane language.
THREAT	Describes an intention to inflict pain, injury, or violence against an individual or group.
TOXICITY EXPERIMENTAL	A rude, disrespectful, or unreasonable comment that is likely to make people leave a discussion.
SEVERE TOXICITY EXPERIMENTAL	A very hateful, aggressive, disrespectful comment or otherwise very likely to make a user leave a discussion or give up on sharing their perspective. This attribute is much less sensitive to more mild forms of toxicity, such as comments that include positive uses of curse words.
IDENTITY ATTACK EXPERIMENTAL	Negative or hateful comments targeting someone because of their identity.
INSULT EXPERIMENTAL	Insulting, inflammatory, or negative comment towards a person or a group of people.
PROFANITY EXPERIMENTAL	Swear words, curse words, or other obscene or profane language.
THREAT EXPERIMENTAL	Describes an intention to inflict pain, injury, or violence against an individual or group.
SEXUALLY EXPLICIT	Contains references to sexual acts, body parts, or other lewd content.
FLIRTATION	Pickup lines, complimenting appearance, subtle sexual innuendos, etc.
ATTACK ON AUTHOR	Attack on the author of an article or post.
ATTACK ON COMMENTER	Attack on fellow commenter.
INCOHERENT	Difficult to understand, nonsensical.
INFLAMMATORY	Intending to provoke or inflame.
LIKELY TO REJECT	Overall measure of the likelihood for the comment to be rejected according to the NYT's moderation.
OBSCENE	Obscene or vulgar language such as cursing.
SPAM	Irrelevant and unsolicited commercial content.
UNSUBSTANTIAL	Trivial or short comments.

Figure 0-1.Attributes and definition[44].

Objective

Our goal is to identify semantic biases in the Perspective API communication filtering algorithm by analyzing the Twitter communications of different institutional users. This study analyses all attributes predicted by Perspective API.

Data collection

The dataset consists of 400 tweets. The dataset contains 100 tweets for each user: the World Health Organization (WHO), the World Trade Organization (WTO), the International Monetary Fund (IMF), and the North Atlantic Treaty Organization (NATO). Each tweet has an assigned field: Tweets from the WHO, WTO, IMF, and NATO have been assigned the fields of Health, Trade, Finance, and Defense, respectively. Each tweet was provided as input to the Perspective API, which produced a score for each attribute (Figure 0-2, Figure 0-3, Figure 0-4, Figure 0-5).

Statistical analysis tools

We use different statistical tools to conduct the analysis of tweets, as follows:

1. A priori Power Analysis to estimate the smallest sample size needed for the experiment.
2. Cumulative distribution function (CDF) to describe the probability distribution of each attribute.
3. Statistical tests (t-test or Kruskal-Wallis test for comparisons of means/medians calculated from the Perspective API scores for each attribute).

The R software environment is the open-source software environment used for the analysis.

Results and Discussion

Figure 0-6 shows that Perspective evaluates tweets as false positive (score > 0). Discrimination can have a greater impact on INCOHERENT, LIKELY TO REJECT, and SPAM (probability greater than 0.7, which is the threshold for filtering online communication). User communications in these fields can be filtered more often than in others on these attributes. Figure 0-7 shows that defense is the most discriminated category in INCOHERENT and LIKELY TO REJECT (INCOHERENTDefense = 0.94, LIKELY TO REJECTDefense = 0.79), whereas commerce is the most discriminated category in SPAM (SPAMTrade = 0.96). Figure 0-8 and Figure 0-9 show the differences between fields in LIKELY TO REJECT AND SPAM, while Figure 0-10 shows the statistical differences between fields (p-values <0.05 are considered statistically significant differences).

The results suggest that discrimination is derived from the semantics of the user (secondary characteristics). From the ethical point of view, the fairness component "Fair equality of opportunity" is not respected [45].

TWEETS	TOXICITY	SEVERE_ TOXICITY	IDENTITY_ ATTACK	INSULT	PROFANITY	SEXUALLY_ EXPLICIT	THREAT	FLIRTATION	ATTACK_ ON_AUTHOR	ATTACK_ ON_ COMMENTER	INCOHERENT	INFLAMMATORY	LIKELY_ TO_REJECT	OBSCENE	SPAM	UNSUBSTANTIAL
🎉Happy 25th WTO Anniversary to Panama 🇵🇦! #Panama became the 132nd WTO Member on 6 September 1997. More info 👉 https://t.co/lQSw0MfNpH https://t.co/imU98mtwkr	0,01	0,05	0,05	0,04	0,05	0,06	0,11	0,50	0,00	0,00	0,83	0,00	0,90	0,01	1,00	0,06
"Ambassador Samba was a champion within the WTO community for trade-led development. He will be missed at the WTO, but his legacy and contributions will be remembered, as will his eloquence", said DG @NOIweala as she expressed her condolences, shared by the WTO staff.	0,01	0,02	0,04	0,08	0,02	0,02	0,03	0,28	0,00	0,00	0,20	0,06	0,10	0,01	0,07	0,21
Trade Policy Analyst at the Ministry of Foreign Affairs, Mauritius 🇲🇺, Shazia Bi Kurmoo, explains how the simulations taught during the course will be most beneficial to delivering on tasks related to Mauritius' ongoing negotiations. Read below 👇 https://t.co/LC58AuDqCe	0,00	0,03	0,10	0,06	0,03	0,03	0,07	0,43	0,02	0,05	0,97	0,08	0,79	0,09	0,96	0,40
🏊 Join us for a virtual roundtable discussion on "Digital Trade and Digital Transformation in Asia", as panellists explore digital transformation challenges & opportunities from an Asian perspective. 🗓 August 24 🕒 10:00 - 12:00 CEST Register 👉 https://t.co/bO8iWvHoll https://t.co/MEWC2Sw3wY	0,00	0,05	0,14	0,05	0,09	0,07	0,12	0,46	0,00	0,00	0,91	0,00	1,00	0,00	1,00	0,05

Figure 0-2. Tweets and scores for WTO

TWEETS	TOXICITY	SEVERE_ TOXICITY	IDENTITY_ ATTACK	INSULT	PROFANITY	SEXUALLY_ EXPLICIT	THREAT	FLIRTATION	ATTACK_ ON_AUTHOR	ATTACK_ ON_ COMMENTER	INCOHERENT	INFLAMMATORY	LIKELY_ TO_REJECT	OBSCENE	SPAM	UNSUBSTANTIAL
Set a reminder for our @TwitterSpaces chat about the impacts of floods on health situation in #Pakistan 🌊 🗓 5 September 🕒 13:00 CEST 🎙 @WHOPakistan's Dr Palitha Mahipala and @DrMariaNeira https://t.co/ARLqbI2RGl	0,00	0,05	0,11	0,04	0,09	0,06	0,19	0,46	0,00	0,00	0,98	0,03	0,97	0,04	0,94	0,21
#SexualHealth doesn't get old! There is no upper age limit for good sex & it can help keep you healthy into your later years. More info 👉 https://t.co/7oG8HOyz22 https://t.co/t4HVOZZiqb	0,23	0,26	0,19	0,14	0,31	0,88	0,18	0,54	0,00	0,05	0,66	0,05	0,92	0,16	1,00	0,17
@DrTedros "But even in high-income countries, 30% of #healthworkers and 20% of older people remain unvaccinated. These vaccination gaps pose a risk to all of us. So, please get vaccinated if you are not, and get a booster if it's recommend that you have one"-@DrTedros #COVID19	0,04	0,04	0,06	0,08	0,03	0,04	0,13	0,47	0,06	0,73	0,44	0,28	0,45	0,40	0,24	0,34
@DrTedros "Now to #COVID19, where we are now seeing a welcome decline in reported deaths globally. However, with colder weather approaching in the northern hemisphere, it's reasonable to expect an increase in hospitalizations and deaths in the coming months"-@DrTedros	0,05	0,03	0,04	0,03	0,03	0,02	0,19	0,32	0,02	0,10	0,55	0,14	0,23	0,10	0,31	0,39

Figure 0-3. Tweets and scores for WHO

TWEETS	TOXICITY	SEVERE_ TOXICITY	IDENTITY_ ATTACK	INSULT	PROFANITY	SEXUALLY_ EXPLICIT	THREAT	FLIRTATION	ATTACK_ ON_AUTHOR	ATTACK_ ON_ COMMENTER	INCOHERENT	INFLAMMATORY	LIKELY_ TO_REJECT	OBSCENE	SPAM	UNSUBSTANTIAL
The AREAER Online database provides the IMF's exclusive survey of countries' trade and exchange regimes. With over 18 years of archival content and a query tool, you can compare specific policies across countries and years. https://t.co/55hZDya8FF #IMFPublications https://t.co/iTyJMM1laB	0,01	0,02	0,03	0,04	0,03	0,03	0,03	0,32	0,00	0,00	0,47	0,00	0,98	0,01	1,00	0,06
RT @IMFAfrica: The IMF is working with Zambia to restore macroeconomic stability and foster higher, resilient, and more inclusive growth. I...	0,01	0,07	0,10	0,11	0,12	0,07	0,08	0,36	0,01	0,03	0,37	0,02	0,25	0,00	0,09	0,21
Singapore's recovery from the pandemic is outperforming similar economies, with economic output topping pre-crisis levels last year, but progress has been uneven. Our Country Focus article has some recommendations. https://t.co/omtutXnBg1 https://t.co/WL03LOszHe	0,01	0,03	0,10	0,04	0,03	0,05	0,07	0,37	0,03	0,00	0,25	0,03	0,20	0,01	0,45	0,15
Singapore's recovery from the pandemic is outperforming similar economies, with economic output topping pre-crisis levels last year, but progress has been uneven. Our Country Focus article has some recommendations. https://t.co/omtutXo95z https://t.co/4ckwUlqiHG	0,01	0,02	0,05	0,04	0,04	0,07	0,05	0,37	0,03	0,00	0,25	0,03	0,20	0,01	0,45	0,15
1 in 5 people in Armenia 🇦🇲 works in farming. But changing weather is forcing Armenian farmers to adapt to climate change. Here's how: https://t.co/uk6F5EpwZn https://t.co/SIlGIFY6h5	0,02	0,06	0,24	0,11	0,10	0,07	0,11	0,33	0,00	0,02	0,97	0,09	0,72	0,05	0,94	0,50

Figure 0-4. Tweets and scores for IMF

TWEETS	TOXICITY	SEVERE_ TOXICITY	IDENTITY_ ATTACK	INSULT	PROFANITY	SEXUALLY_ EXPLICIT	THREAT	FLIRTATION	ATTACK_ ON_AUTHOR	ATTACK_ ON_ COMMENTER	INCOHERENT	INFLAMMATORY	LIKELY_ TO_REJECT	OBSCENE	SPAM	UNSUBSTANTIAL
RT @NATOBrazeB: What does it take to #ProtectTheFuture? We're asking young artists from across the Alliance to answer that question.Winning...	0,02	0,08	0,09	0,11	0,15	0,09	0,14	0,32	0,01	0,09	0,58	0,06	0,24	0,16	0,17	0,40
RT @SpainNATO: #GoodNight @NATO allies! #WeAreNATO 🇪🇸 📸 @hesyja https://t.co/tqYnhlLLHT	0,01	0,11	0,19	0,15	0,13	0,10	0,17	0,41	0,00	0,00	0,97	0,03	0,99	0,11	0,78	0,35
Soldiers from Slovenia 🇸🇮, Montenegro 🇲🇪, North Macedonia 🇲🇰 and Italy 🇮🇹 take part in a regular #CBRN training exercise involving decontamination of vehicles, equipment and personnel 📸 https://t.co/cuczmYJ9XA	0,01	0,02	0,04	0,02	0,02	0,03	0,05	0,40	0,00	0,00	0,89	0,01	0,97	0,02	1,00	0,11
🇺🇸 U.S. Air Force B-52 integrates with 🇩🇰 Danish and 🇸🇪 Swedish Air Force during a pre-planned Bomber Task Force mission #WeAreNATO Bomber Task Force operations are conducted with #NATO allies and partners to demonstrate and strengthen shared commitment to global security https://t.co/sYY1co2jWk	0,01	0,05	0,14	0,06	0,07	0,06	0,12	0,43	0,00	0,00	1,00	0,02	0,92	0,00	0,84	0,29
RT @dylanpwhite: ICYMI: Last week NATO Secretary General Jens Stoltenberg visited a North Warning System radar site in Canada's #Arctic—vit...	0,01	0,02	0,03	0,03	0,06	0,06	0,09	0,43	0,00	0,00	0,56	0,01	0,43	0,00	0,17	0,53

Figure 0-5. Tweets and scores for NATO

Attribute	Median	IQR	Min	Max
TOXICITY	0.01	0.02	$0.23 \cdot 10^{-2}$	0.37
SEVERE TOXICITY	0.05	0.08	$0.26 \cdot 10^{-2}$	0.59
IDENTITY ATTACK	0.09	0.11	$0.63 \cdot 10^{-2}$	0.57
INSULT	0.08	0.09	$0.55 \cdot 10^{-2}$	0.54
PROFANITY	0.09	0.13	$0.35 \cdot 10^{-2}$	0.62
THREAT	0.12	0.09	$0.93 \cdot 10^{-2}$	0.75
TOXICITY EXPERIMENTAL	0.01	0.12	$0.23 \cdot 10^{-2}$	0.37
SEVERE TOXICITY EXPERIMENTAL	$0.13 \cdot 10^{-2}$	0.11	$0.03 \cdot 10^{-2}$	0.21
IDENTITY ATTACK EXPERIMENTAL	0.02	0.02	0.01	0.21
INSULT EXPERIMENTAL	0.01	0.12	$0.14 \cdot 10^{-2}$	0.21
PROFANITY EXPERIMENTAL	$0.77 \cdot 10^{-2}$	0.11	$0.23 \cdot 10^{-2}$	0.45
THREAT EXPERIMENTAL	0.01	0.13	$0.98 \cdot 10^{-2}$	0.13
SEXUALLY EXPLICIT	0.07	0.28	$0.57 \cdot 10^{-2}$	0.92
FLIRTATION	0.37	0.25	0.14	0.92
ATTACK ON AUTHOR	$0.79 \cdot 10^{-2}$	0.27	$0.16 \cdot 10^{-4}$	0.53
ATTACK ON COMMENTER	0.03	0.27	$0.56 \cdot 10^{-4}$	0.91
INCOHERENT	0.89	0.02	0.20	1
INFLAMMATORY	0.10	0.00	$0.09 \cdot 10^{-2}$	0.63
LIKELY TO REJECT	0.76	0.01	0.10	0.99
OBSCENE	0.10	0.01	$0.03 \cdot 10^{-2}$	0.94
SPAM	0.93	0.01	0.07	1
UNSUBSTANTIAL	0.48	0.00	0.04	0.96

Figure 0-6. Descriptive statistics for each attribute

Field	INCOHERENT	LIKELY TO REJECT	SPAM
Health	0.89 [0.12]	0.76 [0.28]	0.94 [0.29]
Trade	0.89 [0.10]	0.78 [0.22]	0.96 [0.09]
Finance	0.84 [0.13]	0.71 [0.28]	0.94 [0.19]
Defense	0.94 [0.08]	0.79 [0.26]	0.75 [0.33]

Figure 0-7. Descriptive statistics for each attribute in relation to the specific field

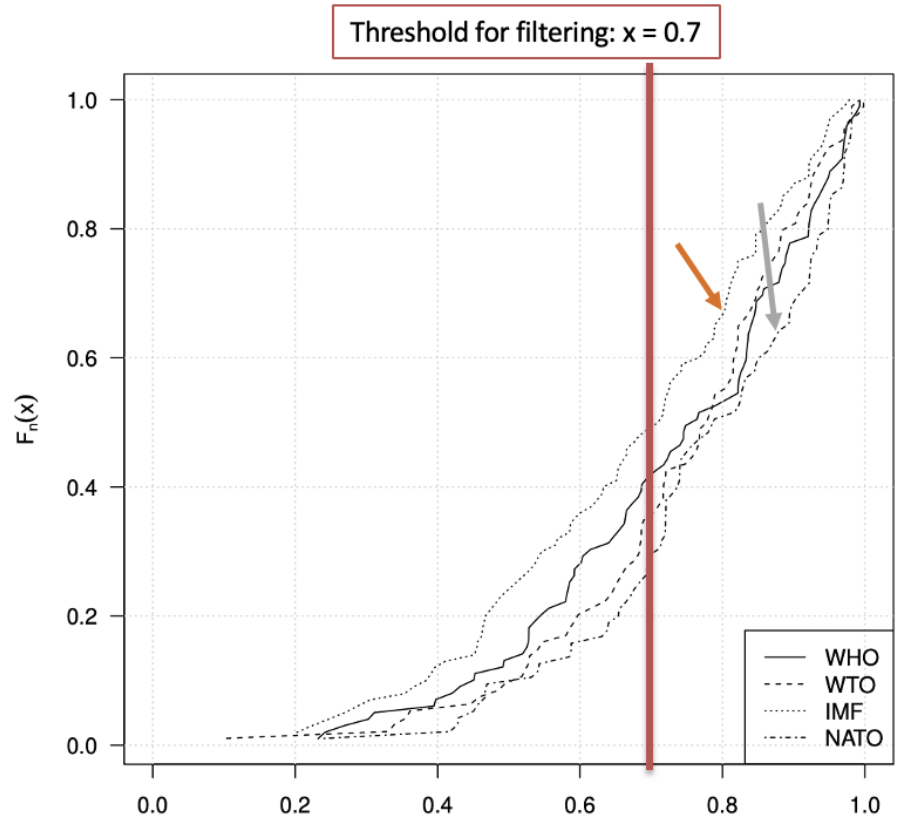

Figure 0-8. Cumulative distribution function of LIKELY TO REJECT. WHO (solid line), WTO (dotted line), IMF (dotted line), NATO (dotted line). Statistical difference between IMF and NATO (p-value<0.05).

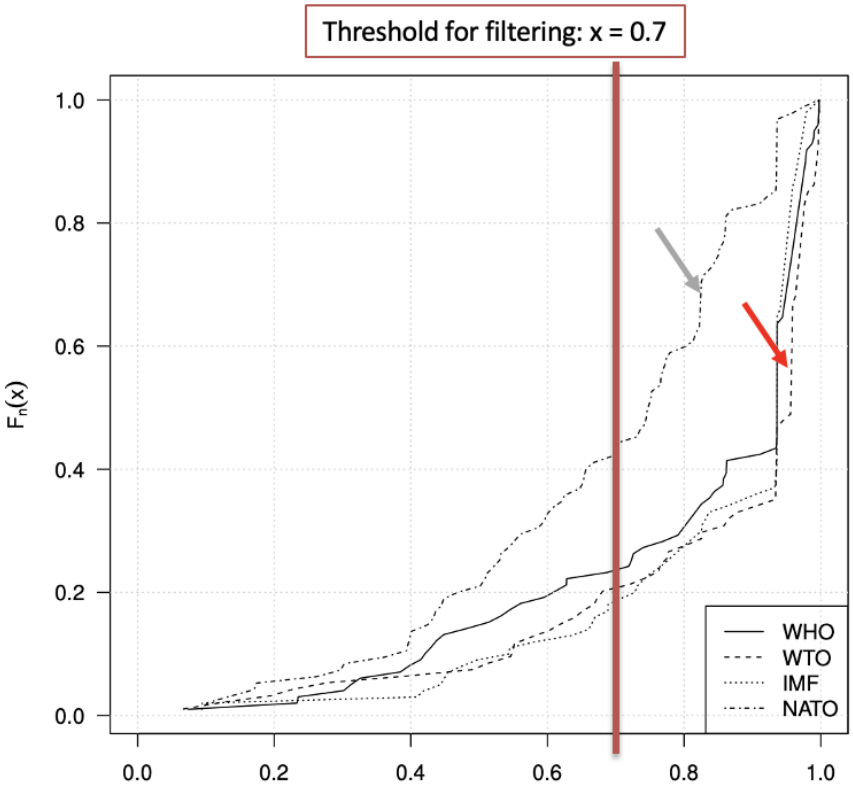

Figure 0-9. Cumulative SPAM distribution function. WHO (solid line), WTO (dotted line), IMF (dotted line), NATO (dotted line). Statistical difference between WTO and NATO (p-value<0.05).

Attribute	Health-Trade	Health-Finance	Health-Defense	Trade-Finance	Trade-Defense	Finance-Defense
TOXICITY	< 0.001	0.012		< 0.001	< 0.001	< 0.001
SEVERE TOXICITY	0.009	< 0.001	0.02	< 0.001	< 0.001	< 0.001
IDENTITY ATTACK	0.007	< 0.001		0.003	< 0.001	< 0.001
INSULT	0.009	< 0.001		0.001	< 0.001	< 0.001
PROFANITY		< 0.001	0.001	< 0.001	< 0.001	< 0.001
SEXUALLY EXPLICIT		< 0.001	0.007	< 0.001	< 0.001	< 0.001
THREAT	< 0.001	< 0.001	0.02	< 0.001	< 0.001	< 0.001
FLIRTATION	0.01		< 0.001	< 0.001	< 0.001	< 0.001 (*)
ATTACK ON AUTHOR		< 0.001		< 0.001		< 0.001
ATTACK ON COMMENTER	0.04	0.01				0.03
INCOHERENT		0.02	< 0.001	0.009	< 0.001	< 0.001
INFLAMMATORY	0.001			0.04	0.007	
LIKELY TO REJECT		0.03		0.01		< 0.001
OBSCENE	0.005			0.009	0.007	
SPAM			< 0.001	0.04	< 0.001	< 0.001
UNSUBSTANTIAL					< 0.001	
TOXICITY EXPERIMENTAL	< 0.001	0.01		< 0.001	< 0.001	< 0.001
SEVERE TOXICITY EXPERIMENTAL	< 0.001	0.006	0.002	0.001	< 0.001	< 0.001
IDENTITY ATTACK EXPERIMENTAL	< 0.001	0.001			< 0.001	< 0.001
INSULT EXPERIMENTAL	< 0.001	< 0.001		0.008	< 0.001	< 0.001
PROFANITY EXPERIMENTAL	< 0.001	< 0.001			< 0.001	< 0.001
THREAT EXPERIMENTAL	< 0.001	< 0.001	< 0.001	0.005	< 0.001	< 0.001

Figure 0-10. p-VALUES OF UNPAIRED TWO-SAMPLES WILCOXON TEST AND T-TEST (*) ACROSS HEALTHCARE, TRADE, FINANCE, AND DEFENSE (p-VALUES ≤ 0.05 ARE CONSIDERED STATISTICALLY SIGNIFICANT DIFFFERENCES).

In conclusion, this study attacks the Perspective API's machine learning-based communication filtering algorithm to identify and analyze bias and discrimination as vulnerabilities. This study confers bias and discrimination to the semantics of communication with users. Therefore, this study can be considered as human-in-the-loop feedback for the continuous design of a trustworthy model. In conclusion, this research highlights the need for an ethically sound automated moderation algorithm according to the Ethics by Design approach.

Chapter 3: Mitigation strategies

This chapter explores the strategies for mitigating ethical risks. In Section 3.1, we start by discussing blockchain technology in the context of cybersecurity ethics, specifically focusing on its use to counter threats such as DDOS and ransomware attacks. Section 3.2 addresses business intelligence as a means of mitigating the risks associated with the loss of confidentiality, integrity, and availability of data. We also consider a healthcare application related to this field of data protection ethics. The section 3.3 concludes with an analysis of various techniques for mitigating risks to the fairness principle in adversarial machine learning.

3.1 Blockchain for the ethics of cybersecurity

Blockchain can be defined as a decentralized and distributed digital ledger (DLT), capable of ensuring the security and immutability of transactions and the data contained in it thanks to the use of cryptography. Blockchain technology, as a form of DLT, has significant implications for the ethical aspects of cybersecurity. The inherent features of blockchain, such as decentralization, encryption, consensus, and immutability, contribute to mitigate ethical risks and enhance ethical principles. In fact, the decentralized and distributed nature of blockchain can contribute to enhance the availability, while the encryption can favor confidentiality of data. Decentralization, distribution, encryption, and immutability features can enhance the integrity of data. Instead, the consensus mechanisms and cryptographic algorithms ensure accountability of data [46]. However, the immutability of blockchain raises ethical concerns regarding the right to be forgotten, as it conflicts with privacy regulations like the General Data Protection Regulation (GDPR) that require the ability to delete personal data upon request [47]. However, the Data Block Matrix (DBM) is a variant of distributed ledger technology that addresses the conflict between blockchain's immutability and privacy requirements [47]. Unlike traditional blockchain systems, DBM allows for controlled revision or deletion of data while maintaining the integrity-preserving properties of distributed ledgers. The key feature of DBM is its ability to support privacy rules by enabling the deletion of a user's private data upon request, which is essential for compliance with regulations like GDPR [47], [48]. This capability expands the range of blockchain applications, particularly in sectors where data privacy is crucial, such as healthcare and logistics management.

While blockchain can enhance transparency and accountability (integrity and traceability), it also necessitates a balance with privacy, as the visibility of transactions could potentially expose sensitive information [49], [50]. Additionally, the anonymity provided by blockchain can complicate the ethical management of user identity and data privacy [51].

In the following section, we consider specific case studies of DDos and ransomware attacks providing some useful mitigation strategies of these attacks with the use of blockchain.

3.1.1 DDos attacks prevention

Blockchain technology has been identified as a promising approach to mitigate DDoS attacks [52], [53]. The decentralized nature of blockchain provides a framework for distributed defense mechanisms, which aligns well with the distributed nature of DDoS attacks [54]. Moreover, blockchain's inherent features such as immutability and transparency can enhance the security and integrity of mitigation strategies [55]. However, the application of blockchain in DDoS mitigation is still in the early stages of development. While the potential is recognized, the solutions are considered to be in their infancy, and further research is needed to fully realize and deploy effective blockchain-based DDoS mitigation techniques [53], [54].

In the following section, we examine the positive and negative impacts of blockchain technology on denial-of-service attack. Specifically, to determine the advantageous impacts of blockchain, we use the ethical principles outlined in the Formosa et al. framework for cybersecurity (Beneficence, Non Maleficence, Justice, Explicability, And Autonomy) [4].

The positive impacts of blockchain on denial of service are as:

- Blockchain’s decentralized structure aligns with DDoS attacks, which also operate through distributed networks of compromised devices. This alignment is advantageous as a decentralized framework allows a broader scope of defense mechanisms, contrasting with traditional centralized approaches that are often a target in DDoS scenarios [54]. From an ethical perspective, this approach supports **Beneficence** (acting for the good of affected users and systems) and **Justice** (ensuring fair defense mechanisms by decentralizing protective measures and reducing dependency on a single point of failure).
- Blockchain’s transparency and immutability allow the storage of secure, tamper-resistant data, which is crucial in distinguishing between legitimate traffic and malicious traffic in

DDoS mitigation (the major challenge in DDos attack). This feature ensures that any record of data access or modification is maintained, enhancing accountability [53]. In ethical terms, this also aligns with **Beneficence** (improving the security posture of systems) and **Justice** (providing all participants in the network with a verifiable history of data integrity).

- Integrating blockchain with IP reputation systems is proposed to improve accuracy in identifying and mitigating malicious sources. This integration could enhance decision-making in cybersecurity defenses, strengthening the precision of responses to attacks [54]. This impact supports **Beneficence** by potentially reducing collateral damage in legitimate traffic and **Justice** by promoting fair, data-driven DDoS defense mechanisms.

The negative impacts of blockchain on denial of service are as:

- As DDoS attacks grow in volume and sophistication, blockchain systems must evolve to match these changes. Blockchain's rigid structure can hinder adaptability in responding to fast-evolving attack patterns [53], [56]. This challenge raises **Explicability** concerns, as stakeholders may not fully understand or trust a blockchain system that cannot adapt quickly, and **Non-Maleficence** (avoiding harm), as outdated or rigid defenses may inadvertently allow certain types of attacks to succeed.
- Blockchain solutions necessitate collaboration among various stakeholders who may need to share sensitive data, such as IP addresses. Transparency in shared data must be balanced with security, as inaccuracies or misuse of reputation data could lead to unjust blocking of legitimate traffic [54]. This challenge is ethically complex, touching on **Explicability** (making systems transparent and understandable to stakeholders) and **Non-Maleficence** (avoiding harm that inaccurate data could inflict on legitimate users).
- Many blockchain-based DDoS solutions remain theoretical, focusing more on conceptual frameworks than on practical, implementable strategies [57][54] . This issue limits blockchain's actual effectiveness in real-world mitigation and impacts **Beneficence**, as underdeveloped solutions cannot reliably protect users.

3.1.2 Ransomware attack prevention

Blockchain technology has been identified as a promising solution to mitigate ransomware attacks, which are a significant threat to cybersecurity across various sectors, including healthcare and IoT networks [58], [59], [60], [61].

Blockchain is generally considered secure and tamper-proof, but it is not immune to all forms of cyber-attacks (e.g., selfish-mining and block withholding) [62].

Despite these concerns, the decentralization and encryption inherent in blockchain provide a robust framework for preventing data breaches and ensuring data integrity [63].

The use of blockchain has been proposed in various frameworks and systems to enhance security measures against ransomware, with some focusing on specific sectors like smart healthcare and IoT [58], [59], [60], [64], [65].

The positive impacts of blockchain on ransomware are as:

- Blockchain's decentralized architecture and strong cryptographic protocols offer enhanced security and resilience. By removing a central point of failure, blockchain can reduce the vulnerability of systems to ransomware attacks that target centralized databases. Additionally, smart contracts (self-executing contracts with terms coded directly into the blockchain) and consensus mechanisms (processes by which blockchain participants agree on transaction validity) provide secure frameworks that could proactively prevent unauthorized access and malicious behavior [66]. This aligns with the ethical principles of **beneficence**—contributing positively to cybersecurity by reducing potential points of exploitation—and **justice**, as these protections can equitably benefit all participants in the network without favoring one group over another.
- Blockchain's transparency allows all network participants to see and verify transactions in a publicly accessible ledger. While this transparency does not prevent ransomware attacks directly, it could enhance tracking efforts, making it harder for criminals to hide illicit transactions. For instance, authorities could analyze patterns and identify potentially suspicious transactions more effectively. This aligns with **explicability**, as transparency helps make the flow of funds within the network clear and understandable, allowing stakeholders to verify and track transactions that might otherwise go unnoticed.

The negative impacts of blockchain on ransomware are as:

- Blockchain transactions are immutable (unchangeable) and anonymous, which can help cybercriminals move illicit funds with reduced traceability. This anonymity is appealing to ransomware operators, who demand payments in cryptocurrencies, knowing the funds are harder to track and can be moved across borders with minimal oversight [67], [68].. This presents a conflict with **non-maleficence**—the principle of avoiding harm—as blockchain's structure, while beneficial in many legitimate contexts, also enables malicious activities like ransomware attacks.
- While blockchain offers promising tools for enhancing security, it is not immune to cyber threats. Attacks on blockchain networks, including 51% attacks (where an attacker controls more than half of the mining power, enabling them to alter the blockchain) and smart contract vulnerabilities, illustrate the need for robust cybersecurity measures. Developing such measures is complex, as they must leverage blockchain's strengths (decentralization and transparency) while also addressing its vulnerabilities [69], [70]. This concern ties into **non-maleficence** by highlighting the need to mitigate potential harm that may result from over-reliance on blockchain in security-sensitive contexts without adequate safeguards.
- Blockchain's immutability can also raise privacy concerns and conflicts with legal requirements, such as the GDPR, which grants individuals the right to request the deletion of their data [67], [71].. The immutable nature of blockchain data makes such deletion impossible, which may create challenges in balancing the **justice** principle (protecting rights fairly) with **autonomy** (respecting individuals' control over their data). Innovations like DBM approaches are being explored to address these concerns by creating hybrid systems that attempt to combine blockchain's benefits with the flexibility needed to respect privacy laws. The DBM is a variant of distributed ledger technology that addresses the privacy challenges associated with traditional blockchain systems while maintaining their integrity-preserving properties [47]. Unlike conventional blockchain, which is immutable and makes it impossible to delete data, DBM allows for controlled revision or deletion of data, making it compliant with privacy regulations that require the ability to remove personally identifiable information [47]. Interestingly, this innovation solves the blockchain privacy conflict, expanding the range of potential blockchain applications by allowing exception management [47]. The DBM has been implemented as a configurable option for Hyperledger

Fabric, with a proof-of-concept application for data sharing in a healthcare environment. Other potential applications include logistics management and digital currency [47]. Therefore, the Data Block Matrix represents a significant advancement in blockchain technology, addressing one of its major limitations while preserving its core benefits.

In conclusion, while blockchain's decentralized architecture, cryptographic security, and transparency offer promising tools for mitigating DDoS and ransomware attacks, these applications are still in early development. Blockchain's capacity to distribute data and create tamper-resistant records could help reduce single points of failure and support secure environments. However, challenges remain, such as ensuring scalability, managing privacy concerns, and addressing potential misuse by malicious actors. Continued research and innovation are crucial to fully realize blockchain's potential as a robust solution in cybersecurity, balancing its strengths with safeguards that mitigate its inherent risks.

3.2 Business intelligence for the ethics of data protection

Business intelligence (BI) involves the process of collecting, analyzing, and interpreting large volumes of data to provide valuable insights that support strategic decision-making within organizations [72]. This section explores how BI serves as a powerful tool for mitigating data protection risks, with a focus on safeguarding individual rights and freedoms.

BI plays a critical role in analyzing and monitoring risks, enabling organizations to make informed decisions that prioritize data protection. A core function of BI is risk monitoring, as it helps identify threats to data R, I, and D, along with ensuring secure processing activities related to data subjects. Here, *processing* refers to any operation, automated or otherwise, applied to personal data or sets of personal data [73].

Through BI, organizations can implement continuous monitoring systems that not only detect vulnerabilities but also anticipate potential data security risks. This approach enhances accountability by reinforcing measures for those handling personal data. BI's dynamic nature enables real-time insights into data usage patterns, highlighting anomalies or potential risk indicators. By catching these early warning signs, organizations can proactively prevent data breaches or misuse, thereby protecting individuals' rights.

In this context, BI goes beyond data analysis, acting as a safeguard to ensure that data management practices meet ethical standards and comply with data protection laws. It promotes a governance

framework rooted in transparency and accountability, thereby respecting and protecting the rights of data subjects.

3.2.1. BI in action: a data protection case study in healthcare

BI has become essential for organizations due to the exponential growth of data (big data), which requires robust tools to analyze and transform this data into actionable insights. BI enables companies to process and analyze massive datasets, extracting valuable information that supports effective, timely decision-making. This capability is crucial for *knowledge workers*, professionals responsible for formulating strategies and making decisions that impact the organization in the medium and long term. Therefore, BI is a support to knowledge workers, who are able to make timely and efficient decisions based on analytical methodologies and mathematical models that are able to transform data into information and information into knowledge (Table 0-1).

The architecture of a BI system typically involves several stages:

1. **Data Sources**: unifying and integrating data from various origins (e.g., internal databases, external sources) to create a consolidated base for analysis.
2. **Data Warehouses and Data Marts**: storing and organizing data in multidimensional structures, enabling efficient, in-depth analysis.
3. **Data Exploration**: conducting passive analysis, such as querying, reporting, and applying statistical methods to generate descriptive insights.
4. **Data Mining**: engaging in active analysis, employing algorithms to discover hidden patterns, correlations, or trends within the data.
5. **Decision-Making**: using the insights generated to make informed choices about future actions.

Each stage plays a distinct role in converting data into insights that support strategic and operational decisions.

Our case study of BI applied to data protection in a healthcare facility highlights BI's role as a powerful tool for a Data Protection Officer (DPO) tasked with mitigating risks associated with data processing. In this context, BI tools, such as Microsoft Power BI, help integrate and explore data sources, enhancing the DPO's ability to monitor and assess risks linked to data R, I, D.

Table 0-1. Definition of data, information and knowledge.

	Definition
Data	Structured encoding of single primary entities or transactions involving two or more primary entities
Information	Outcome of data extraction and processing activities
Knowledge	Information transformed into knowledge and used to make decisions and develop corresponding actions

3.1.2 Power BI for implementing Business Intelligence in data protection

We have collected data relating to the processing register of the University Hospital of the Marche region concerning: the registers, the evaluation of the registers, the structure of the processing, the risk of loss of R, I, and D of the data (the risk is calculated using the quantitative method in Chapter 2 by using SQL in the JOBS DP software), the summary assessment of the need for DPIA, the list of services related to the GDPR register, the threat probability assessment, the risk mitigation measures, and detailed calculations (Figure A-11, Figure A-12, Figure A-13, Figure A-14, Figure A-15, Figure A-16, Figure A-17, Figure A-18, Figure A-19, Figure A-20, Figure A-21, Figure A-22)

The work carried out for business intelligence has seen the data analysis phase with data pre-processing and processing. As a result of the data analysis, the data will be visualized through the creation of reports and dashboards.

In particular, the work is divided into:

1) Writing R code for data preprocessing (Appendix B)
2) Writing DAX code for the risk calculation algorithm (Appendix C)
3) Build dashboard prototypes to extract actionable insights from data
4) Satisfying the requirement of automaticity on the business intelligence prototype (Table 0-2)

Table 0-2. Requirements, feasibility, and tools for using Power BI as a risk mitigation tool

Requirement	**Feasibility**	**Tool**
Automaticity of each process so that the user can • See the report, dataset, and updated dashboard • See how the risk index changes over time • Alerts on changes in the situation or when the risk index is above a certain value	Yes	Power BI Service. For the alert, you should switch to Power BI Premium mode.
Formula for calculating risk with Dax	Yes	Power BI Desktop

How the Business Intelligence Tool Works

The following description provides an overview of how Power BI Desktop and Power BI Service are utilized to create a business intelligence tool that meets specific data analysis and risk calculation requirements (Figure 0-1).

Data Preprocessing (Using R)

1. Data Preparation: Data is preprocessed using R code, which includes:
    a. Loading data from the company server: GDP_REGISTRI, GDP_REGISTRI_VALUTAZ, GD_STRUTTURA_TRATT_RISCHIO, GDP_VAL_SINT_DPIA, GDP_REGISTRI_SERVIZI, GDP_REGISTRI_MINACCE, GDP_REGISTRI_MISURE_MIT, and GDP_REGISTRI_CALCOLI (Figure A-11, Figure A-12, Figure A-13, Figure A-14, Figure A-15, Figure A-16, Figure A-17, Figure A-18, Figure A-19, Figure A-20, Figure A-21, Figure A-22)
        - Removing unnecessary columns for streamlined analysis.
        - Merging tables by joining them on the ID_REGISTRO field.
        - Eliminating duplicate columns to reduce redundancy.
        - Handling missing values by removing NA entries from tables.
        - Adding missing ID_REGISTRO values in relevant columns to align data for analysis.
        - Starting the risk calculation algorithm.

Power BI Desktop Workflow

1. Loading R Code: The R code is loaded and executed within Power BI Desktop, generating a ready-to-process table. The code can be refreshed as needed via the refresh button in Power BI.
2. Risk Calculation: Risk is calculated using DAX, where severity, probability, and overall risk are computed based on the formula. Factors in the GDP_REGISTRI_CALCOLI table are utilized, and risk values (R, I, D) are compared with those calculated using SQL in the JOBS DP software. Intermediate calculation values can be adjusted according to the DPO's preferences.
3. Data Visualization: Calculated risk values and other relevant data are visualized using comparison charts, filters, and additional reporting and dashboard tools in Power BI.
4. Saving and Publishing: The .pbix file is saved and published to Power BI Service.

Power BI Service Workflow

1. Viewing: The .pbix file containing the processed data, reports, and dashboard is available for viewing.
2. Scheduling Updates: Automatic updates are set to run at a preferred frequency.
3. Data Modification for Risk Recalculation: Data from the most recent update can be exported to Excel, where calculation indexes can be modified as needed. The dataset is then reloaded into Power BI Desktop, with updated risk values published back to Power BI Service.
4. Risk Monitoring: For historical comparison and ongoing risk monitoring, datasets with past refresh dates can be exported and uploaded to Power BI Desktop.
5. App Creation: An app is created to share the tool with multiple users, making the data and dashboards accessible to a wider team (Figure 0-1)

Automation

The Power BI Service enables a high level of automation in data processing and updating:

- Data Storage: Source data from the company server is stored in a designated OneDrive folder (e.g., "C:/Users/……xlsx"), while the data model, reports, and dashboard are saved in a .pbix file on OneDrive.
- Process Flow:

- Source data stored in OneDrive is preprocessed using R within Power BI Desktop. This preprocessing step creates the necessary data model for report and dashboard building.
- The .pbix file, containing the data source, model, reports, and dashboard, is saved to OneDrive and published to Power BI Service.
- Updates to both data and dashboards (scheduled or on-demand) are conducted via Power BI Service, using a gateway connection to the data source for seamless synchronization.

Figure 0-1. Built-in tools

3.2.3 BI results: enhancing decision-making and risk assessment

This study includes a comprehensive suite of 4 dashboards and 3 reports designed to analyze and visualize various aspects of risk assessment and data processing. Each dashboard focuses on a specific dimension, providing insights and comparisons essential for informed decision-making:

1) Register (Figure 0-2)
2) Final risk R, I, D (all treatments) (Figure 0-3)
3) Severity (all treatments) (Figure 0-4)
4) Probability (all treatments) (Figure 0-5)

5) Inherent risk vs ultimate risk (Figure 0-6)
6) Risk R in SQL vs DAX (Figure 0-7)
7) Risk I SQL vs DAX (I) (Figure 0-8)
8) Risk D SQL vs DAX (D) (Figure 0-9)

Figure 0-2 shows the list of logs. Each register contains a brief description of the treatment, for example: ID_REGISTRO 1 contains the treatment of care, diagnosis, prevention, rehabilitation by means of an electronic file; ID_REGISTRO 2 contains research and study activities in the medical, biomedical and epidemiological fields based on the data of the electronic health record and so on... Figure 0-3 and Figure 0-6 show the RID risk trend for (final) processing and (INHERENT) processing for each register. The risks are defined as follows:

- The **risk** of (final) processing is the magnitude of the risk relating to the "security of processing", i.e. the loss of R, I, and D of data and processing. This risk consists of two types of risk: inherent and residual.

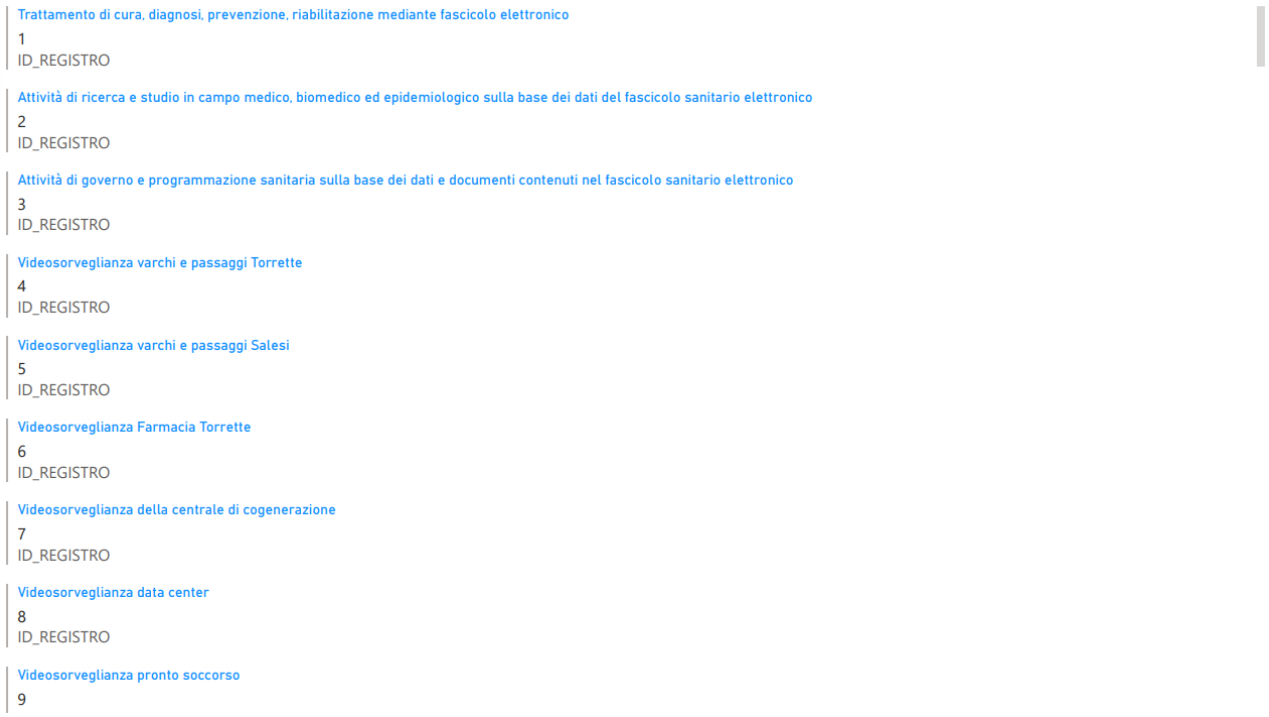

Figure 0-2. Elenco del registri (da 1 a 9) con descrizione sintetica del trattamento

- The **risk inherent in the processing** (INHERENT) is represented by the risk deriving from the processing in terms of R, I, and D in the absence of any mitigation intervention with general and/or specific technical and organizational measures.
- The **residual risk of the processing** (RID Risk) is represented by the risk deriving from the processing in terms of R, I, and D, considering the effect of general and/or specific mitigation measures.

Figure 0-7, Figure 0-8, and Figure 0-9 show the coincidence (except for some values) of the final RID risk values for each ID_REGISTRO (number of ID_REGISTRI = 169) in the calculation modes with DAX and SQL.

Figure 0-4 shows the trend of intrinsic (GI) and final (G) RID gravity for each ID_REGISTRO according to these definitions:

- The final severity is the severity given by the intrinsic severity and the multiplication factors which are: category of data subjects, age group, volume of data (number of data subjects and amount of data), category of processing, frequency of processing and category of recipients.
- Intrinsic severity is the severity given by the confidentiality of the category of personal data and the purposes of the processing. It is the purpose of the processing that allows the RID triad to be calculated.

Figure 0-5 presents the trend of the final RID probability (Mean of Prob.X.overall.tip13) for each registry (ID_REGISTRO). The final probability is defined with factors that depend on the type of process (paper or digital). If the processing is paper based, the probability is given by two factors: the value corresponding to the structuring of document management policies and procedures and paper archives; the probability amplification factor. If the processing is digital, the probability is given by three factors: the value corresponding to the structuring of the software systems and the technological chain; the Backup & Restore System; the probability amplification factor.

Calculation of the probability of computerized process management

- Take the average of the RID values extracted from GDP_REGISTRI_MINACCE with TIPO_PROCESSO = 1.
- Calculate the amplification coefficient by summing 1+ (R/I/D coefficient divided by 100) of the movements on GDP_REGISTRI_VALUTAZ with TYPOLOGY = 18 (in practice, if you have a row with R=5, I=4, D=0 the result will be R=1.05, I = 1.04, D=1)

- Multiply the values at the first point by the amplification coefficients, if the result is greater than 1 we assume 1.

Calculation of the probability of paper-based process management

Take the average of RID values extracted from GDP_REGISTRI_MINACCE with TIPO_PROCESSO = 2

Calculation of the overall probability

- If the processing is mixed and the percentage of automated management is indicated in (GDP_REGISTRI. PERC_GEST_AUTO), then the probability is calculated as (LINE AMOUNT 10 * PERC_GEST_AUTO/100) + (LINE AMOUNT 20 * (100-PERC_GEST_AUTO)/100)
- If the processing is mixed and the percentage of automated management is not indicated in (GDP_REGISTRI. PERC_GEST_AUTO), then the probability is calculated as the average of the sum of the values of rows 10 and 20 is calculated
- If the processing is only computer or only paper, the values calculated in line 10 or 20 are taken directly
- Multiply the resulting values by the amplification coefficients deriving from the "transfer to third countries" flag present in the record on GDP_REGISTRI_VALUTAZ with TYPOLOGY = 13

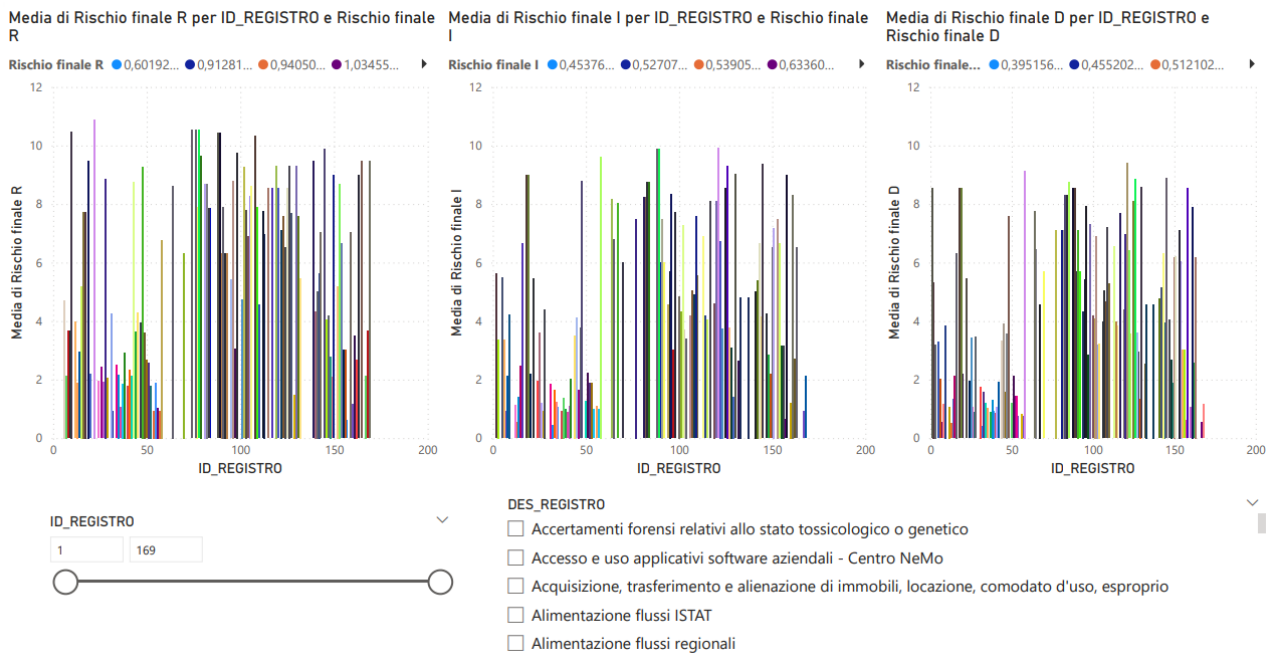

Figure 0-3. Dashboard with final risk of loss of confidentiality (R), integrity (I) and availability (D) of data and processing for each registry (ID_REGISTRO)

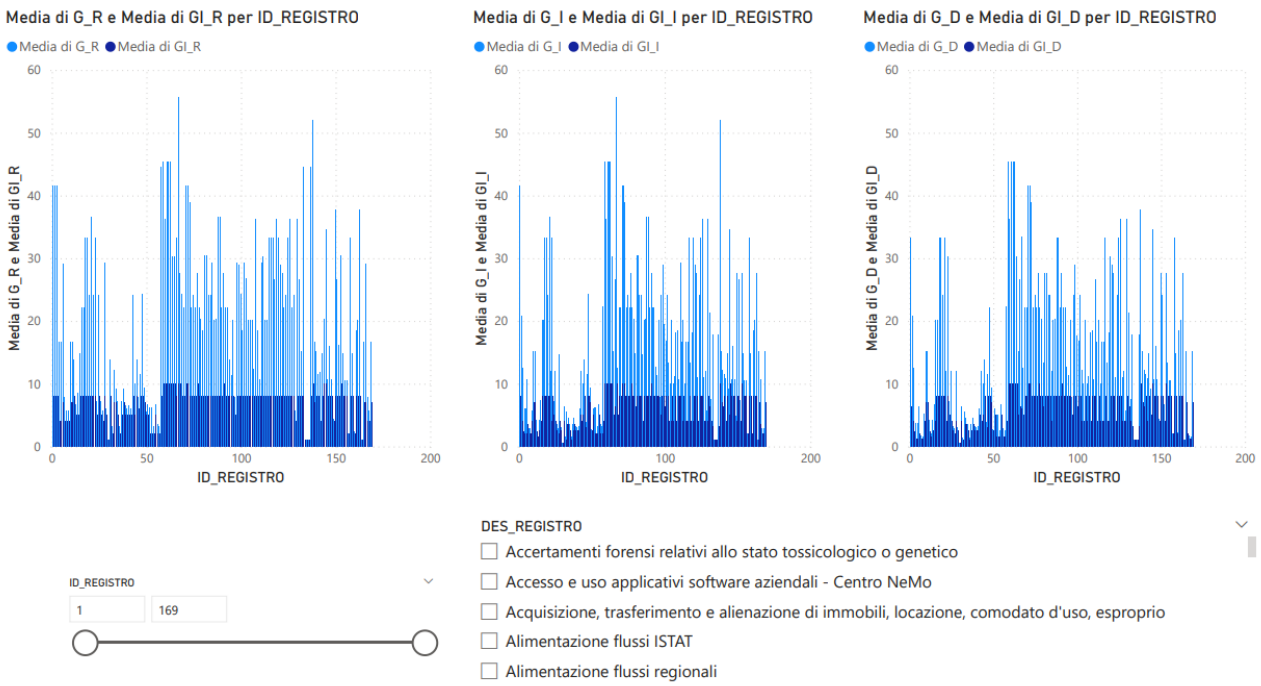

Figure 0-4. Dashboard with intrinsic vs. final severity of loss of confidentiality (GI_R vs. G_R average), integrity (GI_I vs. G_I average), and availability (GI_D vs. G_D average) of the data and processing for each registry (ID_REGISTRO)

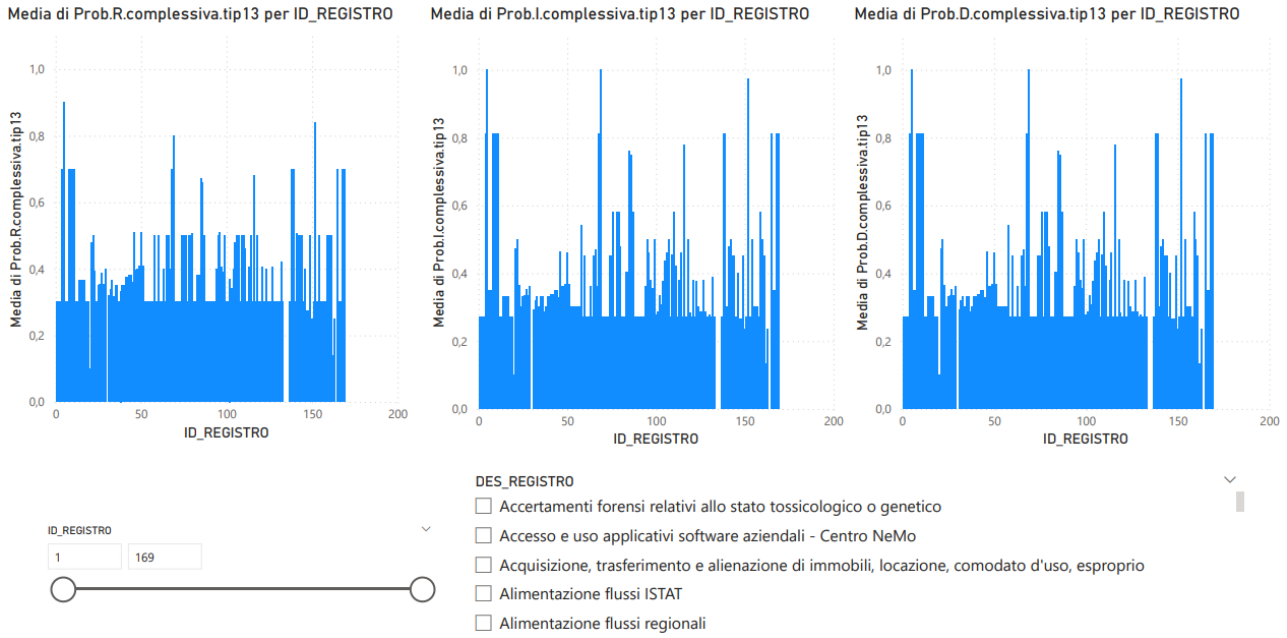

Figure 0-5. Dashboard with final probability of loss of confidentiality (Average of Prob.R.overall.tip13), integrity (Average of Prob.I.overall.tip13) and availability (Average of Prob.D.overall.tip13) of the data and processing for each registry (ID_REGISTRO)

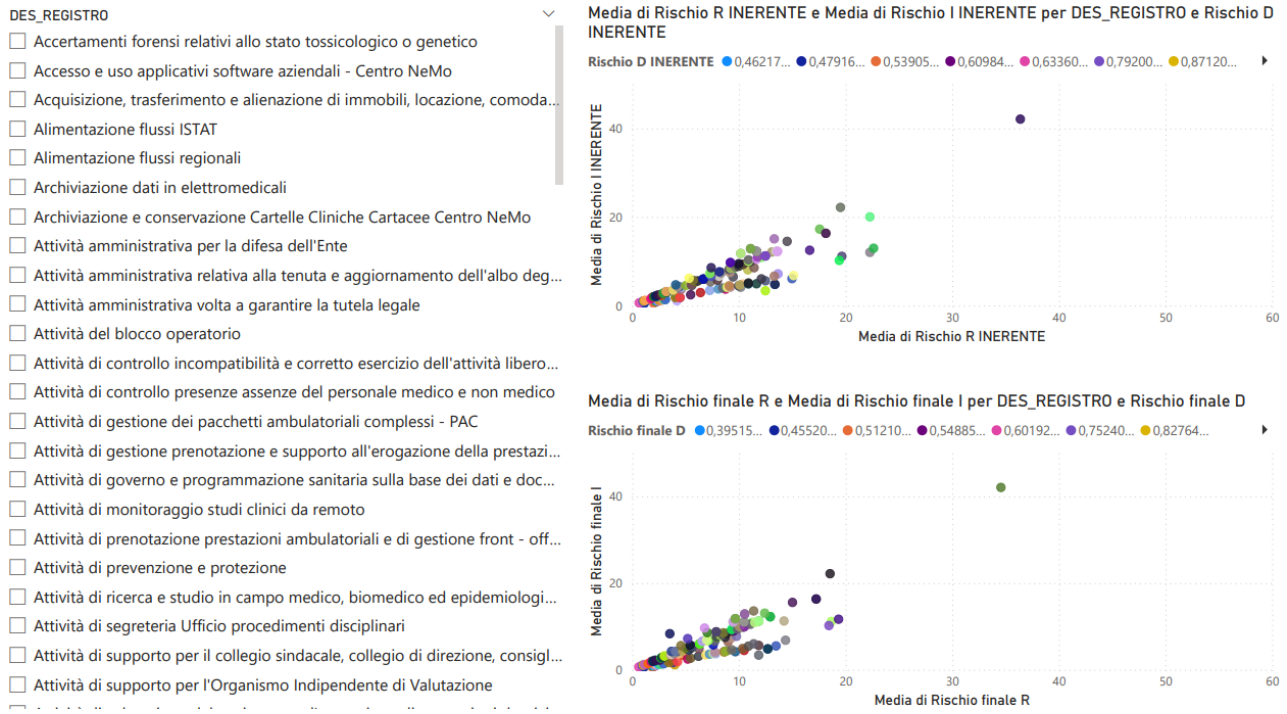

Figure 0-6. Top: dashboard with inherent risk of loss of confidentiality (Inherent Risk Average), integrity (INHERENT Risk Average) and availability (INHERENT Risk Average) of the data and processing for each registry (DES_REGISTRO). Down: dashboard with final risk of loss of confidentiality (Final Risk Mean R), integrity (Final Risk Average I) and availability (Final Risk Average D) of the data and processing for each registry (DES_REGISTRO)

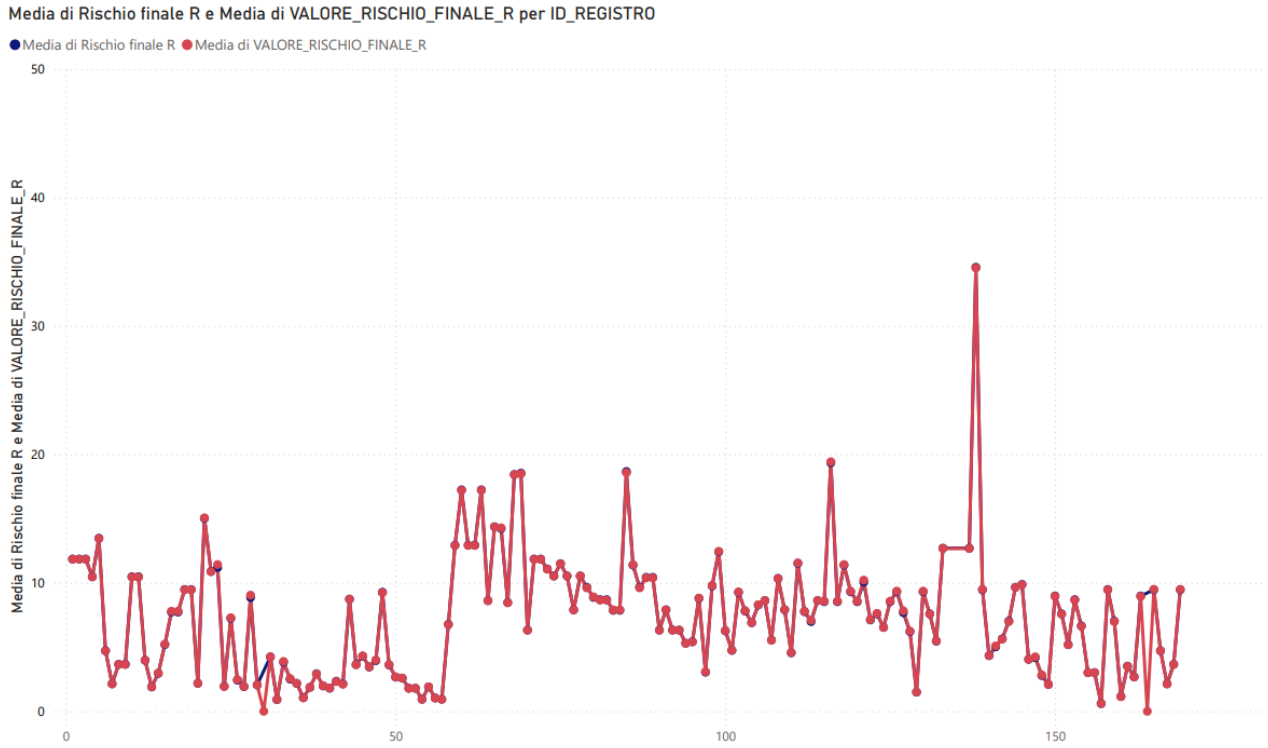

Figure 0-7. Final risk of confidentiality loss calculated with DAX in blue (Final Risk Average R) vs red SQL (Average of VALORE_RISCHIO_FINALE_R) for each log (ID_REGISTRO)

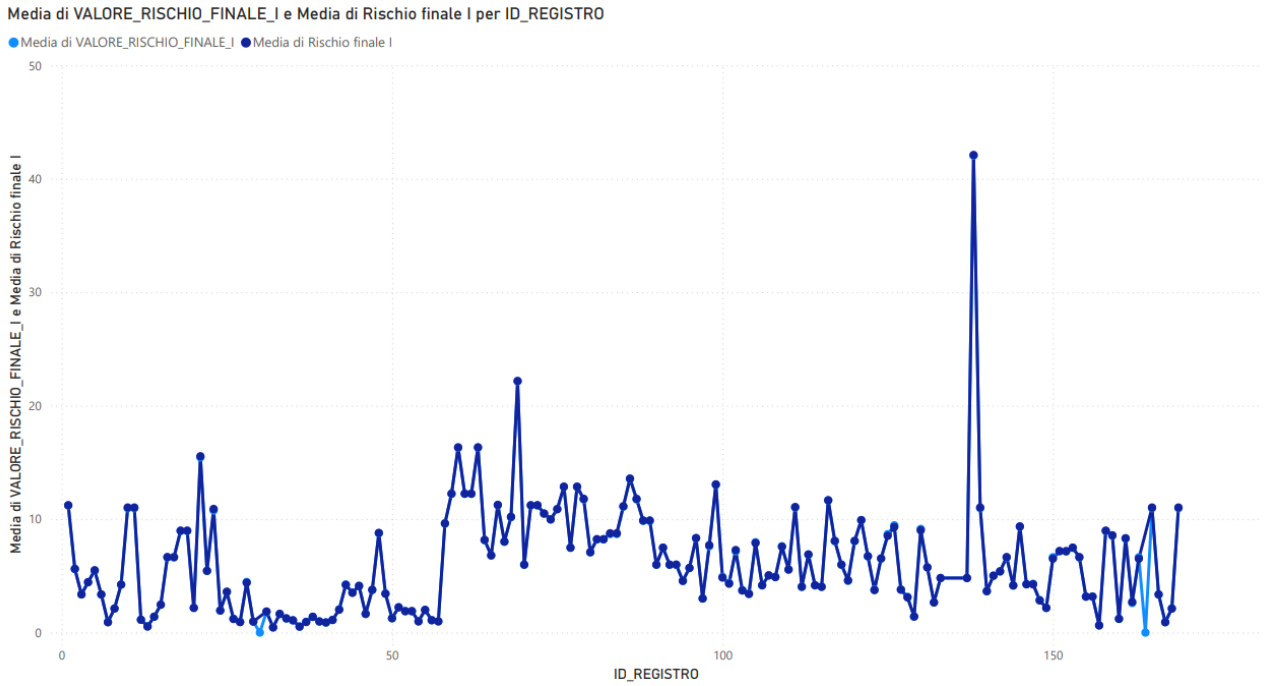

Figure 0-8. Final risk of loss of integrity calculated with DAX in blue (Final Risk Average I) vs SQL in light blue (Average of VALORE_RISCHIO_FINALE_I) for each register (ID_REGISTRO)

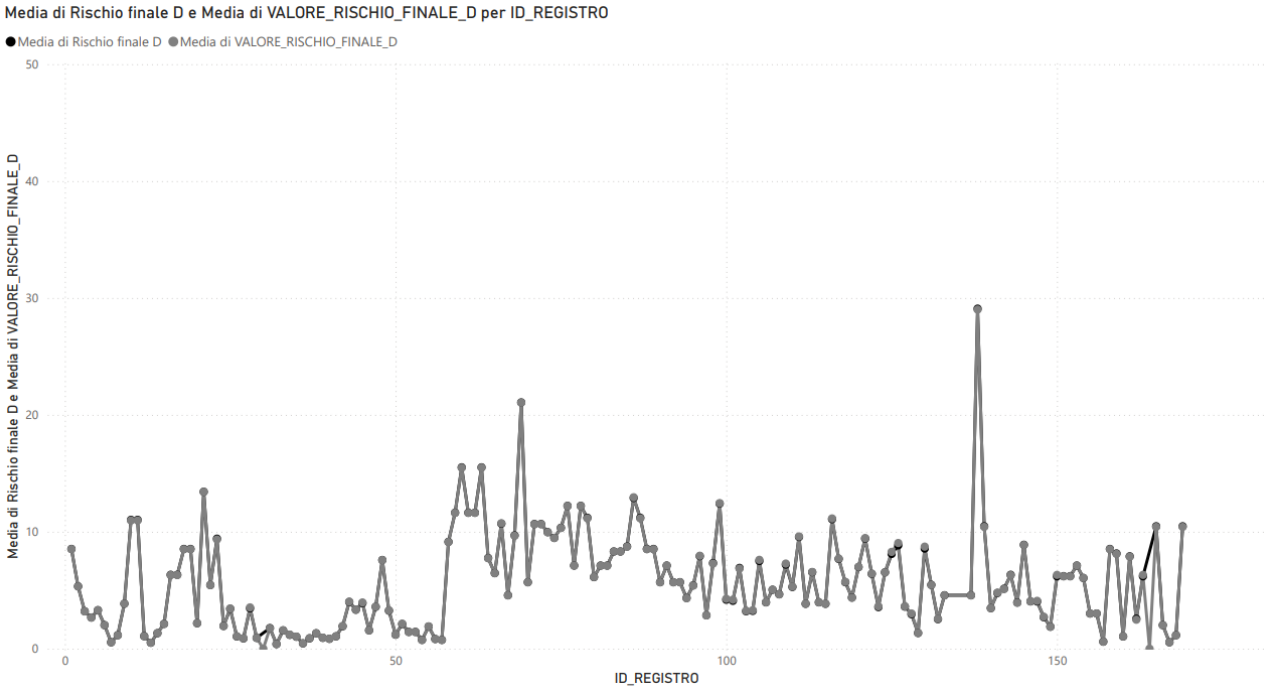

Figure 0-9. Final risk of loss of availability calculated with DAX in black (Average of Final Risk D) vs SQL in gray (Average of VALORE_RISCHIO_FINALE_D) for each register (ID_REGISTRO)

3.2.4 The advantages of Power BI in managing ethical risks: from data pre-processing to real-time visualization

Business intelligence (BI) with Power BI is an effective tool for monitoring ethical risks in data processing, offering a robust platform for risk management. Power BI's functionalities make it well-suited for tasks that involve complex data transformations and calculations, while remaining accessible and user-friendly.

The key advantages of Power BI for monitoring ethical risk are as follows:

1) **Data Pre-Processing and Calculation Capabilities**: Power BI supports integration with R for data pre-processing, allowing data collection, integration, and cleaning (handling missing values, removing duplicates, correcting errors). This preprocessing step ensures that data used for risk analysis is accurate and reliable. Furthermore, Power BI's integration of DAX enables sophisticated calculations for risk assessment metrics, such as the RID values, which are critical in evaluating ethical risk factors in data processing.
2) **Dynamic and Contextualized Calculations with DAX**: DAX is particularly well-suited for dynamic environments, as it enables calculations that adapt based on user filters and selections in real-time within Power BI reports. Unlike SQL, which is optimized for structured and static data in relational databases, DAX allows for complex, contextualized calculations that adjust interactively, a feature critical in BI reports. This functionality makes Power BI an ideal choice for organizations that require flexibility in viewing data across multiple contexts (Figure 0-7, Figure 0-8, Figure 0-9).
3) **Data Filtering and Visualization**: Power BI provides an intuitive interface for filtering and visualizing both entire datasets and partial datasets, which is essential for ethical risk monitoring. Users can filter by specific fields, such as ID_registro and DES_REGISTRO (Figure 0-3, Figure 0-4, Figure 0-5, Figure 0-6.), allowing for targeted analyses that help uncover potential risks in specific data segments. This functionality ensures that ethical risks are evaluated not only at an aggregate level but also within detailed subsets, making risk analysis both granular and comprehensive.
4) **Real-Time Risk Visualization and Alert System**: Power BI's dashboards and reports are updated instantly as new data becomes available, ensuring that risk levels are continuously monitored. This real-time update feature is essential for a proactive approach to ethical risk management. Additionally, Power BI can send alerts when specific risk thresholds are

crossed, enabling timely responses to potential compliance issues and protecting the organization from potential data security incidents.

5) **Accessibility Across Multiple Platforms**: Power BI offers flexibility in terms of data accessibility. Reports and dashboards can be viewed online through Power BI Service, on mobile devices with Power BI Mobile, and through integration with other Microsoft applications. This multi-platform accessibility allows for seamless risk monitoring, regardless of location, and ensures that the Data Protection Officer (DPO) and other stakeholders remain informed about risk levels.
6) **Ease of Use and Broad User Access**: Power BI is known for its user-friendly, intuitive interface, which makes it accessible to both technical and non-technical users. The drag-and-drop functionality for creating reports and dashboards reduces the learning curve, empowering knowledge workers, DPOs, and other stakeholders to easily create visualizations without advanced technical skills. This broad accessibility democratizes data analysis within the organization, making it easier for multiple teams to participate in ethical risk management.
7) **Integration with the Microsoft Ecosystem**: Power BI is deeply integrated with Microsoft tools such as Excel, Azure, Teams, and SharePoint, which enhances collaboration and data flow across platforms. For organizations already using the Microsoft ecosystem, this integration reduces operational friction, enabling seamless data sharing, risk reporting, and collaborative decision-making across departments.
8) **Cost-Effectiveness**: Power BI is recognized as one of the most affordable BI tools available. The Power BI Desktop version is free, allowing for offline data analysis and reporting, while Power BI Pro and Premium versions offer more advanced capabilities at competitive prices. This cost-effective pricing structure allows organizations of varying sizes to adopt Power BI for their business intelligence needs without a significant financial burden.

Power BI stands out as a leading business intelligence tool, with a range of features that make it highly suitable for monitoring ethical risks in data processing. Its ease of use, affordability, integration with the Microsoft ecosystem, and adaptability in dynamic reporting environments give it a competitive edge over other BI tools like Tableau, Qlik Sense, Looker, SAP BusinessObjects, and IBM Cognos. For organizations already invested in the Microsoft ecosystem, Power BI becomes especially advantageous, facilitating collaborative, real-time, and accessible risk monitoring.

The choice of BI tool ultimately depends on specific organizational requirements, including data complexity, budget, and the level of technical expertise needed. However, Power BI's feature set and accessibility make it a powerful and flexible option for organizations focused on managing and mitigating ethical risks in data processing effectively.

3.3 Different techniques in evasion attacks for adversarial machine learning

In this section, we will find some potential evasive attack mitigation techniques that have proven resilient when an attacker attacks a machine learning system. The attacker's goal is to generate malicious adversarial examples: test samples whose classification can be changed at the time of implementation to an arbitrary class of the attacker's choice with minimal perturbation [19]. The machine learning (ML) model can be tricked into classifying the adversarial example into the target class selected by the attacker, while humans still recognize it as part of the original class (Figure 0-10).

3.3.1 Mitigation of evasion attacks

Potential mitigation techniques are adversarial training, randomized smoothing, formal verification techniques. The real challenge of these techniques is the trade-off between the robustness and accuracy (classification accuracy is the proportion of correctly classified examples) of the system plus a computational cost. Therefore, the development of machine learning systems that can withstand evation-type attacks and simultaneously maintain accuracy is an open challenge.

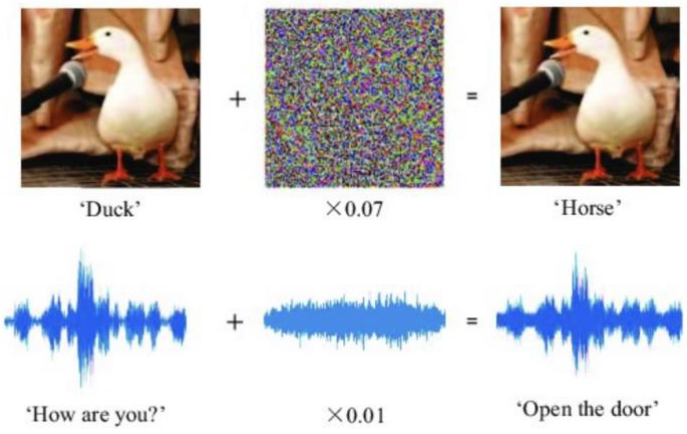

Figure 0-10. Adversarial examples in computer vision (top graph) and speech recognition (bottom graph) [74].

- Adversarial training, a technique introduced by Goodfellow et al. [18] and further refined by Madry et al. [75], is a comprehensive approach that enhances the training dataset by incorporating adversarial examples generated iteratively during the training process using their correct labels (Figure 0-11). The robustness of the trained model increases with the strength of the adversarial attacks used to create these examples. Notably, models developed through adversarial training exhibit more semantic significance compared to conventional models [76] . However, this advantage often comes with a trade-off in the form of reduced model accuracy on unaltered data. Furthermore, the iterative generation of adversarial examples during training makes adversarial training a computationally intensive process.
- Randomized smoothing, a technique introduced by Lecuyer et al. [77] and subsequently enhanced by Cohen et al. [78], is an approach that converts any classifier into a certifiable robust smooth classifier. This is achieved by generating the most probable predictions under Gaussian noise perturbations. The robustness of this technique leads to a decrease in accuracy in the system.
- Formal verification presents an approach to certify the adversarial robustness of neural networks, utilizing techniques from formal methods. While formal verification techniques show considerable promise for certifying neural network robustness, they face challenges in scalability and computational costs.

In addition to the trade-off between robustness and system accuracy, adversarial training also compromises the balance between robustness and fairness. The following case study, Antagonistic Debiasing presents this trade-off challenge between robustness and fairness.

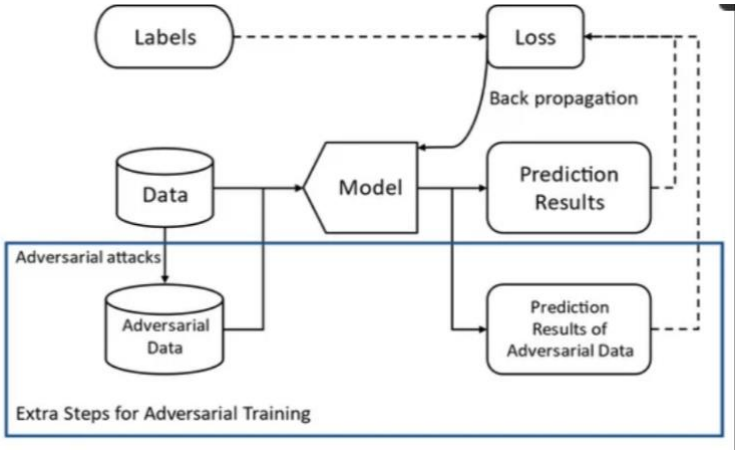

Figure 0-11. Adversarial training in the neural network model [79].

3.3.2 Case study: adversarial de-biasing

Adversarial Debiasing is an in-processing mitigation algorithm designed to promote fairness. This technique leverages adversarial training to reduce bias, involving the simultaneous training of a predictor and a discriminator. The predictor is responsible for making accurate predictions about the target variable, while the discriminator tries to predict a protected variable (such as gender or ethnicity) based on the predictor's predictions. The main goal is to maximize the accuracy of the predictor on the target variable, while reducing the effectiveness of the discriminator in recognizing the protected variable.

In this case study, the evaluation of fairness is based on two fairness metrics: disparate impact (DI) and statistical parity difference (SPD).

1. **Disparate impact**: measures the ratio of favorable outcomes for a non-privileged group to those of the privileged group.

$$DI = P\left(\hat{Y} = 1 \middle| A = non-privileged\right) / P\left(\hat{Y} = 1 \middle| A = privileged\right)$$

An ideal value for this metric is 1, which means that the ratio of favorable outcomes is the same for both groups.

1. **Statistical parity difference**: This is very similar to a disparate impact. Instead of ratio, it measures the difference in favorable outcomes between non-privileged groups and privileged groups.

$$SPD = P(\hat{Y} = 1 \mid A = non-privileged) - P(\hat{Y} = 1 \mid A = privileged)$$

An ideal value for this metric is 0, , which means that the difference in favorable outcomes is the same for both groups.

For this case study, we use Orange Data Mining software.

We use the well-known "Adult" dataset. The Adult dataset consists of 48,824 instances with 15 attributes describing demographic details from the 1994 census. The main task is to predict whether an individual earns more than $50,000 per year, with "sex" as the protected attribute and "male" as the privileged protected attribute.

We train two adversarial debiasing models, one with and one without debiasing, and compare them to a commonly used machine learning algorithm known as Random Forests (Figure 0-12). The adversarial debiasing models require the configuration of certain parameters (Figure 0-13).

- **Neurons in hidden layers** specifies the number of neurons in each hidden layer of the neural network.
- **Use Debiasing** determines whether debiasing should be applied. If this option is not selected, it will operate as a standard neural network model.
- **Adversary loss weight** defines the weight of the adversary loss in the total loss function. The adversary loss corresponds to the discriminator's loss function. A higher weight will make the model focus more on reducing the discriminator's ability to predict the protected variable, potentially at the expense of the predictor's accuracy in predicting the target variable.

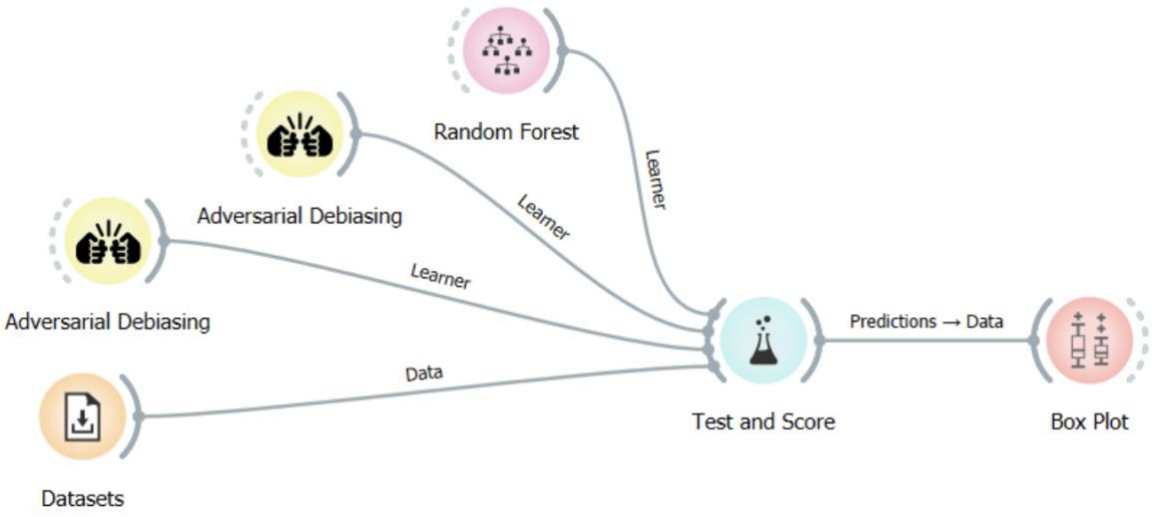

Figure 0-12. Machine learning system [80]

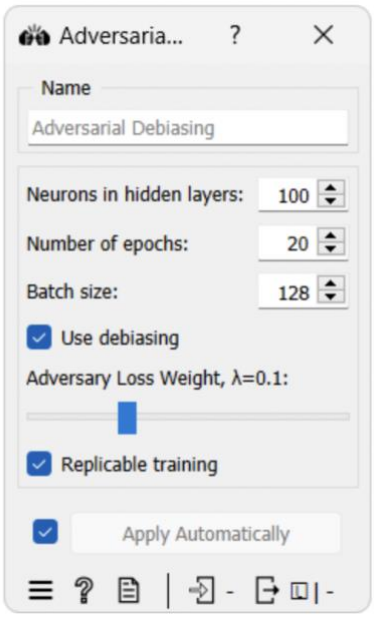

Figure 0-13. Adversarial de-biasing that minimizes bias. [80]

The results show that implementing debiasing techniques enhanced fairness metrics (Disparate Impact and Statistical Parity Difference) whereas in the absence of debiasing, the outcomes closely resembled those of the Random Forest model. In fact, disparate Impact shifted from 0.294 to 1.051,

while Statistical Parity Difference changed from -0.180 to 0.006, suggesting minimal bias between groups regarding favorable outcomes (Figure *0-14*).

In addition, the application of debiasing methods may result in a slight reduction in accuracy, because the model aims to balance accuracy and fairness. This trade-off is acceptable because the accuracy remains comparable to that of the scenario without debiasing measures, and simultaneously, the bias is minimized.

The box plot at the top chart (Figure 0-15) shows that without debiasing, males (the privileged group) are classified as earning “>50K” more often than females (the unprivileged group), indicating bias, as reflected by suboptimal Disparate Impact (0.294) and Statistical Parity Difference (-0.180) metrics. In the box plot at the bottom chart, with debiasing applied, males and females are classified into the favorable class at similar rates, with metrics (Disparate Impact of 1.051 and Statistical Parity Difference of 0.006) close to their optimal values, showing minimal bias.

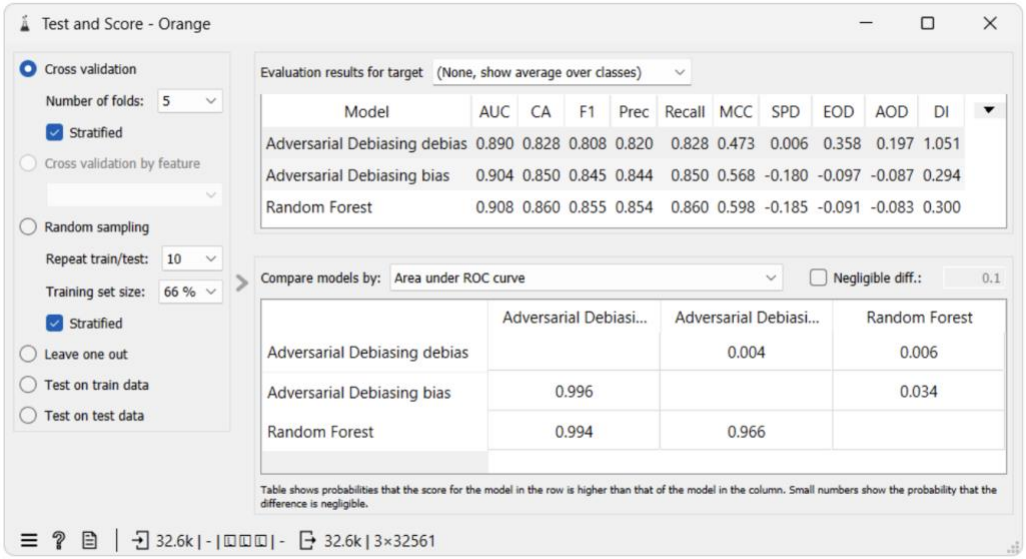

Figure 0-14. Fairness metrics (SPD and DI) and accuracy (CA) in different models [80]

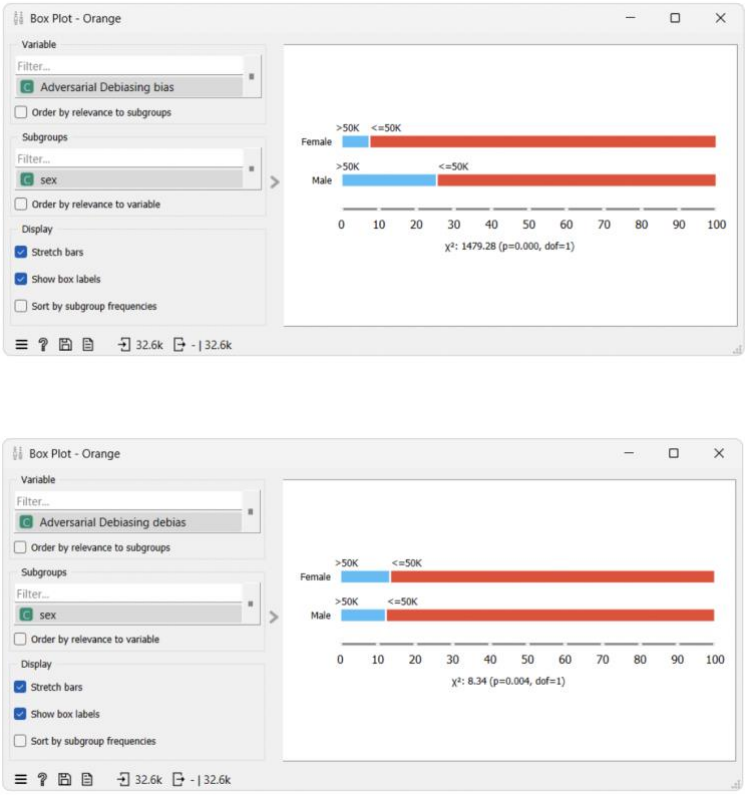

Figure 0-15. Fairness in sex with antagonistic de-biasing that minimizes bias (bottom chart) vs antagonistic de-biasing that doesn't minimize bias (top chart) [80]

Predictive AI can be at risk of evasive malicious attacks that target data integrity (Section 1.4). There are various methods to mitigate this risk and strengthen the ethics of AI. One method is antagonistic training, which has advantages and disadvantages in terms of accuracy and robustness. The study of adversarial de-biasing, which is used to defend the model against attacks that seek to introduce bias through malicious examples, has these disadvantages and advantages in terms of accuracy and fairness. Predictive AI can be at risk of evasion attacks targeting data integrity (Section 1.4). Several methods exist to mitigate this risk and strengthen AI ethics. One method is adversarial training, which has both advantages and disadvantages in terms of accuracy and robustness. The study of adversarial de-biasing presents both disadvantages and advantages in terms of accuracy and fairness. Nevertheless, it is a valid mitigation method for defending the model against attacks that seek to introduce bias through malicious examples.

Chapter 4: Ethics of cybersecurity and data protection in Quantum AI

The quantum era can bring benefits in terms of the reduction of parameters in artificial intelligence, with possible ethical benefits. While future applications in the healthcare world may involve AI technology, the quantum era also brings with it a strong demand for new frameworks for data protection and cybersecurity. Therefore, this study aims to identify the new ethical aspects of data protection and cybersecurity in the quantum era. In particular, the study identified what new ethical aspects related to data collection and processing emerge, and what possible ethical issues will arise in future cybersecurity strategies involving quantum and post-quantum cryptography. The responsible integration of cybersecurity measures into quantum AI strategies and the development of robust quantum data governance frameworks will ensure the protection of digital assets in the quantum era.

4.1 Hybrid quantum machine learning: cybersecurity challenges and ethical considerations

Quantum AI can be approached through various configurations, including classical data paired with classical algorithms, classical data with quantum algorithms, quantum data with quantum algorithms, and quantum data with classical algorithms. In the previous chapters, we explore the foundational setup of classical data and classical algorithms, providing a reference framework. In this chapter we delve into the second configuration: classical data with quantum algorithms. This latter configuration can be called hybrid quantum machine learning, where the data collected are classical, and through quantum encoding, they become quantum type. Quantum data are processed through a quantum circuit and then fall back into the classical world through the measurement process. Summarizing the various steps that exist in a hybrid quantum machine learning system, we find a classic type of preprocessing, a quantum type of processing, and a classic type of post-processing. The steps are as follows:

1) Classical preprocessing: data collection, data cleaning and preparation, feature selection or extraction, and data encoding for quantum input
2) Quantum processing: quantum encoding of classical data–quantum circuit operations–quantum algorithm execution

3) Classical post-processing: measurement of quantum states, data interpretation and analysis, result evaluation and validation, model refinement or optimization.

This combines the strengths of classical and quantum computing paradigms, leveraging classical data collection and preparation techniques with the computational power of quantum algorithms for processing. Finally, the system returned to the classical domain for final analysis and interpretation of the results. Below, we present more technical steps of pre-processing, processing, and post-processing with reference to quantum aspects.

1) Pre-processing:
    - Encodes classical data into quantum representations
    - Prepares quantum state vectors of qubits
2) Processing:
    - Performs quantum operations on the encoded data
    - Utilizes quantum superposition and entanglement for computations
3) Post-processing:
    - Conducts measurements on the quantum output
    - Interprets results for the learning model's predictions

Having outlined the technical underpinnings of quantum machine learning systems, it is necessary to examine the cybersecurity risks these innovations face, underscoring the need for robust protective measures in this emerging field.

Let's imagine that we initially have a classic machine learning system with classical data. These machine learning systems could be hit by an attacker (a cracker) who carried out an attack and managed to enter the system. We therefore ask ourselves what cyber-attacks are present today in order to understand what risks a classic machine learning system is posed to. At the same time, let's now imagine that we have designed and developed a hybrid machine learning system (classical-quantum) and ask ourselves what cyber-attacks could affect the system, how they may differ from previous attacks carried out in a classical world and what are the future protection techniques. Transversally to cyber-type attacks in the classical world, we will also examine about what ethical frameworks for cybersecurity have been developed and how they should change, also considering post-quantum cryptography and quantum cryptography

Therefore, this study aims to clarify what are the aspects of data protection and the security measures that should be adopted when a hybrid machine learning system (classical-quantum) is designed and developed, taking into account the ethical issues within it.

4.2 Cryptographic techniques

Cryptographic techniques are methods and algorithms used to secure communication, protect data, and ensure privacy in various digital systems. These techniques are typically divided into two main categories: symmetric key cryptography and asymmetric key cryptography. Additionally, there are other techniques such as hash functions and digital signatures. The primary cryptographic techniques are as follows:

1) Symmetric Key Cryptography (Secret Key Cryptography)

In symmetric cryptography, the same key is used for both encryption and decryption. The key must be shared secretly between the communicating parties. Common techniques are:

- AES (Advanced Encryption Standard): a widely used block cipher that encrypts data in fixed-size blocks of 128 bits, with key sizes of 128, 192, or 256 bits.
- DES (Data Encryption Standard): an older block cipher that encrypts data in 64-bit blocks using a 56-bit key. It is now considered insecure.
- 3DES (Triple DES): a variant of DES that applies the DES algorithm three times to each data block to enhance security.

2) Asymmetric Key Cryptography (Public Key Cryptography)

In asymmetric cryptography, two keys are used: a public key for encryption and a private key for decryption. This allows secure communication without sharing a secret key in advance. Common techniques are:

- RSA (Rivest–Shamir–Adleman): one of the first and most widely used public-key cryptosystems. It relies on the difficulty of factoring large numbers.
- ECC (Elliptic Curve Cryptography): a modern public-key cryptosystem based on elliptic curves, offering the same level of security as RSA but with much smaller key sizes.
- Diffie-Hellman (DH) Key Exchange: a method for securely exchanging cryptographic keys over a public channel.
- El Gamal Encryption: an encryption algorithm based on the Diffie- Hellman key exchange and used in many cryptographic protocols.

3) Hash Functions

Hash functions transform input data of any size into a fixed-size output, known as a hash or digest. They are primarily used for data integrity and verification, not encryption. Common Techniques:

- SHA-2 (Secure Hash Algorithm 2): a family of hash functions, including SHA-256 and SHA-512, used in many cryptographic applications.
- SHA-3: the latest member of the Secure Hash Algorithm family, based on the Keccak algorithm.
- MD5 (Message Digest Algorithm 5): a widely used but now insecure hash function.

4) Digital Signatures

Digital signatures are a type of cryptographic algorithm that provides authentication and integrity for a message or document. A digital signature ensures that the message has not been altered and verifies the identity of the sender. Common Techniques:

- RSA-based Digital Signatures: RSA can be used for both encryption and digital signatures, ensuring message integrity and sender authentication.
- ECDSA (Elliptic Curve Digital Signature Algorithm): a digital sig- nature algorithm based on elliptic curves, offering the same benefits as RSA with smaller keys.
- DSA (Digital Signature Algorithm): a federal standard for digital signatures, based on discrete logarithms.

These techniques form the backbone of secure communication and are applied in many fields, including online transactions, secure messaging, authentication systems, and data protection

4.3 Post quantum cryptography vs quantum cryptography

We have previously seen the classic encryption techniques used to secure systems, data and communications. Now let's see the new techniques of quantum and post-quantum cryptography we have to use to secure classic systems with the advent of the quantum computing Quantum cryptography exploits the properties of quantum mechanics for cryptographic purposes such as quantum key distribution (QKD). QKD is a method that allows two parties to generate a shared,

secret key using quantum properties of particles, ensuring the security of the key exchange. QKD exploits the non-cloning theorem: when qubits are observed, they collapse into the classical domain; therefore, a perfect copy operation of a qubit found in a superposition state cannot be performed. Therefore, QKD can be considered the 2nd perfect cipher after the one-time pad designed during World War II. Quantum cryptography, a set of information protection techniques based on the use of quantum technologies by legitimate users, is contrasted with Post-Quantum Cryptography (or Quantum-Safe Cryptography), in which legitimate users use non-quantum technologies to defend themselves against attackers equipped with quantum technologies. To date, the security of cryptographic systems is based on the difficulty of solving certain mathematical problems with limited computation on the part of the attacker. One mathematical problem in which security is based is the difficulty of computing the factorization of large integers with classical computers. With a classical computer, one would have to perform all the tests and therefore have exponential complexity, that is, the computation time is exponential in log N, where N is an integer. Using a quantum computer (Shor's Algorithm of 1994), this time becomes a polynomial in log N. This reduces the speed of factoring the numbers, and thus, the ease of attack. Another problem for the attacker is searching for an unordered list for an item. With a classical computer, one would have to make all attempts, and this involves complexity proportional to N, where N is the length of the list, that is, the computation time is proportional to N. With a quantum computer (Grover's Algorithm of 1996), the time taken to find an item in an unordered list is proportional to N, speeding up the attacker's time to launch the attack. In the worst-case scenario, many of the security systems used today are attacked. These systems are: RSA public key cryptosystem based on integer factorization used in SSL/TLS, online banking, ATM,...; El Gamal public key cryptosystem based on discrete logarithm, used in SSL/TLS,...; DH and Elliptic- curve Diffie–Hellman (ECDH) key exchange protocols based on discrete logarithm used in SSL/TLS, NFC/contactless,...; others systems are Elliptic Curve Cryptography (ECC), DSA, and (ECDSA). According to the National Security Agency, the practical solution to attacks by quantum computers is not quantum cryptography because of the number of devices required for communication, increased infrastructural costs, and risks of attacks from within (such as denial of service).Therefore, the solution is post-quantum cryptography, that is, changing mathematical problems and defending ourselves with classical computers through mathematical methods to which quantum attacks do not apply. This is considered the most practical solution by NIST. This solution is also shared by ANSSI, BSI, NLNCSA, and SNCSA/SAF. Therefore, since 2016, the National

Institute of Standards and Technology (NIST) has initiated a process for the development and standardization of one or more public key cryptographic algorithms. The subsequent section outlines the ethical considerations relevant to these two approaches.

4.3.1 Ethical issues in post quantum cryptography vs quantum cryptography

There are two options to secure systems against quantum attacks with different ethical consequences [81]. The first option is to strengthen current systems using post-quantum algorithms (e.g., lattice systems, coding- based systems, super singular isogenies, hash- based signatures, etc.) to immediately guarantee higher levels of security. This option, however, implies a) a significant increase in controls and a progressive diminishment of the privacy and autonomy of companies and citizens b) a decrease in the transparency of government activities c) the moral risk of paternalism, that is, the government deciding what is best for citizens and companies without consulting them. The second option is to accelerate the migration from our current cryptographic systems to fully quantum systems. The risk is that this migration is too slow and therefore a) incomplete, exposing it to possible attacks b) does not cover some of the interested parties, that is, users and companies that have not yet developed the ap- propriate technology. Both consequences could significantly exacerbate disparities between countries and citizens, as well as between those who currently possess quantum cryptographic systems and those who do not, thereby compromising their security and privacy. Such disparities could also engender novel forms of exploitation and undermine human dignity. Replacing existing safety systems could take over 20 years [82], and sensitive data may still be at risk. Moreover, quantum key distribution poses challenges of excessive privacy and security, potentially enabling criminal activities and necessitating a reevaluation of privacy, welfare ethics, and the acceptability of non-breakable cryptography. The use of quantum computing forces us to differentiate between two aspects of privacy: privacy as the control of information and privacy as the construction of one's social identity [83]. In classical communication systems, privacy is linked to the control of information dissemination and, consequently, to one's social identity, which is how we are perceived and regarded by others. However, in the quantum world, this relationship changes. On one hand, managing information becomes more complicated due to security, as previously dis- cussed in this report. On the other hand, quantum computing demands that we reconsider the development of our digital identity in alternative ways. This is an uncharted territory that necessitates further research.

4.4 Security attacks, vulnerabilities, and defenses in quantum machine learning

Different attacks are aimed at hitting different features and various measures have been identified to mitigate such attacks [84], [85]. Below we see some examples.

The study [4] discusses various vulnerabilities in quantum machine learning, focusing on hardware, compilation, and fault injection attacks. • Identical Coupling Hardware: quantum cloud providers like IBM, D-Wave, and IonQ use different hardware with varying quality. Users cannot distinguish between identical coupling maps, leading to possible allocation of low- quality hardware, which negatively impacts Quantum Machine Learning (QML) performance. • Compilation Quality: compilers convert high-level quantum programs into hardware-specific gate sets. While companies like IBM and Rigetti provide compilers, third-party tools may optimize performance but pose security risks, including theft of intellectual property (IP) and malicious code insertion. • Embedding Sensitive Information: sending QML circuits with proprietary data to third-party compilers exposes them to potential IP theft. Sensitive circuits containing financial data or proprietary algorithms are vulnerable when sent to untrusted cloud providers. • Unreliable Hardware Allocator: third-party providers may as- sign poor-quality hardware, degrading QML performance. The Quantum Physically Unclonable Function (QuPUF) was proposed to verify hardware identity before execution, protecting against this threat. • Compilation- Oriented Attacks: adversaries could steal information from QML circuits. To counter this, circuits can be split and sent to different compilers, in- creasing security. Additionally, inserting dummy SWAP gates in circuits helps prevent reverse engineering. • Fault Injection Attacks: crosstalk errors and external adversaries can degrade QML performance by injecting faults into shared hardware. Buffer qubits can mitigate this by isolating programs running on neighboring qubits, improving reliability. These methods aim to enhance security in quantum computing environments.

The study [5] outlines vulnerabilities and attack vectors unique to quantum machine learning models, divided into two parts: 1. Quantum attack vectors. These exploits target specific aspects of QML models that don't have classical equivalents. Key examples include: o Fault Injections: a "quantum Trojan virus" can covertly insert gates into a Quantum Neural Network (QNN) to alter its performance. The attack remains undetected until the virus is activated, leading to nearly 100o Exploiting Quantum Noise: nn shared quantum computing environments, attackers can degrade model performance or launch denial-of-service attacks by inducing hardware errors, such as cross-

talk in superconducting systems or repeated operations in ion-trap systems. Buffering or recognizing malicious circuit patterns can help mitigate these risks. 2. Scaling Pitfall. As quantum systems grow larger and more complex, they become more sensitive to small perturbations, which can threaten the reliability of quantum classification. This increased sensitivity demands more resources for verification, potentially negating quantum advantages. Research also highlights that adversarial perturbations become more effective in higher-dimensional systems, increasing the risk for QML classifiers as qubit counts rise.

This study [5] outlines various defenses against attacks on QML models, categorizing them into three main areas: adversarial training, differential privacy, and formal verification. 1. Adversarial Training. It enhances quantum models by exposing them to crafted malicious inputs, improving their robustness. Common classical methods include: • Fast Gradient Sign Method (FGSM). It perturbs inputs using gradient signs to maximize model error. • Basic Iterative Method (BIM). An enhanced FGSM that applies perturbations iteratively. • Projected Gradient Descent (PGD). It keeps adversarial examples within a valid range. • Momentum Iterative Method (MIM). It uses momentum to evade local maxima. Research shows QML models are vulnerable to these classical attacks, but adversarial training significantly enhances their defenses. Notably, studies have demonstrated that quantum classifiers can outperform classical ones in

robustness as they scale. Recent advancements include a quantum adversarial metric learning approach and random unitary encoding strategies that improve adversarial resilience. 2. Data Privacy. As data breaches become more prevalent, ensuring data privacy in training datasets is critical. Two main strategies are: • Differential Privacy. This framework protects individual data by adding noise to computations, helping models generalize without compromising privacy. Research has shown that in- creasing rotation noise in quantum training boosts classifier robustness.

Inherent Privacy. Overparameterized QML models offer inherent protection against gradient inversion attacks, especially in federated learning settings. 3. Formal Verification. To ensure robustness against unidentified adversaries, formal verification is essential. Key concepts include:

Ensuring robustness relies on having a generative model for sample distributions and resilient classifiers. This approach links randomness, generative models, and cryptographic protocols. • MILP Verification. A novel method transforms robustness verification into a mixed-integer linear program, using quantum algorithms for efficient neural network certification. • Lipschitz Continuity. This involves deriving an analytical measure for robustness using semi-definite

programming, providing practical tools for verifying QML models. Overall, these defenses form a comprehensive framework for securing QML applications against adversarial threats, enhancing both privacy and model robustness.

4.5 Data protection in hybrid-quantum machine learning

This section examines a possible strategy for protecting data collected and analyzed in a hybrid machine learning system using a technique, called Fully Homomorphic Encryption (FHE), which falls under post-quantum cryptography. FHE is gaining traction as a valuable tool for privacy- preserving machine learning, extending its potential into the realm of quantum machine learning (QML). FHE allows computations to be per- formed directly on encrypted data, thus maintaining privacy and security, which is especially crucial given the threats posed by quantum computers to traditional cryptographic schemes [86], [87].

The concept of Quantum Fully Homomorphic Encryption (QFHE) allows for secure quantum computations while preserving the privacy of input data. This is important as QML systems often involve sensitive information. One challenge in applying FHE to quantum computing is the verification of quantum circuits, complicated by the no-cloning theorem. However, advancements such as verifiable QFHE (vQFHE) enable non-interactive delegation and verification of quantum computations [86].

Despite these advancements, applying FHE to quantum contexts remains a nascent field, facing challenges such as performance overhead and the complexity of quantum algorithms [88]. Notably, the integration of FHE with machine learning models may require significant modifications and could impact performance, particularly in deep learning applications like Convolutional Neural Networks (CNNs) and Extreme Gradient Boosting (XGBoost) [16], [87], [89].

Emerging frameworks allow for training and inference of privacy-preserving machine learning models without interactive decryption, presenting possibilities for adaptation to quantum algorithms. However, challenges persist, including maintaining a balance between security, accuracy, and computational complexity.

In conclusion, while FHE and QFHE provide promising avenues for se- cure quantum machine learning applications, particularly in sensitive fields such as healthcare and genomics, ongoing research is essential to address the complexities and performance issues involved.

4.6 Overview of the main points

This study discusses the various aspects of cybersecurity and ethics in the context of classical and quantum computing, focusing on hybrid quantum machine learning systems. It outlines the pre-processing, processing, and post-processing steps involved in such systems and the potential cyber risks they face, including malware, denial of service attacks, and adversarial attacks on machine learning models. Considering these attacks, the study also covers cryptographic techniques used in classical computing, such as symmetric and asymmetric key cryptography, hash functions, and digital signatures. It then delves into the ethics of cybersecurity, highlighting the need for detailed ethical frameworks to guide decision-making in practice. The discussion then shifts to quantum cryptography, comparing post-quantum cryptography and quantum cryptography. It explores the ethical implications of these approaches, such as the trade-offs between security, privacy, and the potential for exploitation. The study also ex- amines security attacks, vulnerabilities, and defenses specific to quantum machine learning, including fault injection attacks, exploiting quantum noise, and the challenges posed by scaling quantum systems. Various defense strategies are discussed, such as adversarial training, differential privacy, and formal verification. Finally, the summary touches on the potential of Fully Homomorphic Encryption (FHE) and its quantum counterpart (QFHE) for privacy-preserving quantum machine learning, while acknowledging the challenges and complexities involved in their implementation. Consequently, this investigation examines the comprehensive landscape that will emerge with the advent of quantum AI and contributes to the research on future risks to data protection and cybersecurity that quantum computing and, specifically, quantum AI will encounter, the associated ethical considerations, and technical and organizational measures for the mitigation of future risks.

Conclusion

In an era of rapid technological advancement, marked by the rise of artificial intelligence and quantum technologies, addressing the ethical challenges in cybersecurity and data protection has become more critical than ever. This study has provided a comprehensive overview of strategies for assessing, mitigating, and adapting to emerging ethical risks, highlighting the necessity of a methodical and adaptive approach.

The initial analysis of ethical frameworks for cybersecurity revealed the complexities of managing risks related to data protection and the introduction of predictive algorithms and machine learning, which raise new questions about fairness, transparency, and accountability. The proposed ethical risk assessment strategies incorporate both qualitative and quantitative approaches, demonstrating how diverse models can address specific needs, from data protection analysis to IT security and AI implementation. The distinction between defensive and offensive strategies, such as managing ethical biases in automated filtering systems, underscores the importance of active oversight of ethical risks throughout the entire digital technology lifecycle.

Risk mitigation solutions, particularly through blockchain and Business Intelligence (BI), illustrate how advanced tools and methodologies can enhance resilience against cyber threats while improving the ability to monitor and assess ethical risks. In the context of BI applied to data protection, tools like Microsoft Power BI have shown how real-time analytics and data visualization can support more informed and timely decision-making. Furthermore, the adaptive approach to countering evasion attacks in machine learning highlights the importance of techniques like de-biasing, which is essential for maintaining the reliability of intelligent systems.

Finally, the transition to Quantum AI applications introduces new technical and ethical complexities, especially concerning post-quantum and quantum cryptography. The specific vulnerabilities and defenses developed for quantum machine learning require a focused effort on data protection and the formulation of new ethical standards. Advanced cryptographic approaches and data protection strategies for hybrid quantum systems demonstrate the need to prepare for increasingly sophisticated threats in this domain.

In conclusion, this study emphasizes the importance of continuously developing and updating ethical frameworks for cybersecurity, adapting them to the evolving landscape of emerging technologies. Balancing technological progress with the safeguarding of fundamental ethical values is essential for building a secure, fair, and responsible digital future.

References

[1] O. Radley-Gardner, H. Beale, and R. Zimmermann, Eds., *Fundamental Texts On European Private Law*. Hart Publishing, 2016. doi: 10.5040/9781782258674.

[2] National Institute of Standards and Technology, "The NIST Cybersecurity Framework (CSF) 2.0," National Institute of Standards and Technology, Gaithersburg, MD, NIST CSWP 29, Feb. 2024. doi: 10.6028/NIST.CSWP.29.

[3] I. Flechais and G. Chalhoub, "Practical Cybersecurity Ethics: Mapping CyBOK to Ethical Concerns," in *Proceedings of the 2023 New Security Paradigms Workshop*, in NSPW '23. New York, NY, USA: Association for Computing Machinery, Dec. 2023, pp. 62–75. doi: 10.1145/3633500.3633505.

[4] P. Formosa, M. Wilson, and D. Richards, "A principlist framework for cybersecurity ethics," *Computers & Security*, vol. 109, p. 102382, Oct. 2021, doi: 10.1016/j.cose.2021.102382.

[5] K. N. J. Macnish and J. van der Ham, "Ethical Approaches to Cybersecurity," in *The Oxford Handbook of Digital Ethics*, Oxford University Press, 2022. doi: 10.1093/oxfordhb/9780198857815.013.28.

[6] J. Fenech, D. Richards, and P. Formosa, "Ethical principles shaping values-based cybersecurity decision-making," *Computers & Security*, vol. 140, p. 103795, May 2024, doi: 10.1016/j.cose.2024.103795.

[7] B. Sadeghi *et al.*, "Modelling the ethical priorities influencing decision-making in cybersecurity contexts," *Organizational Cybersecurity Journal: Practice, Process and People*, vol. 3, no. 2, pp. 127–149, Jan. 2023, doi: 10.1108/OCJ-09-2022-0015.

[8] S. R. Sindiramutty *et al.*, "Ethical Considerations in Drone Cybersecurity," in *Cybersecurity Issues and Challenges in the Drone Industry*, IGI Global, 2024, pp. 42–87. doi: 10.4018/979-8-3693-0774-8.ch003.

[9] S. Boopathi and A. Khang, "AI-Integrated Technology for a Secure and Ethical Healthcare Ecosystem," in *AI and IoT-Based Technologies for Precision Medicine*, IGI Global, 2023, pp. 36–59. doi: 10.4018/979-8-3693-0876-9.ch003.

[10] U. Hani, O. Sohaib, K. Khan, A. Aleidi, and N. Islam, "Psychological profiling of hackers via machine learning toward sustainable cybersecurity," *Front. Comput. Sci.*, vol. 6, Apr. 2024, doi: 10.3389/fcomp.2024.1381351.

[11] O. A. Lottu *et al.*, "Towards a conceptual framework for ethical AI development in IT systems," *World Journal of Advanced Research and Reviews*, vol. 21, no. 3, Art. no. 3, 2024, doi: 10.30574/wjarr.2024.21.3.0735.

[12] H. H. Tseng, "The Development of the Personal and Professional Values-Integrated Framework as an Aid to Ethical Decision Making," *American Journal of Educational Research*.

[13] S. S. Bush, "Use of practice guidelines and position statements in ethical decision making," *American Psychologist*, vol. 74, no. 9, pp. 1151–1162, 2019, doi: 10.1037/amp0000519.

[14] B. Biggio, B. Nelson, and P. Laskov, "Poisoning Attacks against Support Vector Machines," Mar. 25, 2013, *arXiv*: arXiv:1206.6389. doi: 10.48550/arXiv.1206.6389.

[15] T. Gu, K. Liu, B. Dolan-Gavitt, and S. Garg, "BadNets: Evaluating Backdooring Attacks on Deep Neural Networks," *IEEE Access*, vol. 7, pp. 47230–47244, 2019, doi: 10.1109/ACCESS.2019.2909068.

[16] Y. Liu *et al.*, "Trojaning Attack on Neural Networks," in *Proceedings 2018 Network and Distributed System Security Symposium*, San Diego, CA: Internet Society, 2018. doi: 10.14722/ndss.2018.23291.

[17] B. Biggio *et al.*, “Evasion Attacks against Machine Learning at Test Time,” in *Machine Learning and Knowledge Discovery in Databases*, H. Blockeel, K. Kersting, S. Nijssen, and F. Železný, Eds., Berlin, Heidelberg: Springer, 2013, pp. 387–402. doi: 10.1007/978-3-642-40994-3_25.

[18] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and Harnessing Adversarial Examples,” Mar. 20, 2015, *arXiv*: arXiv:1412.6572. doi: 10.48550/arXiv.1412.6572.

[19] C. Szegedy *et al.*, “Intriguing properties of neural networks,” Feb. 19, 2014, *arXiv*: arXiv:1312.6199. doi: 10.48550/arXiv.1312.6199.

[20] I. Dinur and K. Nissim, “Revealing information while preserving privacy,” in *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, in PODS ’03. New York, NY, USA: Association for Computing Machinery, Jun. 2003, pp. 202–210. doi: 10.1145/773153.773173.

[21] M. L. Hernandez-Jaimes, A. Martinez-Cruz, K. A. Ramírez-Gutiérrez, and C. Feregrino-Uribe, “Artificial intelligence for IoMT security: A review of intrusion detection systems, attacks, datasets and Cloud–Fog–Edge architectures,” *Internet of Things*, vol. 23, p. 100887, Oct. 2023, doi: 10.1016/j.iot.2023.100887.

[22] *Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act) (Text with EEA relevance)*, vol. 151. 2019. Accessed: May 08, 2024. [Online]. Available: http://data.europa.eu/eli/reg/2019/881/oj/eng

[23] P. Formosa, M. Wilson, and D. Richards, “A principlist framework for cybersecurity ethics,” *Computers & Security*, vol. 109, p. 102382, Oct. 2021, doi: 10.1016/j.cose.2021.102382.

[24] K. Weber and N. Kleine, “Cybersecurity in Health Care,” in *The Ethics of Cybersecurity*, M. Christen, B. Gordijn, and M. Loi, Eds., Cham: Springer International Publishing, 2020, pp. 139–156. doi: 10.1007/978-3-030-29053-5_7.

[25] M. Christen, B. Gordijn, K. Weber, I. van de Poel, and E. Yaghmaei, “A Review of Value-Conflicts in Cybersecurity: An assessment based on quantitative and qualitative literature analysis,” *The ORBIT Journal*, vol. 1, no. 1, pp. 1–19, Jan. 2017, doi: 10.29297/orbit.v1i1.28.

[26] D. Reidsma, J. Van Der Ham, and A. Continella, “Operationalizing Cybersecurity Research Ethics Review: From Principles and Guidelines to Practice,” in *Proceedings of the 2nd International Workshop on Binary Analysis Research*, San Diego, CA, USA: Internet Society, 2023. doi: 10.14722/ethics.2023.237352.

[27] M. Christen, B. Gordijn, and M. Loi, Eds., *The Ethics of Cybersecurity*, vol. 21. in The International Library of Ethics, Law and Technology, vol. 21. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-29053-5.

[28] “ethics and data protection - Cerca con Google.” Accessed: Jun. 04, 2024. [Online]. Available: https://www.google.it/search?q=ethics+and+data+protection&client=safari&sca_esv=814dda77967cd28b&channel=iphone_bm&source=hp&ei=IANfZpr2PI-TkdUPq6mPgAI&iflsig=AL9hbdgAAAAAZl8RMTVD1IQa2kx63KKdk1dxdkeSGeil&ved=0ahUKEwiahvmr9MGGAxWPSaQEHavUAyAQ4dUDCAw&uact=5&oq=ethics+and+data+protection&gs_lp=Egdnd3Mtd2l6IhpldGhpY3MgYW5kIGRhdGEgcHJvdGVjdGlvbjIHEAAYgAQYEzIHEAAYgAQYEzIIEAAYExgWGB4yCBAAGBMYFhgeMggQABgTGBYYHjIIEAAYExgWGB4yCBAAGBMYFhgeMggQABgTGBYYHjIIEAAYExgWGB5IkCRQAFiUInAAeACQAQCYAV2gAa8MqgECMja4AQPIAQD4AQGYAhqgArUNwgIFEAAYgATCAg4QLhiABBjHARiOBRivAcICCxAuGIAEGNEDGMcBwgIFEC4YgATCAgsQLhiABBjHARivAcICDRAuGIAEGNEDGMcBGArCAgcQLhiABBgTwgIIEAAYgAQYogSYAwCSBwQyNS4xoAeJ5AE&sclient=gws-wiz

[29] E. Schlehahn, “Cybersecurity and the State,” in *The Ethics of Cybersecurity*, Springer, Cham, 2020, pp. 205–225. doi: 10.1007/978-3-030-29053-5_10.

[30] M. Gharib, P. Giorgini, and J. Mylopoulos, “An Ontology for Privacy Requirements via a Systematic Literature Review,” *J Data Semant*, vol. 9, no. 4, pp. 123–149, Dec. 2020, doi: 10.1007/s13740-020-00116-5.

[31] W. Labda, N. Mehandjiev, and P. Sampaio, “Modeling of privacy-aware business processes in BPMN to protect personal data,” in *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, Gyeongju Republic of Korea: ACM, Mar. 2014, pp. 1399–1405. doi: 10.1145/2554850.2555014.

[32] “Regulation - 2016/679 - EN - gdpr - EUR-Lex.” Accessed: Jun. 06, 2024. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[33] S. Tamjidi and A. Shameli-Sendi, “Intelligence in security countermeasures selection,” *J Comput Virol Hack Tech*, vol. 19, no. 1, pp. 137–148, Mar. 2023, doi: 10.1007/s11416-022-00439-w.

[34] C. J. Hernández-Castro, Z. Liu, A. Serban, I. Tsingenopoulos, and W. Joosen, “Adversarial Machine Learning,” in *Security and Artificial Intelligence: A Crossdisciplinary Approach*, L. Batina, T. Bäck, I. Buhan, and S. Picek, Eds., Cham: Springer International Publishing, 2022, pp. 287–312. doi: 10.1007/978-3-030-98795-4_12.

[35] C. Mehta, P. Harniya, and S. Kamat, “Comprehending and Detecting Vulnerabilities using Adversarial Machine Learning Attacks,” in *2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP)*, Vijayawada, India: IEEE, Feb. 2022, pp. 1–5. doi: 10.1109/AISP53593.2022.9760580.

[36] I. T. L. Computer Security Division, “NIST Releases 2023 Edition of Adversarial Machine Learning Report | CSRC,” CSRC | NIST. Accessed: Jun. 06, 2024. [Online]. Available: https://csrc.nist.rip/News/2024/nist-releases-adversarial-ml-taxonomy-terminology

[37] A. Vassilev, A. Oprea, A. Fordyce, and H. Anderson, “Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations,” National Institute of Standards and Technology, NIST Artificial Intelligence (AI) 100-2 E2023, Jan. 2024. doi: 10.6028/NIST.AI.100-2e2023.

[38] G. Rafaiani, G. Barchiesi, L. Ilari, M. Baldi, and B. Giovanola, “A Quantitative Model for the Assessment of Ethics Risks in Information Technology,” in *2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)*, May 2023, pp. 01–08. doi: 10.1109/ETHICS57328.2023.10155002.

[39] M. Battaglioni, G. Rafaiani, F. Chiaraluce, and M. Baldi, “MAGIC: A Method for Assessing Cyber Incidents Occurrence,” *IEEE Access*, vol. 10, pp. 73458–73473, 2022, doi: 10.1109/ACCESS.2022.3189777.

[40] “CIS Critical Security Controls Version 8.” Accessed: Oct. 13, 2024. [Online]. Available: https://www.cisecurity.org/controls/v8

[41] L. Ilari, G. Rafaiani, M. Baldi, and B. Giovanola, “Ethical Biases in Machine Learning-based Filtering of Internet Communications,” in *2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)*, May 2023, pp. 01–09. doi: 10.1109/ETHICS57328.2023.10154975.

[42] “Automated moderation tool from Google rates People of Color and gays as ‘toxic,’” AlgorithmWatch. Accessed: Oct. 29, 2024. [Online]. Available: https://algorithmwatch.org/en/automated-moderation-perspective-bias/

[43] S. Zannettou, M. Elsherief, E. Belding, S. Nilizadeh, and G. Stringhini, “Measuring and Characterizing Hate Speech on News Websites,” in *Proceedings of the 12th ACM Conference on Web Science*, in WebSci ’20. New York, NY, USA: Association for Computing Machinery, Jul. 2020, pp. 125–134. doi: 10.1145/3394231.3397902.

[44] “About the API - Attributes and Languages.” Accessed: Oct. 29, 2024. [Online]. Available: https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages?language=en_US

[45] B. Giovanola and S. Tiribelli, “Weapons of moral construction? On the value of fairness in algorithmic decision-making,” *Ethics Inf Technol*, vol. 24, no. 1, p. 3, Jan. 2022, doi: 10.1007/s10676-022-09622-5.

[46] R. G. Leema, R. Rajmohan, S. Usharani, K. Kiruba, and P. Manjubala, “Fundamentals of Blockchain and Distributed Ledger Technology (DLT),” in *Recent Trends in Blockchain for Information Systems Security and Privacy*, CRC Press, 2021.

[47] J. D. Roberts, J. F. Defranco, and D. R. Kuhn, “Data Block Matrix and Hyperledger Implementation: Extending Distributed Ledger Technology for Privacy Requirements,” *Distrib. Ledger Technol.*, vol. 2, no. 2, p. 16:1-16:11, Jun. 2023, doi: 10.1145/3585539.

[48] M. K. S. Suripeddi and P. Purandare, “Blockchain and GDPR – A Study on Compatibility Issues of the Distributed Ledger Technology with GDPR Data Processing,” *J. Phys.: Conf. Ser.*, vol. 1964, no. 4, p. 042005, Jul. 2021, doi: 10.1088/1742-6596/1964/4/042005.

[49] P. C. Franks, “Implications of blockchain distributed ledger technology for records management and information governance programs,” *Records Management Journal*, vol. 30, no. 3, pp. 287–299, Jan. 2020, doi: 10.1108/RMJ-08-2019-0047.

[50] H. Yaacob, “Legal Issues In Distributed Ledger Technology (DLT) & Blockchain In Brunei Darussalam,” *iEco | Islamic Economics Journal*, vol. 1, no. 1, Art. no. 1, Jan. 2021, doi: 10.59202/ieco.v1i1.390.

[51] S. Tyagi and M. Kathuria, “Role of Zero-Knowledge Proof in Blockchain Security,” in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, May 2022, pp. 738–743. doi: 10.1109/COM-IT-CON54601.2022.9850714.

[52] R. Swami, M. Dave, V. Ranga, N. Tripathi, A. K. Shaji, and A. Sharma, “Towards Utilizing Blockchain for Countering Distributed Denial-of-Service (DDoS),” in *Revolutionary Applications of Blockchain-Enabled Privacy and Access Control*, IGI Global, 2021, pp. 35–58. doi: 10.4018/978-1-7998-7589-5.ch002.

[53] Research Scholar, Gujarat Vidyapith, Ahmedabad, India. and D. Patel*, “Blockchain Technology towards the Mitigation of Distributed Denial of Service Attacks,” *IJRTE*, vol. 8, no. 6, pp. 961–965, Mar. 2020, doi: 10.35940/ijrte.F7420.038620.

[54] D. Patel and D. Patel, “Collaborative Blockchain Based Distributed Denial of Service Attack Mitigation Approach with IP Reputation System,” in *Database Systems for Advanced Applications. DASFAA 2022 International Workshops*, U. K. Rage, V. Goyal, and P. K. Reddy, Eds., Cham: Springer International Publishing, 2022, pp. 91–103. doi: 10.1007/978-3-031-11217-1_7.

[55] A. Panwar and V. Bhatnagar, “Distributed Ledger Technology (DLT): The Beginning of a Technological Revolution for Blockchain,” in *2nd International Conference on Data, Engineering and Applications (IDEA)*, Feb. 2020, pp. 1–5. doi: 10.1109/IDEA49133.2020.9170699.

[56] S. Wani, M. Imthiyas, H. Almohamedh, K. M. Alhamed, S. Almotairi, and Y. Gulzar, “Distributed Denial of Service (DDoS) Mitigation Using Blockchain—A Comprehensive Insight,” *Symmetry*, vol. 13, no. 2, Art. no. 2, Feb. 2021, doi: 10.3390/sym13020227.

[57] R. Singh, S. Tanwar, and T. P. Sharma, “Utilization of blockchain for mitigating the distributed denial of service attacks,” *SECURITY AND PRIVACY*, vol. 3, no. 3, p. e96, 2020, doi: 10.1002/spy2.96.

[58] I.-S. Lei, S.-K. Tang, and R. Tse, “Integrating consortium blockchain into edge server to defense against ransomware attack,” *Procedia Computer Science*, vol. 177, pp. 120–127, Jan. 2020, doi: 10.1016/j.procs.2020.10.019.

[59] N. Thamer and R. Alubady, “A Survey of Ransomware Attacks for Healthcare Systems: Risks, Challenges, Solutions and Opportunity of Research,” in *2021 1st Babylon International*

Conference on Information Technology and Science (BICITS), Apr. 2021, pp. 210–216. doi: 10.1109/BICITS51482.2021.9509877.

[60] N. Thamer and R. Alubady, "Security against Ransomware Attack in Medical Healthcare Records Using Blockchain Technology," in *2022 Fifth College of Science International Conference of Recent Trends in Information Technology (CSCTIT)*, Nov. 2022, pp. 111–117. doi: 10.1109/CSCTIT56299.2022.10145717.

[61] M. Wazid, A. Kumar Das, and S. Shetty, "BSFR-SH: Blockchain-Enabled Security Framework Against Ransomware Attacks for Smart Healthcare," *IEEE Transactions on Consumer Electronics*, vol. 69, no. 1, pp. 18–28, Feb. 2023, doi: 10.1109/TCE.2022.3208795.

[62] E.-E. Gojka, N. Kannengiesser, B. Sturm, J. Bartsch, and A. Sunyaev, "Security in Distributed Ledger Technology: An Analysis of Vulnerabilities and Attack Vectors," 2021, pp. 722–742. doi: 10.1007/978-3-030-80129-8_50.

[63] J. E. T. Akinsola, M. A. Adeagbo, S. A. Akinseinde, F. O. Onipede, and A. A. Yusuf, "Applications of Blockchain Technology in Cyber Attacks Prevention," in *Sustainable and Advanced Applications of Blockchain in Smart Computational Technologies*, Chapman and Hall/CRC, 2022.

[64] J. Kaur, "A Secure and Smart Framework for Preventing Ransomware Attack," Jan. 20, 2020, *arXiv*: arXiv:2001.07179. doi: 10.48550/arXiv.2001.07179.

[65] A. U. Nwosu and S. B. Goyal, *Blockchain Transforming Cyber-attacks: Healthcare Industry*. 2020.

[66] A. Sunyaev, "Distributed Ledger Technology," in *Internet Computing: Principles of Distributed Systems and Emerging Internet-Based Technologies*, A. Sunyaev, Ed., Cham: Springer International Publishing, 2020, pp. 265–299. doi: 10.1007/978-3-030-34957-8_9.

[67] R. G. Leema, R. Rajmohan, S. Usharani, K. Kiruba, and P. Manjubala, "Fundamentals of Blockchain and Distributed Ledger Technology (DLT)," in *Recent Trends in Blockchain for Information Systems Security and Privacy*, CRC Press, 2021.

[68] S. Shrivastava and A. Sharma, "Distributed Ledger Technology (DLT) and Byzantine Fault Tolerance in Blockchain," in *Soft Computing: Theories and Applications*, R. Kumar, C. W. Ahn, T. K. Sharma, O. P. Verma, and A. Agarwal, Eds., Singapore: Springer Nature, 2022, pp. 971–981. doi: 10.1007/978-981-19-0707-4_86.

[69] E. Baninemeh, S. Jansen, and K. Labunets, "A Security Risk Assessment Method for Distributed Ledger Technology (DLT) based Applications: Three Industry Case Studies," Jan. 22, 2024, *arXiv*: arXiv:2401.12358. doi: 10.48550/arXiv.2401.12358.

[70] U. Cali, O. Elma, and R. Reddi, "Cybersecurity of Renewable Energy Data and Applications Using Distributed Ledger Technology".

[71] S. Gutlapalli, "Commercial Applications of Blockchain and Distributed Ledger Technology," *Engineering International*, vol. 4, pp. 89–94, Dec. 2016, doi: 10.18034/ei.v4i2.653.

[72] D. M. West and J. R. Allen, *Turning Point: Policymaking in the Era of Artificial Intelligence*. Brookings Institution Press, 2020. Accessed: Jun. 20, 2024. [Online]. Available: https://www.jstor.org/stable/10.7864/j.ctvwh8fcb

[73] "Cosa intendiamo per dati personali?" Accessed: Jun. 20, 2024. [Online]. Available: https://www.garanteprivacy.it/home/diritti/cosa-intendiamo-per-dati-personali

[74] Y. Gong and C. Poellabauer, *Protecting Voice Controlled Systems Using Sound Source Identification Based on Acoustic Cues*. 2018. doi: 10.1109/ICCCN.2018.8487334.

[75] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards Deep Learning Models Resistant to Adversarial Attacks," Sep. 04, 2019, *arXiv*: arXiv:1706.06083. Accessed: Oct. 03, 2024. [Online]. Available: http://arxiv.org/abs/1706.06083

[76] D. Tsipras, S. Santurkar, L. Engstrom, A. Turner, and A. Madry, "Robustness May Be at Odds with Accuracy," Sep. 09, 2019, *arXiv*: arXiv:1805.12152. doi: 10.48550/arXiv.1805.12152.

[77] M. Lecuyer, V. Atlidakis, R. Geambasu, D. Hsu, and S. Jana, “Certified Robustness to Adversarial Examples with Differential Privacy,” in *2019 IEEE Symposium on Security and Privacy (SP)*, May 2019, pp. 656–672. doi: 10.1109/SP.2019.00044.

[78] J. Cohen, E. Rosenfeld, and Z. Kolter, “Certified Adversarial Robustness via Randomized Smoothing,” in *Proceedings of the 36th International Conference on Machine Learning*, PMLR, May 2019, pp. 1310–1320. Accessed: Oct. 03, 2024. [Online]. Available: https://proceedings.mlr.press/v97/cohen19c.html

[79] W. Zhao, S. Alwidian, and Q. H. Mahmoud, “Adversarial Training Methods for Deep Learning: A Systematic Review,” *Algorithms*, vol. 15, no. 8, Art. no. 8, Aug. 2022, doi: 10.3390/a15080283.

[80] B. L. Ljubljana University of, “Orange Fairness - Adversarial Debiasing.” Accessed: Oct. 03, 2024. [Online]. Available: https://oldorange.biolab.si/blog/2023/2023-09-19-fairness-adversarial-debiasing/

[81] E. Perrier, “Ethical Quantum Computing: A Roadmap,” Apr. 20, 2022, *arXiv*: arXiv:2102.00759. Accessed: Jul. 29, 2024. [Online]. Available: http://arxiv.org/abs/2102.00759

[82] E. Grumbling and M. Horowitz, Eds., *Quantum Computing: Progress and Prospects*. Washington, D.C.: National Academies Press, 2019. doi: 10.17226/25196.

[83] D. Elliott and E. Soifer, “AI Technologies, Privacy, and Security,” *Front. Artif. Intell.*, vol. 5, Apr. 2022, doi: 10.3389/frai.2022.826737.

[84] S. Kundu and S. Ghosh, “Security Aspects of Quantum Machine Learning: Opportunities, Threats and Defenses,” in *Proceedings of the Great Lakes Symposium on VLSI 2022*, Jun. 2022, pp. 463–468. doi: 10.1145/3526241.3530833.

[85] N. Franco *et al.*, “Predominant Aspects on Security for Quantum Machine Learning: Literature Review,” Apr. 19, 2024, *arXiv*: arXiv:2401.07774. doi: 10.48550/arXiv.2401.07774.

[86] G. Alagic, Y. Dulek, C. Schaffner, and F. Speelman, “Quantum Fully Homomorphic Encryption with Verification,” in *Advances in Cryptology – ASIACRYPT 2017*, T. Takagi and T. Peyrin, Eds., Cham: Springer International Publishing, 2017, pp. 438–467. doi: 10.1007/978-3-319-70694-8_16.

[87] R. Deviani, “The Application of Fully Homomorphic Encryption on XGBoost Based Multiclass Classification,” *JIEET (Journal of Information Engineering and Educational Technology)*, vol. 7, no. 1, pp. 49–58, Jun. 2023, doi: 10.26740/jieet.v7n1.p49-58.

[88] A. Maqousi, M. Alauthman, and A. Almomani, “Homomorphic Encryption Enabling Computation on Encrypted Data for Secure Cloud Computing,” in *Innovations in Modern Cryptography*, IGI Global, 2024, pp. 215–240. doi: 10.4018/979-8-3693-5330-1.ch009.

[89] N. J. Hernandez Marcano, M. Moller, S. Hansen, and R. H. Jacobsen, “On Fully Homomorphic Encryption for Privacy-Preserving Deep Learning,” in *2019 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2019, pp. 1–6. doi: 10.1109/GCWkshps45667.2019.9024625.

[90] P. K. Yeng, B. Yang, M. A. Fauzi, and P. Nimbe, “A framework for exploring incentive methods towards reducing phishing susceptibility in Healthcare: Based on a review and In-the-wild-field study approach,” in *2023 Intelligent Methods, Systems, and Applications (IMSA)*, Jul. 2023, pp. 228–234. doi: 10.1109/IMSA58542.2023.10217499.

[91] J. B. Wright and D. N. Burrell, “Cybersecurity Leadership Ethics in Healthcare,” in *Handbook of Research on Cybersecurity Risk in Contemporary Business Systems*, IGI Global, 2023, pp. 137–148. doi: 10.4018/978-1-6684-7207-1.ch007.

[92] F. Tronnier, S. Pape, S. Löbner, and K. Rannenberg, “A Discussion on Ethical Cybersecurity Issues in Digital Service Chains,” in *Cybersecurity of Digital Service Chains: Challenges, Methodologies, and Tools*, J. Kołodziej, M. Repetto, and A. Duzha, Eds., Cham: Springer International Publishing, 2022, pp. 222–256. doi: 10.1007/978-3-031-04036-8_10.

[93] L. Thomas, I. Gondal, T. Oseni, and S. (Sally) Firmin, “A framework for data privacy and security accountability in data breach communications,” *Computers & Security*, vol. 116, p. 102657, May 2022, doi: 10.1016/j.cose.2022.102657.

[94] B. Sadeghi, D. Richards, P. Formosa, M. McEwan, and M. Hitchens, “How to increase ethical awareness in cybersecurity decision-making,” *ACIS 2022 Proceedings*, Dec. 2022, [Online]. Available: https://aisel.aisnet.org/acis2022/28

[95] J. Rajamäki and A. Hummelholm, “Ethical Resilience Management Framework for Critical Healthcare Information Infrastructure,” 61681. Accessed: May 06, 2024. [Online]. Available: http://www.theseus.fi/handle/10024/745453

[96] I. Priyadarshini and C. Cotton, *Cybersecurity: Ethics, Legal, Risks, and Policies*. New York: Apple Academic Press, 2022. doi: 10.1201/9781003187127.

[97] A. Oruc, “Ethical Considerations in Maritime Cybersecurity Research,” *Ethical Considerations in Maritime Cybersecurity Research*, 2022, doi: 10.12716/1001.16.02.14.

[98] M. I. Maratsi, O. Popov, C. Alexopoulos, and Y. Charalabidis, “Ethical and Legal Aspects of Digital Forensics Algorithms: The Case of Digital Evidence Acquisition,” in *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*, in ICEGOV ’22. New York, NY, USA: Association for Computing Machinery, Nov. 2022, pp. 32–40. doi: 10.1145/3560107.3560114.

[99] G. Lorenzini, D. M. Shaw, and B. S. Elger, “It takes a pirate to know one: ethical hackers for healthcare cybersecurity,” *BMC Med Ethics*, vol. 23, no. 1, p. 131, Dec. 2022, doi: 10.1186/s12910-022-00872-y.

[100] D. Kozhuharova, A. Kirov, and Z. Al-Shargabi, “Ethics in Cybersecurity. What Are the Challenges We Need to Be Aware of and How to Handle Them?,” in *Cybersecurity of Digital Service Chains: Challenges, Methodologies, and Tools*, J. Kołodziej, M. Repetto, and A. Duzha, Eds., Cham: Springer International Publishing, 2022, pp. 202–221. doi: 10.1007/978-3-031-04036-8_9.

[101] C. DeCusatis *et al.*, “A Cybersecurity Awareness Escape Room using Gamification Design Principles,” in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan. 2022, pp. 0765–0770. doi: 10.1109/CCWC54503.2022.9720748.

[102] A. Carlo *et al.*, “‘Understanding Space Vulnerabilities: Developing Technical and Legal Frameworks for AI and Cybersecurity in Space,’” vol. 2022-September, Accessed: May 06, 2024. [Online]. Available: https://avesis.istanbul.edu.tr/yayin/30e9366e-7d7f-4b35-9ee1-7dc038dd55ab/understanding-space-vulnerabilities-developing-technical-and-legal-frameworks-for-ai-and-cybersecurity-in-space

[103] J. Breig and D. Westhoff, “Short Paper: Debating Ethics with Cybersecurity Students,” presented at the GI SICHERHEIT 2022, Gesellschaft für Informatik, Bonn, 2022, pp. 183–192. Accessed: May 06, 2024. [Online]. Available: https://dl.gi.de/handle/20.500.12116/40133

[104] M. Botes and G. Lenzini, “When Cryptographic Ransomware Poses Cyber Threats: Ethical Challenges and Proposed Safeguards for Cybersecurity Researchers,” in *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, Jun. 2022, pp. 562–568. doi: 10.1109/EuroSPW55150.2022.00067.

[105] R. Baskerville, P. Depaoli, and P. Spagnoletti, “Organizing Cybersecurity in Action: A Pragmatic Ethical Reasoning Approach,” in *Organizing in a Digitized World*, S. Za, A. Consorti, and F. Virili, Eds., Cham: Springer International Publishing, 2022, pp. 190–203. doi: 10.1007/978-3-030-86858-1_11.

[106] G. L. Wallace, “Ethics in Technology: Heroes Assemble!,” in *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, Dec. 2021, pp. 1085–1090. doi: 10.1109/CSCI54926.2021.00229.

[107] M. J. Santofimia, F. J. Villanueva, E. Litvinov, I. Viksnin, A. Fernandes, and J. C. Lopez, “Cybersecurity in Active and Healthy Ageing Era,” *Procedia Computer Science*, vol. 192, pp. 2068–2076, Jan. 2021, doi: 10.1016/j.procs.2021.08.214.

[108] M. Dupuis and K. Renaud, “Scoping the ethical principles of cybersecurity fear appeals,” *Ethics Inf Technol*, vol. 23, no. 3, pp. 265–284, Sep. 2021, doi: 10.1007/s10676-020-09560-0.

[109] A. Poulsen, E. Fosch-Villaronga, and O. K. Burmeister, “Cybersecurity, value sensing robots for LGBTIQ+ elderly, and the need for revised codes of conduct,” *Australasian Journal of Information Systems*, vol. 24, Jun. 2020, doi: 10.3127/ajis.v24i0.2789.

[110] K. Nikolskaia and V. Naumov, “Ethical and Legal Principles of Publishing Open Source Dual-Purpose Machine Learning Algorithms,” in *2020 International Conference Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS)*, Sep. 2020, pp. 56–58. doi: 10.1109/ITQMIS51053.2020.9322897.

[111] G. Lucas, “Cybersecurity and Cyber Warfare: The Ethical Paradox of ‘Universal Diffidence,’” in *The Ethics of Cybersecurity*, M. Christen, B. Gordijn, and M. Loi, Eds., Cham: Springer International Publishing, 2020, pp. 245–258. doi: 10.1007/978-3-030-29053-5_12.

[112] M. Loi and M. Christen, “Ethical Frameworks for Cybersecurity,” in *The Ethics of Cybersecurity*, M. Christen, B. Gordijn, and M. Loi, Eds., Cham: Springer International Publishing, 2020, pp. 73–95. doi: 10.1007/978-3-030-29053-5_4.

[113] D. Herrmann and H. Pridöhl, “Basic Concepts and Models of Cybersecurity,” in *The Ethics of Cybersecurity*, M. Christen, B. Gordijn, and M. Loi, Eds., Cham: Springer International Publishing, 2020, pp. 11–44. doi: 10.1007/978-3-030-29053-5_2.

[114] J. Domingo-Ferrer and A. Blanco-Justicia, “Ethical Value-Centric Cybersecurity: A Methodology Based on a Value Graph,” *Sci Eng Ethics*, vol. 26, no. 3, pp. 1267–1285, Jun. 2020, doi: 10.1007/s11948-019-00138-8.

[115] R. Spinello, “Ethics in Cyberspace: Freedom, Rights, and Cybersecurity,” 2019, pp. 444–458. doi: 10.1017/9781108616188.029.

[116] M. Loi, M. Christen, N. Kleine, and K. Weber, “Cybersecurity in health – disentangling value tensions,” *Journal of Information, Communication and Ethics in Society*, vol. 17, no. 2, pp. 229–245, Jan. 2019, doi: 10.1108/JICES-12-2018-0095.

[117] M. Etschmaier, “A Purposeful Systems Design Approach for Cybersecurity,” in *EPiC Series in Computing*, EasyChair, Sep. 2019, pp. 90–100. doi: 10.29007/fq42.

[118] C. T. Holzer and J. E. Lerums, “The ethics of hacking back,” in *2016 IEEE Symposium on Technologies for Homeland Security (HST)*, May 2016, pp. 1–6. doi: 10.1109/THS.2016.7568877.

[119] R. Dipert, “The Ethics of Cyberwarfare,” *Journal of Military Ethics*, vol. 9, Dec. 2010, doi: 10.1080/15027570.2010.536404.

[120] D. Shou, “Ethical Considerations of Sharing Data for Cybersecurity Research,” in *Financial Cryptography and Data Security*, G. Danezis, S. Dietrich, and K. Sako, Eds., Berlin, Heidelberg: Springer, 2012, pp. 169–177. doi: 10.1007/978-3-642-29889-9_15.

[121] G. Faiella *et al.*, “Building an Ethical Framework for Cross-Border Applications: The KONFIDO Project,” in *Security in Computer and Information Sciences*, E. Gelenbe, P. Campegiani, T. Czachórski, S. K. Katsikas, I. Komnios, L. Romano, and D. Tzovaras, Eds., Cham: Springer International Publishing, 2018, pp. 38–45. doi: 10.1007/978-3-319-95189-8_4.

[122] S. B. Nazeer Khan, D. Richards, P. Formosa, and S. Bankins, “To breach or not? Profiling students’ likelihood of breaching university ICT Codes of Conduct: Student Profiling of Breach of ICT Codes of Conduct,” in *Proceedings of the 2023 Australasian Computer Science Week*, in ACSW ’23. New York, NY, USA: Association for Computing Machinery, Mar. 2023, pp. 50–57. doi: 10.1145/3579375.3579382.

[123] C. Posey and R. Folger, “An exploratory examination of organizational insiders’ descriptive and normative perceptions of cyber-relevant rights and responsibilities,” *Computers & Security*, vol. 99, p. 102038, Dec. 2020, doi: 10.1016/j.cose.2020.102038.
[124] M. L. Hernandez-Jaimes, A. Martinez-Cruz, K. A. Ramírez Gutiérrez, and C. Feregrino, “Artificial intelligence for IoMT security: A review of intrusion detection systems, attacks, datasets and Cloud-Fog-Edge architectures,” *Internet of Things*, vol. 23, pp. 1–82, Aug. 2023, doi: 10.1016/j.iot.2023.100887.
[125] “Glossary,” ENISA. Accessed: May 08, 2024. [Online]. Available: https://www.enisa.europa.eu/topics/risk-management/current-risk/risk-management-inventory/glossary
[126] “Create a holistic approach to data protection | IBM.” Accessed: May 08, 2024. [Online]. Available: https://www.ibm.com/resources/the-data-differentiator/data-protection-strategy
[127] “https://www.datenschutz- mv.de/static/DS/Dateien/Datenschutzmodell/SDMV 3en.pdf - Cerca con Google.” Accessed: Jun. 06, 2024. [Online]. Available: https://www.google.it/search?q=https%3A%2F%2Fwww.datenschutz-+mv.de%2Fstatic%2FDS%2FDateien%2FDatenschutzmodell%2FSDMV+3en.pdf&client=safari&sca_esv=9a83bf99618973ac&channel=iphone_bm&source=hp&ei=PoBhZoqMIMesxc8PrObcwQk&iflsig=AL9hbdgAAAAAZmGOTgO-K21n6V4BR0tIZ0eSNerVY5xK&ved=0ahUKEwiKo-j408aGAxVHVvEDHSwzN5gQ4dUDCA4&uact=5&oq=https%3A%2F%2Fwww.datenschutz-+mv.de%2Fstatic%2FDS%2FDateien%2FDatenschutzmodell%2FSDMV+3en.pdf&gs_lp=Egdnd3Mtd2l6Ik9odHRwczovL3d3dy5kYXRlbnNjaHV0ei0gbXYuZGUvc3RhdGljL0RTL0RhdGVpZW4vRGF0ZW5zY2h1dHptb2RlbGwvU0RNViAzZW4ucGRmSABQAFgAcAB4AJABAJgBAKABAKoBALgBA8gBAPgBAvgBAZgCAKACAJgDAJIHAKAHAA&sclient=gws-wiz
[128] W. H. O. R. O. for Europe, “The protection of personal data in health information systems-principles and processes for public health,” Art. no. WHO/EURO:2021-1994-41749-57154, 2021, Accessed: Jun. 06, 2024. [Online]. Available: https://iris.who.int/handle/10665/341374
[129] “GDPR - Decision tree | reviewed.” Accessed: Oct. 10, 2024. [Online]. Available: https://ec.europa.eu/assets/rtd/ethics-data-protection-decision-tree/index.html

A. Appendix: figures and tables on ethics, cybersecurity, and data protection

This appendix provides a series of figures and tables that serve to expand on key ethical, cybersecurity, and data protection concepts discussed throughout the text. **Figure A1** illustrates the European Commission's guidelines on ethics and data protection, establishing a foundational understanding of regulatory frameworks. Further figures, such as **Figures A2–A10**, present examples of qualitative and semi-quantitative methods, showcasing self-assessment models for evaluating trustworthy artificial intelligence.

The appendix also includes detailed GDP (General Data Protection) registers across **Figures A11–A22**, which illustrate stages of data processing, risk assessment structures, and mitigation measures, along with comprehensive calculations for threat probabilities and necessary evaluations for Data Protection Impact Assessments (DPIA).

Complementary tables provide data on cyber-attacks (**Table A1**), extracted ethical principles from key sources (**Table A2**), and examples from key studies (**Tables A3–A9**), offering a diverse perspective on current ethical standards and approaches in cybersecurity. Additional tables provide ethical principles in data protection (**Table A13-A14**), map GDPR principles with other models (**Table A15**), and detail DPIA criteria (**Tables A16–A23**), rounding out a structured view on assessing ethical impacts in data handling. Finally, the tables illustrate the effectiveness of various risk assessment methods, as presented in the European ethics issues table and in self-assessment tools (**Tables A24–A25**).

Table A-1. Different cyber attacks

Classical cyber risk	Cyber risk in classical predictive AI

Malware
• Ransomware
• Adware
• Spyware
• Rootkit
• Virus
• Trojan
	Training Stage
	• Poisoning attacks
	  ○ Data poisoning
	  ○ Model poisoning
	Deployment Stage
	• Evasion attacks

• Worm
• Exploit
• Cryptominer
• Keylogger
	• Privacy attacks
	  ○ Data privacy
	  ○ Model privacy

Man in the Middle • Wi-fi Eavesdropping • DNS Spoofing • HTTPS Spoofing • ARP Spoofing • E-mail Hacking • Session Hacking • SSL Stripping • MITB attack	
Dos/DDos	
Brute Force	
0-Day Exploit	
Social Engineering • Phishing / Spear-Phishing /Whaling • Vishing • Smishing • Baiting • Catfishing • Pretexting • Scareware • Tailgating, piggybacking • Water Holing • Quid Pro Quo	
Web app	

• Injection attacks (SQL injection) • Authentication • Sens. Data Exposure • XML External Entities • Broken Access Control • Security Misconfiguration • Cross Cite Scripting • Insecure Deserialization • Vulnerable Components • Insufficient Logging&Monitoring	

Table A-2. Ethical principles extracted from the selected papers.

Source title	**Source**	**Authors**	**Title**	**Origin of principles (approach)**	**Principles**
1st International Conference of Intelligent Methods, Systems and Applications, IMSA 2023	Scopus	Yeng P.K et Wright J.B (2023) [90]	A framework for exploring incentive methods towards reducing phishing susceptibility in Healthcare: Based on a review and In-the-wild-field study approach	Issues form case studies **(Bottom-up)**	Deception vs informed content, potential harm to participants, infringement upon privacy, violation of participant's right, limited control of participants, justification of withholding information, education purpose, n° of phishing simulation to reducing phishing susceptibility (click/not click).
Handbook of Research on Cybersecurity Risk in Contemporary Business Systems	Scopus	Wright J.B et al. (2023) [91]	Cybersecurity leadership ethics in healthcare	Beauchamp & Childress (2019) **(Top-down)**	Autonomy, non-maleficence, beneficence, and justice

Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes	Scopus	Tronnier F. et al. (2022) [92]	A Discussion on Ethical Cybersecurity Issues in Digital Service Chains	Formosa et al. + issues are specific to the digital service chain **(Top-down; bottom-up)**	Non-maleficence, Beneficence, Autonomy, Justice, Explicability (Formosa et al) + Principles of the Digital Service Chain
Computers and Security	Scopus	Thomas L. et al. (2022) [93]	A framework for data privacy and security accountability in data breach communications	Ethics in data breach **(Top-down)**	Privacy, security, and accountability
ACIS 2022 - Australasian Conference on Information Systems, Proceedings	Scopus	Sadeghi B. et al. (2022) [94]	How to increase ethical awareness in cybersecurity decision-making	Formosa et al. + awareness **(To-down)**	Non-maleficence, Beneficence, Autonomy, Justice, Explicability (Formosa et al.) + Awareness
WSEAS Transactions on Biology and Biomedicine	Scopus	Rajamäki J.; Hummelholm A. (2022) [95]	Ethical Resilience Management Framework for Critical Healthcare Information Infrastructure	Christen, Gordijn and Loi + Weber and Kleine (principlism based on Beauchamp and Childress's four principles of biomedical ethics) + technical resilience **(Top-down)**	Equality, fairness, freedom, or privacy + citizens' trust and confidence in the digital infrastructure, in policy makers and in state authorities + respect for autonomy, nonmaleficence, beneficence, and justice + efficiency and quality of services, the privacy of information and confidentiality of communication, usability of services, and safety (Technical principles mapped on the previous 4) + technical resilience
Cybersecurity: Ethics, Legal, Risks, and Policies	Scopus	Priyadarshini I.; Cotton C. (2022) [96]	CYBERSECURITY: Ethics, Legal, Risks, and Policies	General, specific (case studies) and professional issues (**Top-down, bottom-up, pragmatic)**	Vulnerability disclosure, encryption issues, automated security tools, sale restriction, incident responses, roles, and responsibilities; copyright infringement, piracy, intellectual property, computer fraud and misuse, plagiarism, cyberbullying, identity theft, ransomware attacks (they are crimes). It's not all of them, because only these open access
TransNav	Scopus	Oruc A. (2022) [97]	Ethical Considerations in Maritime Cybersecurity Research	6 Principles Exposed for Maritime sector + 4 Categories of Ethical Dilemmas **(Bottom-up; Top-down)**	Integrity, professional responsibility, accountability, confidentiality, legality, and openness. Dilemmas in developing tools and services, conducting research in support of state- sponsored cyberattacks, and sharing details in a research paper concerning cyber incidents that have occurred.

ACM International Conference Proceeding Series	Scopus	Maratsi M.I. et al. (2022) [98]	Ethical and Legal Aspects of Digital Forensics Algorithms: The Case of Digital Evidence Acquisition	From 3 requirements to ethical principles **(Bottom-up)**	Preservation of user privacy, preservation of the minimality principle, preservation of integrity, accuracy, legitimacy, and forensic soundness
BMC Medical Ethics	Scopus	Lorenzini G. et al. (2022) [99]	It takes a pirate to know one: ethical hackers for healthcare cybersecurity	Principles from case study (ethical hacking: penetration testing) **(Bottom-up)**	Privacy and freedom (positive)
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	Scopus	Kozhuharova D. et al. (2022) [100]	Ethics in Cybersecurity. What Are the Challenges We Need to Be Aware of and How to Handle Them?	Not to cause harm, data privacy and mitigation of security breaches + principles for IOT, cloud computing, and new technology **(Top-down, bottom-up, pragmatic)**	Not to cause harm, data privacy and mitigation of security breaches. For internet of things: privacy and consent; security. For cloud computing: ownership of data; security. For new technology: testing of future products (technical); participant consent.
2022 IEEE 12th Annual Computing and Communication Workshop and Conference, CCWC 2022	Scopus	Decusatis C. et al. (2022) [101]	A Cybersecurity Awareness Escape Room using Gamification Design Principles	IEEE code of ethics **(Pragmatic)**	To uphold the highest standards of integrity, responsible behavior, and ethical conduct in professional activities. To treat all persons fairly and with respect, to avoid harassment or discrimination, and to avoid injuring others. To strive to ensure this code is upheld by colleagues and co- workers.
Proceedings of the International Astronautical Congress, IAC	Scopus	Carlo A. et al. (2022) [102]	Understanding Space Vulnerabilities: Developing Technical and Legal Frameworks for AI and Cybersecurity in Space	Space sector **(Bottom-up, pragmatic)**	Cybersecurity and privacy
Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)	Scopus	Breig J.; Westhoff D. (2022) [103]	Short Paper: Debating Ethics with Cybersecurity Students	Moral theories, communities, individual (case studies) **(Top-down, bottom-up, pragmatic)**	Virtue, Utilitarian and Deontological Ethics; Community's Code of Conduct and Individual's Moral Compass
Proceedings - 7th IEEE European Symposium on Security and Privacy Workshops, Euro S and PW 2022	Scopus	Botes M.; Lenzini G. (2022) [104]	When Cryptographic Ransomware Poses Cyber Threats: Ethical Challenges and Proposed Safeguards for Cybersecurity Researchers	Menlo report 2012 (from Biomedical research ethics. Belmont Report). International Law Protections in Cyberspace (it does not cover all). International Code of Conduct for Information	Respect for Persons (informed consent), Beneficence, Justice (fairness and equity), Respect for Law and Public Interest (compliance; transparency and accountability). The latter is new compared to Belmont Report.

				Security (It is not international). **(Top-down)**	
Lecture Notes in Information Systems and Organisation	Scopus	Baskerville R. et al. (2022) [105]	Organizing Cybersecurity in Action: A Pragmatic Ethical Reasoning Approach	Pragmatism **(Pragmatic)**	Pragmatism attributes judgement-of-value to any action performed in a specific context. Pragmatic ethics cannot exist as a set of rules or principles; representatives of the 'communities' rather than committee
Proceedings - 2021 International Conference on Computational Science and Computational Intelligence, CSCI 20	Scopus	Wallace G.L. (2021) [106]	Ethics in Technology: Heroes Assemble!	Case study **(Bottom-up)**	Ethics of employee. Training of employee
Procedia Computer Science	Scopus	Santofimia M.J. et al. (2021) [107]	Cybersecurity in active and healthy ageing era	(Directive, act, ENISA…) + other issues **(Pragmatic)**	Unauthorized disclosure, Deception, Disruption, Usurpation
Computers and Security	Scopus	Formosa P. et al. (2021) [4]	A principlist framework for cybersecurity ethics	AI4People **(Top-down)**	Autonomy, Explicability, Non-maleficence, Justice, and Beneficence.
Ethics and Information Technology	Scopus	Dupuis M.; Renaud K. (2021) [108]	Scoping the ethical principles of cybersecurity fear appeals	6 ethical principles. Theories from all stakeholders **(Top-down, bottom-up)**	Obtain Institutional Review Board approval, make cybersecurity benefits salient, Deception must be justified, provide feasible recommended cybersecurity action, calibrate during deployment, Debrief at conclusion of experiment.
International Library of Ethics, Law and Technology	Scopus	Weber K.; Kleine N. (2020) [24]	Cybersecurity in Health Care	Beauchamp and Childress' four principles of biomedical ethics + technical aims + other ethical values **(Top-down, bottom-up, pragmatic)**	Autonomy, Non-Maleficence, Justice, Beneficence. Efficiency and quality of services (technical), Usability of services (technical), safety (technical). Fairness and equality, privacy and trust, freedom and consent, and dignity and solidarity.
Australasian Journal of Information Systems	Scopus	Poulsen A. et al. (2020) [109]	Cybersecurity, value sensing robots for LGBTIQ+ elderly, and the need for revised codes of conduct	Cybersecurity as a principle of AI (Theories [bioethics] and communities [IEEE, HGLE]. 4 principles related to cybersecurity. **(Top-down, pragmatic)**	Cybersecurity as a principle of AI, Privacy, Safe systems, Internet connectivity, Determinants of security for LGBTIQ+ elders

Proceedings of the 2020 IEEE International Conference "Quality Management, Transport and Information Security	Scopus	Nikolskaia K.; Naumov V. (2020) [110]	Ethical and legal principles of publishing open-source dual-purpose machine learning algorithms	Ethics of AI, publication of open-source code **(Top-down)**	Ethics of AI
International Library of Ethics, Law and Technology	Scopus	Lucas G. (2020) [111]	Cybersecurity and Cyber Warfare: The Ethical Paradox of 'Universal Diffidence'	Ethical issues: confidentiality **(Top-down)**	Confidentiality
International Library of Ethics, Law and Technology	Scopus	Loi M.; Christen M. (2020) [112]	Ethical Frameworks for Cybersecurity	Menlo report + Belmont report + Human factors + Consent and Justice **(Top-down)**	Confidentiality, integrity, availability; Respect for Persons, Beneficence, and Justice; Human factors (user acceptance, confusion, frustration, cognitive workload, error/risk reduction and the optimization of error-tolerant systems); Consent and Justice.
International Library of Ethics, Law and Technology	Scopus	Herrmann D.; Pridöhl H. (2020) [113]	Basic Concepts and Models of Cybersecurity	Information securities (confidentiality; integrity; availability of data) and system securities **(Top-down)**	Information securities (confidentiality; integrity; availability of data) and system securities
Science and Engineering Ethics	Scopus	Domingo-Ferrer J.; Blanco-Justicia A. (2020) [114]	Ethical Value-Centric Cybersecurity: A Methodology Based on a Value Graph	Moral theories **(Top-down)**	Autonomy, Security, Privacy, and Fairness.
Next-Generation Ethics: Engineering a Better Society	Scopus	Spinello R.A. [115]	Ethics in Cyberspace: Freedom, Rights, and Cybersecurity	Privacy vs public safety **(Bottom-up)**	Privacy vs public safety
Journal of Information, Communication and Ethics in Society	Scopus	Loi M. et al. (2019) [116]	Cybersecurity in health – disentangling value tensions	4 principles of bioethics. **(Top-down)**	Respect for autonomy, non-maleficence, beneficence, and justice
EPiC Series in Computing	Scopus	Etschmaier M.M. (2019) [117]	A purposeful systems design approach for cybersecurity	7 threats (ethical principles) **(Bottom-up)**	Threats to privacy (deprivation of access to reality and the truth, threats to personal freedom and to property rights), threats to businesses users, threats to affinity groups and formal social organizations, threats to sovereign states, threats to the global human sphere, obstacles to a sustainable cyberspace

2016 IEEE Symposium on Technologies for Homeland Security, HST 2016	Scopus	Holzer C.T.; Lerums J.E. (2016) [118]	The ethics of hacking back	Case studies: law enforcement and military **(Bottom-up)**	proportionality and minimizing harm
The Applied Ethics of Emerging Military and Security Technologies	Scopus	Dipert R.R. (2016) [119]	The ethics of cyberwarfare	Case studies **(Bottom-up)**	Prevention of harm
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes	Scopus	Shou D. (2012) [120]	Ethical considerations of sharing data for cybersecurity research	Case studies **(Bottom-up)**	Openness, balancing privacy rights, enabling sound scientific experimentation (for data sharing)
SECURITY IN COMPUTER AND INFORMATION SCIENCES, EURO-CYBERSEC 2018	WoS	Faiella, G et al. (2018) [121]	Building an Ethical Framework for Cross-Border Applications: The KONFIDO Project	eHealth code of ethics and GDPR) **(Pragmatic)**	Trust, privacy and security, proportionality, ownership and data control, equity and dignity.
PROCEEDINGS OF 2023 AUSTRALIAN COMPUTER SCIENCE WEEK, ACSW 2023	WoS	Khan, SBN et al. (2023) [122]	To breach or not? Profiling students' likelihood of breaching university ICT Codes of Conduct Student Profiling of Breach of ICT Codes of Conduct	Formosa et al. + awareness and training, behavior (human factor) **(Top-down)**	Formosa et al., awareness and training, behavior (human factor)
COMPUTERS & SECURITY	WoS	Posey, C et al. (2020) [123]	An exploratory examination of organizational insiders' descriptive and normative perceptions of cyber-relevant rights and responsibilities	Deonance theory **(Top-down)**	Personal norms and ethics (behavior). Deonance Theory (right vs. responsibility: descriptive and normative) Behaviors of insiders are not always interchangeable
INTERNET OF THINGS	WoS	Hernandez-Jaimes, ML et al. (2023) [124]	Artificial intelligence for IoMT security: A review of intrusion detection systems, attacks, datasets and Cloud-Fog-Edge architectures	Technical (confidentiality, integrity, availability) + other principles **(Bottom-up)**	Ethical matters of IoMT (including security: availability, integrity, and confidentiality), transparency, solidarity, human dignity, and responsibility)

Table A-3. Kozhuharova et al. (2022)

Principle/*Issue*	**Definition/Explanation**	**Case study**	**Mitigation**
Privacy and Consent	(Both not well defined) Privacy: protection of an individual's personal information and their right to control what data is collected, stored, and processed by devices they use. Consent: voluntary and well-informed agreement given by an individual to allow their personal data to be collected, processed, and used by IoT devices or other entities. (freely given, specific, informed and unambiguous, Withdrawal)	Internet of Things (smart watches smart kitchen appliance, smart water bottles, digitalized personal assistants)	GDPR Compliance
Security/safety	(not well defined) Protection of IoT (Internet of Things) devices and systems from risks such as malfunction, cyber intrusions, unauthorized access, and data breaches.	Internet of Things (Hacked Baby Monitors, IoT Children's Dolls as Surveillance Devices, Household Devices with Surveillance Weaknesses, implantable cardiac devices)	Regulators must work in cooperation with businesses to ensure threats are avoided or mitigated
Ownership of Data	Not well defined. Rights and control that individuals or entities have over the data they generate, upload, or store on cloud service platforms.	Cloud Computing (cloud service platforms)	Regulation, informed Decision-Making
Security	Not explicitly provide a definition. Protection of data, information, and systems from unauthorized access, breaches, cyber intrusions, and other potential threats that could compromise the confidentiality, integrity, and availability of sensitive information stored in cloud environments.	Cloud Computing	Security checks, encryption, cybersecurity expertise
Testing of future products (technical)	The potential risks the test subjects are exposed to (even unforeseen risks)	New technologies. how certain Bluetooth earphones and their waves can influence the brain of their users	• create a list of requirements (functional, design, performance, ethical, data protection and etc.) • implementing technical and Organisational Measures • Carrying Out an Impact Assessment (**the Standard**

			Data Protection Model (SDM) • Adopting Privacy Enhancing Technologies
Participant consent is not always clearly given		New technologies. Uber is testing its self-driving cars in real life environments within urban populated areas such as San Francisco, Phoenix, Pittsburgh [34]. While the direct participants in the testing have consented to take part, all pedestrians in the testing areas have not.	• Create a list of requirements (functional, design, performance, ethical, data protection and etc.) • implementing technical and Organisational Measures • Carrying Out an Impact Assessment (the Standard Data Protection Model (SDM) • Adopting Privacy Enhancing Technologies

Table A-4. Formosa et al. (2021)

Principle	**Definition/Explanation**	**Case study**	**Mitigation**
Autonomy	Cybersecurity technologies should be used in ways that respect human autonomy. Humans should be able to make informed decisions for themselves about how that technology is used in their lives.	• penetration testing • denial of service attack • ransomware attacks • cybersecurity system administration	
Explicability	Cybersecurity technologies should be used in ways that are intelligible, transparent, and comprehensible, and it should also be clear who is accountable and responsible for its use.	• penetration testing • denial of service attack • ransomware attacks • cybersecurity system administration	
Non-Maleficence	Cybersecurity technologies should not be used to intentionally harm humans or to make our lives worse overall.	• penetration testing • denial of service attack • ransomware attacks • cybersecurity system administration	
Justice	Cybersecurity technologies should be used to promote fairness, equality, and impartiality. It should not be used to unfairly discriminate, undermine solidarity, or prevent equal access.	• penetration testing • denial of service attack • ransomware attacks • cybersecurity system administration	

Beneficence	Cybersecurity technologies should be used to benefit humans, promote human well-being, and make our lives better overall.	• penetration testing • denial of service attack • ransomware attacks • cybersecurity system administration	

Table A-5. Weber and Klein (2020)

Principle	**Definition/Explanation**	**Case study**	**Mitigation**
Autonomy	• Respecting an individual's right to make their own decisions and act on them freely	• Cardiac Pacemakers and Other Implantable Medical Devices • Electronic Health Card (eHC) in Germany and Elsewhere	
Non-Maleficence	• Not to harm or ill-treat anyone		
Justice	• Guarantee of Fair Opportunities • Prevention of Unfair Discrimination • Efficient Use of Scarce Resources • Shared Responsibility		
Beneficence	• Contribute to the welfare and well-being of others		
Privacy of Information and confidentiality of communication	•		
Efficiency and Quality of Services (technical)	• Efficency: the ability to achieve maximum output or results with the minimum use of resources (optimize resource use. Justice and Beneficence) • Quality: standard of care provided to patients or individuals accessing healthcare facilities (enhance healthcare outcomes. Beneficence)	In the context of ICT systems in healthcare, the goal is to use information technology to streamline processes, reduce administrative burdens, and optimize resource allocation. By improving efficiency, the healthcare system can operate more smoothly and effectively, leading to benefits such as reduced waiting times, faster access to information, and cost savings.	Introduction of new treatments, procedures, or technologies that lead to better health outcomes for patients.
Usability of Services (technical)	• Degree of effectiveness, efficiency, and satisfaction with which users (patients, medical staff, and administrators) can accomplish their intended		

	tasks when interacting with a system (Nonmaleficence, Justice, Quality and Efficiency)		
Safety (technical)	• Reduction of Health-Threatening Risks • Alignment with Nonmaleficence • Alignment with Beneficence		
Fairness and Equality	• Equitable Access • Protection from Unfair Treatment • Emphasis on Justice		
Privacy and Trust	• Privacy: safeguarding patients' personal information and sensitive health data • Trust: patients' confidence in the healthcare system to prioritize their well-being and protect their privacy	Big data	
Freedom and Consent	• Freedom: unrestricted choice of using new technologies and the unhindered choice of how these technologies are utilized • Consent: informed consent, which empowers patients to make autonomous decisions about their healthcare options, aligning with the ethical principle of autonomy.		
Dignity and Solidarity	• Inherent Worth and Value • Universal and Inalienable Right • Upholding Dignity • Principle of beneficence		

Table A-6. Domingo-Ferrer (2019)

Principle	Definition/Explanation	Case study	Mitigation
Autonomy	• The state of being free from external control or influence (New Oxford 2015) • The capacity of an individual to make an informed, uncoerced decision (self-determination)		Hamiltonian paths: 1. Privacy helps autonomy, autonomy helps security, security helps fairness 2. Autonomy helps privacy, privacy helps security, security helps fairness 3. Autonomy helps security, security helps privacy, privacy helps fairness
Security	• The state of being free from danger or threat (New Oxford 2015)		
Privacy	• The state or condition of being free from being observed or disturbed by other people" (New Oxford 2015) • The right to be left alone" (Warren and Warren 1890) • The right of an individual to decide what information about himself should be communicated to others and under what circum-stances" (Westin 1970)		
Fairness	• The impartial and just treatment or behavior without favoritism or discrimination (New Oxford 2015)		

Table A-7. Hernandez-Jaimes, ML et al. (2023)

Type of attack	Principle	Definition/Explanation	Case study	Mitigation
Denial of service	Availability of data and health services (attacks in the application, transport, and network layers)	Confidentiality protects the patient's health status and treatment details. Furthermore, personal information from unauthorized users during the storing or transmitting of IoMT-based data. Integrity guarantees that medical information is not corrupted or deleted during storing or transmission of IoMT-based data. Availability ensures the continuous operation of medical devices, services, and clinical records of patients. Additionally, play a critical role in promptly responding to health emergencies.		• New standards and laws • Intrusion Detection and Prevention System
Distributed denial of service	Availability of data and health services (attacks in the application, transport, and network layers)			
Man in the middle	Confidentiality and integrity (attacks in the network and transport layers)			
Ransomware	Integrity and availability (attacks in the application layer)		• University of Vermont Medical Center (2020) • University Hospital of Düsseldorf (2020) • the DCH Health System in West Alabama (2019) • the Indiana hospital system (2018) • The National Health Service (NHS) of Britain (2017, Wannacry)	
Information gathering	Confidentiality (attack in the network layer)			
Malware	Integrity and availability (attack in the application layer)			
Injection	Confidentiality, integrity, and availability (attack in the application layer)			
Password	Confidentiality, integrity, and availability (attack in the application layer)			

Table A-8. Yeng et Wright (2023) [90]

Type of attack	Issue	Definition/Explanation	Case study	Mitigation
Phishing simulation	Deception vs informed content	Providing participants with all relevant information about a study (informed consent) and deliberately withholding certain details to avoid biasing the results (deception)		Precaution by refraining from collecting personal info (ex: substitution)
	Potential harm to participants	In the context of deception vs informed consent		Establish a secure connection
	Infringement upon privacy (	In the context of deception vs informed consent		Encryption method (GDPR) for data protection
	Violation of participant's	In the context of deception vs informed consent		
	Limited control of participants	In the context of deception vs informed consent		
	Justification of withholding information	In the context of deception vs informed consent		
	Phishing susceptibility (click/not click)	State of mind when you attempt to click/not click on malicious links		Perform N° of phishing simulation

Table A-9. Lorenzini et al. (2022)

Type of attack	Principle/Issue	Definition/Explanation	Case study	Mitigation
Penetration testing	Privacy	*Pen-testers do not disseminate or leak data.*		Official code of ethics
	Freedom	Freedom invokes free and open access to information with the pedagogical goal of equally allowing humans to educate themselves.		Financial resources for the employment of pen-tester

	Penetration testing is not common in healthcare facilities	It partly depends on the fears and prejudices towards the hacking practice, on the lack of a clear, and official, code of ethics for this profession, and on the limited financial resources that these facilities can devote to cybersecurity		Hire a company providing pen-tests services
	Not official ethics code for pen-tester			Draft an international code of ethics that can less arbitrarily define and describe moral principles, standards, expectations, and best practices

Table A-10. ENISA [22], [125]

Principle	**Definition/Explanation**	**Case study**	**Mitigation**
Confidentiality	The protection of communications or stored data against interception and reading by unauthorized persons.		Cybersecurity certification
Integrity	The confirmation that data which has been sent, received, or stored are complete and unchanged.		Cybersecurity certification
Availability	The fact that data is accessible and services are operational.		Cybersecurity certification
Authenticity	Property that an entity is what it claims to be.		Cybersecurity certification
Availability	secure ICT products, ICT services and ICT processes		
Non-repudiation			

Table A-11. ISO/IEC PDTR 13335-1(However, the standard is not free of charge, and its provisions are not publicly available).

Principle	**Definition/Explanation**	**Case study**	**Mitigation**
Confidentiality	The property that information is not made available or disclosed to unauthorized individuals, entities, or processes.		

Integrity	The property that data has not been altered or destroyed in an unauthorized manner.		

Table A-12. IBM. Data protection [126]

Principle/Issue/risk	Definition/Explanation	Case study	Mitigation
Ownership; (Upholding data integrity)	Who owns the data you're using. Just because a user gives you data doesn't mean you own it.		Consent, data protection and data respect.
Transparency; (Mistrust)	Being clear with customers about how their data is used.		Empower users to understand the purposes and lifecycle of customer data
Privacy	When a company collects information, stores it and analyzes it, that information should not be used, stored, shared, maintained, retained or disposed of outside of the agreed purposes for which it was originally obtained.		Data ethics and security policies
Intention (Responsible AI; Trust in your technology, your processes and the outcomes of their data use)	Be clear on your purpose when using data and machine intelligence.		Tools to increase our trust in AI—such as explainer toolkits, taxonomies of AI techniques and AI governance solutions
Prevention (Patience, loyalty and faith)		Data breaches, ransomware attacks and slipups. One IBM study found that companies that have fully deployed AI and automation as part of their security strategy save an average of USD 3.05 million in data breach costs compared to those who have yet to do so.	Risk management; AI and automation as part of their security strategy

Table A-13. The standard data protection model. A method for Data Protection advising and controlling on the basis of uniform protection goals. Version 3.0a Adopted by the 104. Conference of the Independent Data Protection Supervisory Authorities [127]

Principle/Issue/risk	Definition/Explanation	Case study	Mitigation
Availability	The requirement that access to personal data and their processing is possible without delay and that the data can be		• Creation of backups of data, process states, configurations, data structures, transaction histories, etc. according to a

	used properly in the intended process. From GDPR: • B1.18 Availability • B1.20 Recoverability • B1.19 Resilience B1.22 Rectification and Mitigation of data protection breaches		tested concept (B1.20 Recoverability), • Protection against external influences (malware, sabotage, force majeure) (B1.18 Availability, B1.19 Resilience, B1.22 Rectification and mitigation of data protection violations), • Documentation of data syntax (B1.18 Availability, B1.20 Recoverability), • Redundancy of hardware, software and infrastructure (B1.20 availability, B1.19 resilience), • Implementation of repair strategies and backup processes (B1.19 Resilience, B1.20 Recoverability, B1.22 Rectification and mitigation of data breaches), • Preparation of a contingency plan for restoring processing activity (B1.19 Resilience, B1.20 Recoverability), Representation arrangements for absent employees (B1.18 Availability).
Integrity	On the one hand, integrity refers to to the requirement that information technology processes and systems continuously comply with the specifications that were defined for them to perform their intended functions (B1.6 Integrity). On the other hand, integrity refers to the property that the data to be processed remain intact (B1.6 Integrity), complete, correct and up-to-date (B1.4 Correctness). From the GDPR:		• Restriction of write and modification permissions (B1.6 Integrity) • Use of checksums, electronic seals and signatures in accordance with a cryptographic concept (B1.6 Integrity, B1.4 Accuracy, B1.23 Appropriate monitoring of processing, B1.22 Removal and mitigation of data breaches) • Documented assignment of authorizations and roles (B1.6 Integrity)

	B1.23 Adequate monitoring of processing B1.22 Rectification and mitigation of data protection breaches B1.19 Resilience B1.16 Freedom from errors and discrimination		• Erasure or rectifying of incorrect data (B1.4 Accuracy) • Hardening of IT systems so that they have no or as few secondary functionalities as possible (B1.6 Integrity, B1.19 Resilience) • Processes for maintaining the timeliness of data (B1.4 Accuracy) • Processes for identification and authentication of persons and equipment (B1.6 Integrity) • Definition of the intended behavior of processes and regular tests to determine and document functionality, risks, security gaps and side effects of processes (B1.6 Integrity, B1.16 Freedom from errors and discrimination in profiling, B1.19 Resilience) • Determination of the target behavior of processes and procedures and regular performance of tests to ascertain or determine the current state of processes (B1.6 Integrity, B1.16 Freedom from errors and discrimination in profiling, B1.23 Appropriate monitoring of processing, B1.19 Resilience) • Protection against external influences (espionage, hacking) (B1.6 Integrity, B1.19 Resilience, B1.22 Rectification and mitigation of data protection violations)

Confidentiality	The requirement that no unauthorised person can access or use personal data (B1.7 Confidentiality). From the GDPR: B1.19 Resilience B1.22 Remedy and mitigation of data protection violations		• Definition of a concept for role-based access control according to the necessity principle on the basis of identity management by the controller (B1.7 Confidentiality) • Implementation of a secure authentication procedure (B1.7 Confidentiality) • Limitation of authorized personnel to those who are verifiably responsible (locally, professionally), qualified, reliable (if necessary, with security clearance) and formally approved, and with whom no conflict of interests may arise in the exercise of their duties (B1.7 Confidentiality) • Specification and monitoring of the use of authorized resources, in particular communication channels (B1.7 Confidentiality, B1.22 Remedy and mitigation of data breaches) • specified environments (buildings, rooms) equipped for processing activities (B1.7 Confidentiality) • Definition and monitoring of organizational processes, internal regulations and contractual obligations (obligation to maintain data secrecy, confidentiality agreements, etc.) (B1.7 Confidentiality, B1.22 Elimination and mitigation of data protection violations) • Encryption of stored or transferred data and processes for managing and protecting cryptographic information (cryptographic

			concept) (B1.7 Confidentiality) • Protection against external influences (espionage, hacking) (B1.7 Confidentiality, Resilience, B1.22 Removal and mitigation of data protection violations)
Unlinkability	The requirement that personal data shall not be merged, i. e. linked. From the GDPR: B1.2 Purpose limitation		• Restriction of processing, use and transfer permissions (B1.2 Purpose limitation) • program-wise omission or deactivation of interfaces in processing methods and components (B1.2 Purpose limitation) • regulatory measures to prohibit backdoors and quality assurance audits for compliance in software development (B1.2 Purpose limitation) • Separation according to organizational/departmental boundaries (B1.2 Purpose limitation) • Separation by means of role concepts with graduated access rights on the basis of identity management by the controller and a secure authentication process (B1.2 Purpose limitation) • Approval of user-controlled identity management by the controller (B1.2 Purpose limitation) • Use of purpose specific pseudonyms, anonymization services, anonymous credentials, processing of pseudonymous or

			anonymized data (B1.2 Purpose limitation) • regulated processes for amending the purposes of the processing (B1.2 Purpose limitation)
Transparency	The requirement that both data subjects (B1.1 Transparency for data subjects) and system operators (B1.23 Adequate monitoring of processing) and competent supervisory bodies (B1.8 Accountability and verifiability) shall be able to identify to varying degrees which data are collected and processed when and for what purpose in a processing activity, which systems and processes are used to determine where the data are used and for what purpose, and who has legal responsibility for the data and systems in the various phases of data processing. From the GDPR: B2 Consent management		• Documentation in the sense of an inventory of all processing activities in accordance with Art. 30 GDPR (B1.8 Accountability and Verifiability) • Documentation of the components of processing activities, in particular business processes, databases, data flows and network plans, IT systems used for this purpose, operating procedures, descriptions of processing activities, interaction with other processing activities (B1.8 Accountability and verifiability) • Documentation of tests, of the release and, where appropriate, the data protection impact assessment of new or modified processing activities (B1.8 Accountability and Verifiability) • Documentation of the factors used for profiling, scoring or semi-automated decisions (B1.8 Accountability and Verifiability) • Documentation of contracts with internal employees, contracts with external service providers and third parties from whom data is collected or transmitted,

			business distribution plans, responsibility regulations (B1.8 Accountability and Verifiability) • Documentation of consents, their revocation and objections (B2 Consent Management) • Logging of accesses and changes (B1.23 Adequate monitoring of processing, B1.8 Accountability and Verifiability) • Versioning (B1.23 Appropriate monitoring of processing, B1.8 Accountability and verifiability) • Documentation of processing by means of protocols on the basis of a logging and evaluation concept (B1.23 Appropriate monitoring of processing, B1.8 Accountability and Verifiability) • Documentation of the data sources, e. g. the implementation of information duties towards data subjects where their data were collected and the handling of data breaches (B1.1 Transparency for data subjects, B1.8 Accountability and verifiability) • Notification of data subjects in the event of data breaches or further processing for another purpose (B1.1 Transparency for data subjects) • Traceability of the activities of the controller for granting data subjects' rights (B1.1 Transparency for data subjects)

			• Consideration of the information rights of data subjects in the logging and evaluation concept (B1.1 Transparency for data subjects) • Provision of information on the processing of personal data to data subjects (B1.1 Transparency for data subjects)
Intervenability	The requirement that the data subjects' rights to notification, information, rectification (B1.11 Possibility of rectification of data), erasure (B1.12 Erasure of data), restriction (B1.13 Restriction of processing of data), data portability (B1.14 Data portability), objection and obtaining the intervention in automated individual decisions (B1.15 Possibility of intervention in processes of automated decisions) are granted without undue delay and effectively if the legal requirements exist (B1.10 Support in the exercise of data subjects' rights) and data controller is obliged to implement the corresponding measures. From the GDPR: • B1.9 Identification and authentication • B3 Implementation of supervisory orders • B1.22 Remedying and mitigating data protection breaches • B2 Consent management B1.17 Data protection-friendly default settings		• Measures for differentiated consent, revocation and objection options (B2 Consent management) • Creation of necessary data fields, e. g. for blocking indicators, notifications, consents, objections, counterstatements (B1.11 Possibility of correcting data, B1.13 Limitability of processing, B1.17 Data protection through pre-settings, B2 Consent Management, B3 Implementation of Supervisory Orders) • Documented processing of faults, problem handling and changes to processing activities as well as to technical and organizational measures (B1.22 Rectification and Mitigation of data protection violations, B1.13 Restriction of processing, B3 Implementation of Supervisory Orders) • Possibility of deactivating individual functionalities without affecting the overall system (B1.22 Removal and mitigation of data protection violations, B1.13 Limitability

			of processing, B3 Implementation of supervisory orders) • Implementation of standardized query and dialogue interfaces for data subjects to assert and/or enforce claims (B1.10 Support in exercising data subjects' rights) • Operation of an interface for structured, machine-readable data for the retrieval by data subjects (B1.10 Support in exercising data subjects' rights, B1.14 Data portability) • Identification and authentication of persons who wish to exercise data subjects' rights (B1.9 Identification and authentication) • Establishment of a Single Point of Contact (SPoC) for data subjects (B1.10 Support in the exercise of data subjects' rights) • Operational possibility of compiling, consistently rectifying, blocking and erasure of all data stored on a person (B1.11 Rectification, B1.12 Erasure, B1.13 Restriction of data processing, B1.14 Data Portability, B3 Implementation of Supervisory Orders) • Provision of options for data subjects in order to be able to set up programs in line with data protection requirements (B1.10 Support in exercising data subjects' rights, B1.17 Data protection by default)

Data Minimisation	The fundamental requirement under data protection law to limit the processing of personal data to what is appropriate, substantial and necessary for the purpose. From the GDPR: B1.3 Data minimisation B1.5 Storage limitation B1.17 Data Protection by Default		• Reduction of recorded attributes of data subjects (B1.3 Data Minimization) • Reduction of processing options in each processing step (B1.3 Data Minimization) • Reduction of the possibility of gaining knowledge of existing data (B1.3 Data Minimization) • Establishing default settings for data subjects which limit the processing of their data to what is necessary for the purpose of the processing. (B1.17 Data protection by default) • Preference for automated processes (not decision processes), which make it unnecessary to gain knowledge of processed data and limit influence in comparison to dialogue controlled processes (B1.3 Data Minimization) • Implementation of data masks that suppress data fields, and automatic blocking and erasure routines, pseudonymization and anonymization processes (B1.3 Data Minimization, B1.5 Storage limitation) • Definition and implementation of an erasure concept (B1.5 Storage limitation) • Rules for the monitoring of processes to change processing activities (B1.3 Data minimization)

Table A-14. WHO. The protection of personal data in health information systems – principles and processes for public health. [128]

Principle/Issue/risk	Definition/Explanation	Case study	Mitigation
Fair, lawful and transparent	Personal data shall be processed fairly, lawfully and in a transparent manner in relation to the data subject. In particular, personal data shall not be processed unless permitted by law, based on a preponderant legal interest of the processor or consented to by the data subject.		• Set up a risk management system for data protection, covering various dimensions of risk such as financial risk or a reputational risk. • Ensure support from the highest management level of the institution. • Report to the highest management level regularly, and prepare regular (annual or similar) reports on data protection. • Set up mid- and long-term financial planning for data protection resources. • Follow internationally recognized standards such as ISO 27001 and audit compliance with these frameworks. • Conduct a DPIA for high-risk activities. • If the institution does not adhere to an overall, holistic data protection framework, start small and develop a concept for the team or department.
Purpose limitation	Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.		
Accuracy	Personal data shall be accurate and, where necessary, kept up to date.		
Data minimization	Personal data shall be adequate, relevant and limited to what is		

	necessary in relation to the purpose for which they are processed.		
Storage limitation	Personal data processed for any purposes shall not be kept for longer than is necessary for those purposes.		
Rights of data subjects	Personal data shall be processed in accordance with the rights of data subjects as stipulated by the applicable data protection laws.		
Integrity and confidentiality	Appropriate physical, technical, legal and organizational measures shall be taken against unauthorized or unlawful processing of personal data and against accidental loss, alteration or damage to personal data.		
International transfer of personal data	Personal data shall not be transferred to a third country or international organization unless that country/organization ensures an adequate level of protection of the rights and freedoms of the data subjects in relation to the processing of personal data.		

Table A-15. Mapping GDPR principles with SDM principles {Citation}

Principles of the GDPR	**Principles of SDM**
Transparency for data subjects (Art. 5 para. 1 lit a, Art. 12 para. 1 and 3 to Art. 15, Art. 34 GDPR)	Transparency
Purpose limitation Art. 5 para. 1 lit. c GDPR	Unlinkability
Data minimisation (Art. 5 para. 1 lit. c GDPR)	Data Minimisation
Accuracy (Art. 5 para. 1 lit. d GDPR),	Integrity

Storage limitation (Art. 5 para. 1 lit. e GDPR),	Data Minimisation
Integrity (Art. 5 para. 1 lit. f GDPR, Art. 32 para. 1 lit. b GDPR),	Integrity
Confidentiality (Art. 5 para. 1 lit. f, Art. 28 para. 3 lit. b, Art. 29, Art. 32 para. 1 lit. b, Art. 32 para. 4, Art. 38 para. 5 GDPR),	Confidentiality
Accountability and Verifiability (Art. 5 para. 2, Art. 7 para. 1, Art. 24 para. 1, Art. 28 para. 3 lit. a, Art. 30, Art. 33 para. 5, Art. 35, Art. 58 par. 1 lit. a and lit. e GDPR)	Transparency
Support in exercising data subjects' rights (Art. 12 para. 2 GDPR)	Intervenability
Identification and Authentication (Art. 12 para. 6 GDPR)	Intervenability
Rectification of data (Art. 5 lit. d, Art. 16 GDPR)	Intervenability
Erasure of Data (Art. 17 para. 1 GDPR)	Intervenability
Restriction of data processing (Art. 18 GDPR)	Intervenability
Data portability (Art. 20, para 1 GDPR)	Intervenability

Possibility to intervene in processes of automated decisions (Art. 22 para 3 GDPR)	Intervenability
Freedom from error and discrimination in profiling (Art. 22 para 3, 4 in connection with recital 71)	Integrity
Data protection-friendly default settings (Art. 25 para 2 GDPR)	Data Minimisation, Intervenability
Availability (Art. 32 para 1 lit. b GDPR)	Availability
Resilience (Art. 32 para. 1 lit. b GDPR),	Availability, Integrity, Confidentiality
Restorability (Art. 32 para 1 lit. b, lit. c GDPR)	Availability
Evaluability (Art. 32 para. 1 lit. d GDPR).	Must be implemented as a process that encompasses all requirements (see Chapter D4 Data Protection Management with SDM).
Remedy and mitigation of data protection breaches (Art. 33, para 3 lit. d, Art. 34 para 2 GDPR)	Integrity, Intervenability, Confidentiality, Availability
Adequate monitoring of the processing (Art. 32, 33, 34 GDPR)	Transparency, Integrity
Consent management (Art. 4 No. 11, Art. 7 and 4 GDPR).	Transparency, Intervenability

Implementation of supervisory orders (Art. 58 para 2 lit. f und lit. j)	Intervenability

Ethics and data protection

05 July 2021

Figure A-1. Ethics and data protection from European Commission.

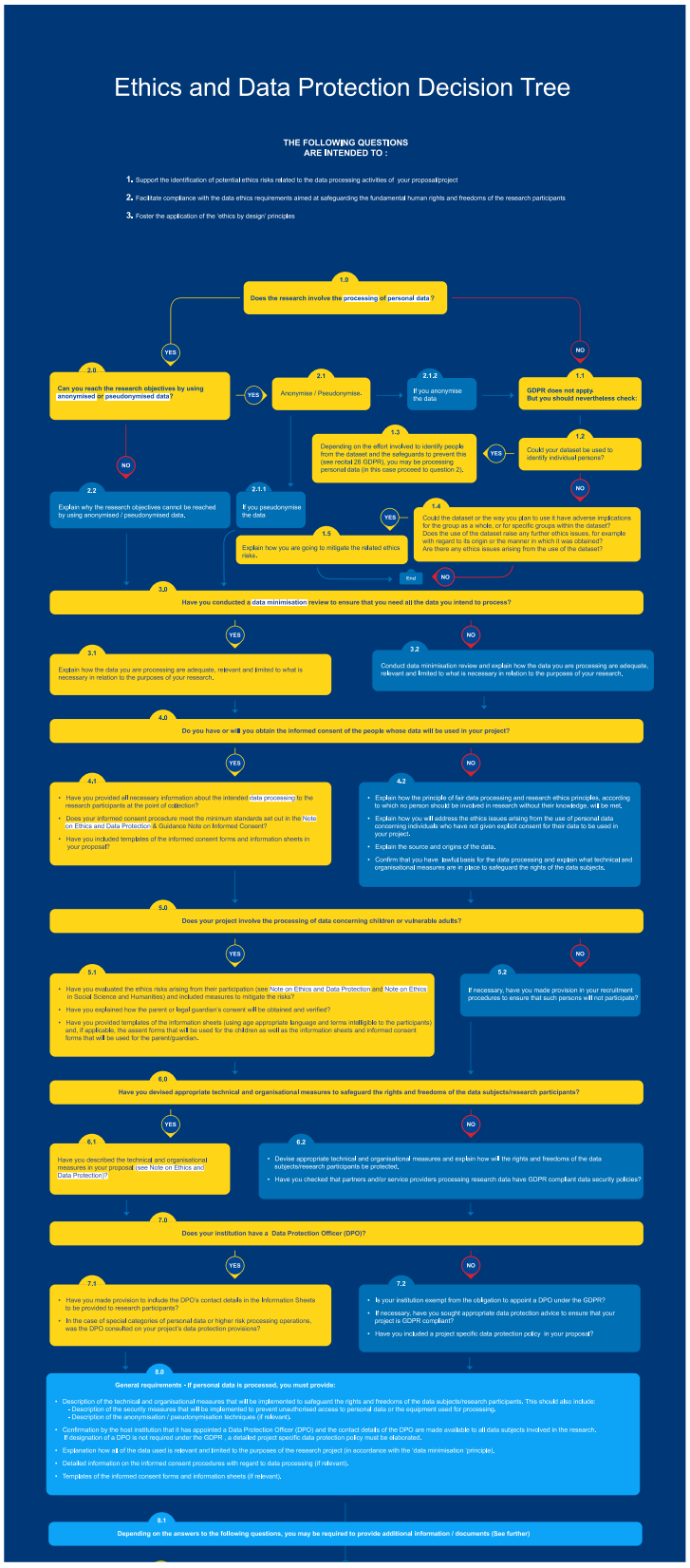

Figure A-2. Example of qualitative method [129]

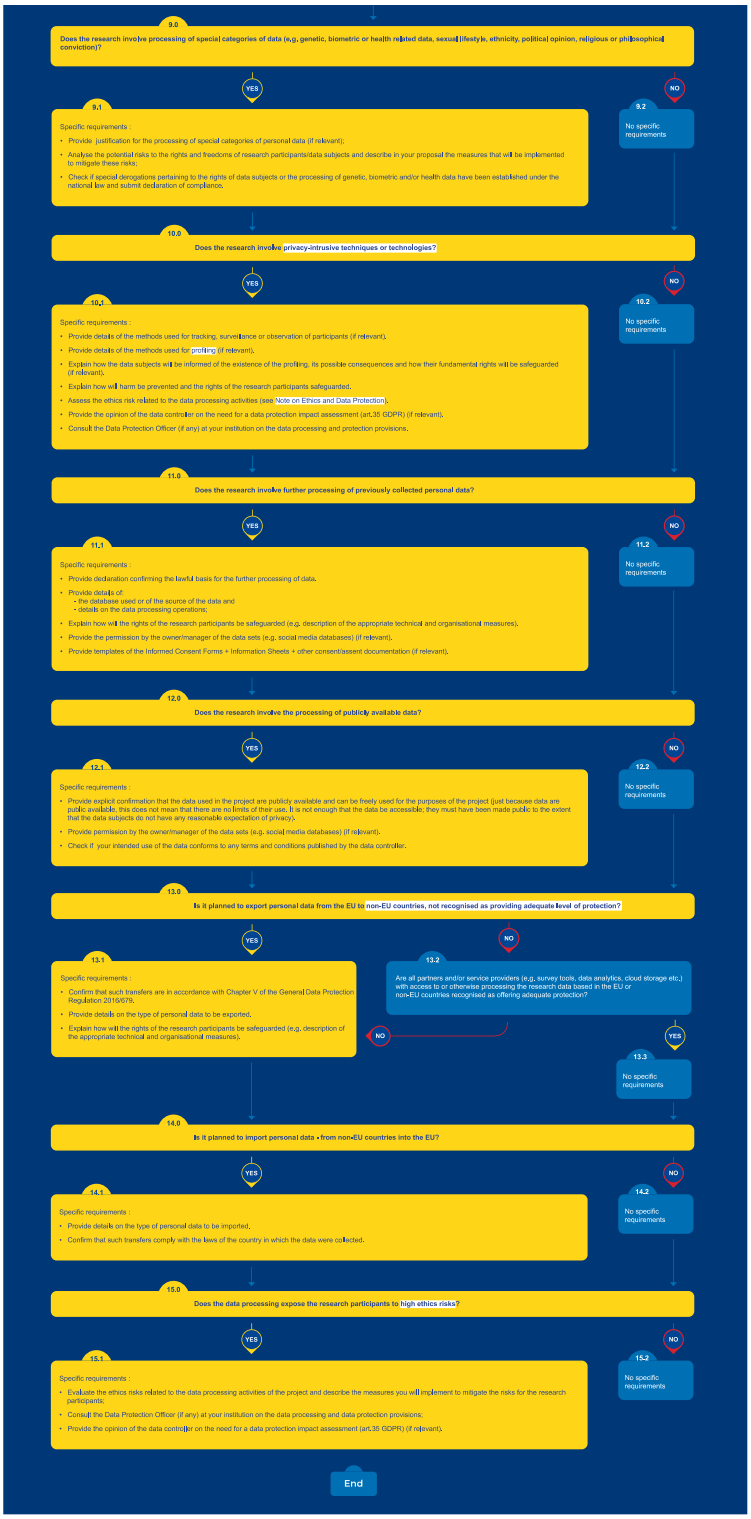

Figure A-3. Example of qualitative method [129]

Table A-16. DPIA criteria

ART.35	WP 29	Our study
A systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person.	Evaluation or scoring, including profiling and predicting, especially from "aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behavior, location or movements" (recitals 71 and 91).	Purpose of the processing
Processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10.	Automated-decision making with legal or similar significant effect: processing that aims at taking decisions on data subjects producing "legal effects concerning the natural person" or which "similarly significantly affects the natural person" (Article 35(3)(a)).	Category of personal data
A systematic monitoring of a publicly accessible area on a large scale.	Systematic monitoring: processing used to observe, monitor or control data subjects, including data collected through networks or "a systematic monitoring of a publicly accessible area" (Article 35(3)(c)).	Category of data subjects

Table A-17. DPIA criteria

ART.35	WP 29	Our study
	Sensitive data or data of a highly personal nature: this includes special categories of personal data as defined in Article 9 (for example information about individuals' political opinions), as well as personal data relating to criminal convictions or offences as defined in Article 10.	Age group of data subects
	Data processed on a large scale: the GDPR does not define what constitutes large-scale, though recital 91 provides some guidance.	Data volume (Number and quantity of data subjects)
	Matching or combining datasets, for example originating from two or more data processing operations performed for different purposes and/or by different data controllers in a way that would exceed the reasonable expectations of the data subject17.	Category of data processing
	Data concerning vulnerable data subjects (recital 75).	Frequencies of data processing
	Innovative use or applying new technological or organisational solutions, like combining use of finger print and face recognition for improved physical access control. etc.	Category of recipient
	When the processing in itself "prevents data subjects from exercising a right or using a service or a contract" (Article 22 and recital 91)	

Table A-18. Severity related to the category of personal data

Category of personal data	R	I	D
--------Common Data--------			
Public data	1	1	1
Generic identification data	2	2	2
Education and work experience	2	3	3
Work	3	2	3
Financial identification data	3	3	3
Photos or videos (that do not link to sensitive data)	4	4	4
Online identifier	4	2	2
Copies of identity documents	7	3	3
--------Special data--------			
Racial or ethnic origin	6	2	2
Political views	6	2	2
Trade union membership	6	4	4
Sexual orientation	7	2	2
Sex life	7	2	2
Philosophical convictions	5	2	2
Religious beliefs	6	2	2
Physical data, photos or videos that link to sensitive data	7	3	2
Biometrics	7	3	4
Judicial (self-certification)	5	3	2
Judicial data	8	3	3
Health data (labour law)	6	4	4
Health data (Diagnosis and treatment)	8	8	8
Health data (Services provided)	7	4	4
Genetic data	10	8	8
Geolocation	8	2	2

Table A-19. Severity related to the age group of data subjects

Age group concerned	R	I	D
Adults (>18 years old)	1,0	1,0	1,0
Adults and/or elderly (>65 years)	1,1	1,1	1,1
Adults and/or under 18 years old (>14 years old)	1,1	1,1	1,1
Adults and/or children under 14 years old	1,2	1,2	1,2
All (from minors to the elderly)	1,2	1,2	1,2

Table A-20. Severity related to the categories of recipients

Categories of Recipients	R	I	D
Internal organization structures	1,0	1,0	1,0
External manager	1,0	1,0	1,0
Social security institutions	1,0	1,0	1,0
Marche Region	1,0	1,0	1,0
Healthcare Authorities	1,1	1,1	1,0
Justice and police services	1,0	1,0	1,0
Public administration	1,0	1,0	1,0
Data subject	1,0	1,0	1,0
Data Subject Advisors	1,1	1,1	1,0
Banks	1,0	1,0	1,0
Ministry of Health	1,0	1,0	1,0
Ministry of Economy and Finance	1,0	1,0	1,0
Marketing companies	1,2	1,2	1,2
Project partners	1,1	1,0	1,0

Table A-21. Severity related to the categories of processing

Categories of processing	R	I	D
Standard data lifecycle management	1,1	1,1	1,1
Collection	1,0	1,0	1,0
Registration	1,0	1,1	1,0
Archiving	1,1	1,1	1,1
Use	1,1	1,0	1,0
Update	1,0	1,0	1,0
Modification	1,0	1,1	1,0
Cancellation	1,0	1,0	1,1
Consultation	1,0	1,0	1,0
Communication	1,1	1,0	1,0
Transfer	1,1	1,1	1,1
Movement	1,1	1,1	1,1

Diffusion	1,4	1,0	1,0
Limitation	1,0	1,0	1,1
Monitoring	1,1	1,0	1,0
Surveillance / Control	1,2	1,0	1,0
Profiling	1,5	1,5	1,5
Interconnection with other databases	1,4	1,0	1,0
Comparison with other databases	1,4	1,0	1,0
Automated decision-making	1,4	1,0	1,0
Automated evaluation process	1,4	1,0	1,0
Extraction	1,0	1,0	1,0
Organization	1,0	1,1	1,1
Structuring	1,0	1,1	1,1
Adaptation	1,0	1,1	1,1
Digital preservation (storage)	1,1	1,0	1,0
Secure Erase	1,0	1,1	1,1
Destruction	1,0	1,1	1,1

Table A-22. Severity related to the frequency of processing

Treatment frequency	**R**	**I**	**D**
Occasional	0,8	1,0	1,0
Standard	1,0	1,0	1,0
Systematic	1,2	1,2	1,2
Daily	1,1	1,0	1,0
Weekly	1,1	1,0	1,0
Monthly	1,0	1,0	1,0
Quarterly	1,0	1,0	1,0
Four-month	1,0	1,0	1,0
Semiannual	1,0	1,0	1,0
Annual	1,0	1,0	1,0

Table A-23. Severity related to the volume of data

Volume of data processed	**R**	**I**	**D**
Low number of data subject	1,0	1,0	1,0
Average number of data subject	1,1	1,0	1,0
High number of data subject	1,2	1,0	1,0
Large scale	1,5	1,2	1,2

Figure A-4.ALTAI

REQUIREMENT #2 Technical Robustness and Safety

A crucial requirement for achieving Trustworthy AI systems is their dependability (the ability to deliver services that can justifiably be trusted) and resilience (robustness when facing changes). Technical robustness requires that AI systems are developed with a preventative approach to risks and that they behave reliably and as intended while minimising unintentional and unexpected harm as well as preventing it where possible. This should also apply in the event of potential changes in their operating environment or the presence of other agents (human or artificial) that may interact with the AI system in an adversarial manner. The questions in this section address four main issues: 1) security; 2) safety; 3) accuracy; and 4) reliability, fall-back plans and reproducibility.

Glossary: Accuracy; AI Bias; AI System; AI Reliability; AI Reproducibility; (Low) Confidence Score; Continual Learning; Data Poisoning; Model Evasion; Model Inversion; Pen Test; Red-team.

Resilience to Attack and Security

- Could the AI system have adversarial, critical or damaging effects (e.g. to human or societal safety) in case of risks or threats such as design or technical faults, defects, outages, attacks, misuse, inappropriate or malicious use?
- Is the AI system certified for cybersecurity (e.g. the certification scheme created by the Cybersecurity Act in Europe)[19] or is it compliant with specific security standards?
- How exposed is the AI system to cyber-attacks?
    - Did you assess potential forms of attacks to which the AI system could be vulnerable?
    - Did you consider different types of vulnerabilities and potential entry points for attacks such as:
        - Data poisoning (i.e. manipulation of training data);
        - Model evasion (i.e. classifying the data according to the attacker's will);
        - Model inversion (i.e. infer the model parameters)
- Did you put measures in place to ensure the integrity, robustness and overall security of the AI system against potential attacks over its lifecycle?
- Did you red-team/pentest the system?
- Did you inform end-users of the duration of security coverage and updates?
    - What length is the expected timeframe within which you provide security updates for the AI system?

Figure A-5. Technical Robustness. ALTAI

General Safety

- Did you define risks, risk metrics and risk levels of the AI system in each specific use case?
    - Did you put in place a process to continuously measure and assess risks?
    - Did you inform end-users and subjects of existing or potential risks?
- Did you identify the possible threats to the AI system (design faults, technical faults, environmental threats) and the possible consequences?
    - Did you assess the risk of possible malicious use, misuse or inappropriate use of the AI system?
    - Did you define safety criticality levels (e.g. related to human integrity) of the possible consequences of faults or misuse of the AI system?
- Did you assess the dependency of a critical AI system's decisions on its stable and reliable behaviour?
    - Did you align the reliability/testing requirements to the appropriate levels of stability and reliability?
- Did you plan fault tolerance via, e.g. a duplicated system or another parallel system (AI-based or 'conventional')?
- Did you develop a mechanism to evaluate when the AI system has been changed to merit
  a new review of its technical robustness and safety?

Figure A-6.Safety ALTAI

REQUIREMENT #3 Privacy and Data Governance

Closely linked to the principle of prevention of harm is privacy, a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy.

Glossary: Aggregation and Anonymisation; AI System; Data Governance; Data Protection Impact Assessment (DPIA); Data Protection Officer (DPO); Encryption; Lifecycle; Pseudonymisation; Standards; Use Case.

Privacy

This subsection helps to self-assess the impact of the AI system's impact on privacy and data protection, which are fundamental rights that are closely related to each other and to the fundamental right to the integrity of the person, which covers the respect for a person's mental and physical integrity.

- Did you consider the impact of the AI system on the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection?
- Depending on the use case, did you establish mechanisms that allow flagging issues related to privacy concerning the AI system?

Data Governance

This subsection helps to self-assess the adherence of the AI system('s use) to various elements concerning data protection.

- Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?
- Did you put in place any of the following measures some of which are mandatory under the General Data Protection Regulation (GDPR), or a non-European equivalent?
    - Data Protection Impact Assessment (DPIA)[23];
    - Designate a Data Protection Officer (DPO)[24] and include them at an early state in the development, procurement or use phase of the AI system;
    - Oversight mechanisms for data processing (including limiting access to qualified personnel, mechanisms for logging data access and making modifications);
    - Measures to achieve privacy-by-design and default (e.g. encryption, pseudonymisation, aggregation, anonymisation);

[23] https://gdpr.eu/data-protection-impact-assessment-template/.
[24] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-officers/.

Figure A-7. Privacy and Data Governance. ALTAI

        - Data minimisation, in particular personal data (including special categories of data);
    - Did you implement the right to withdraw consent, the right to object and the right to be forgotten into the development of the AI system?
    - Did you consider the privacy and data protection implications of data collected, generated or processed over the course of the AI system's life cycle?
- Did you consider the privacy and data protection implications of the AI system's non-personal training-data or other processed non-personal data?
- Did you align the AI system with relevant standards (e.g. ISO[25], IEEE[26]) or widely adopted protocols for (daily) data management and governance?

Figure A-8. Privacy and Data Governance. ALTAI. Following Figure 17.

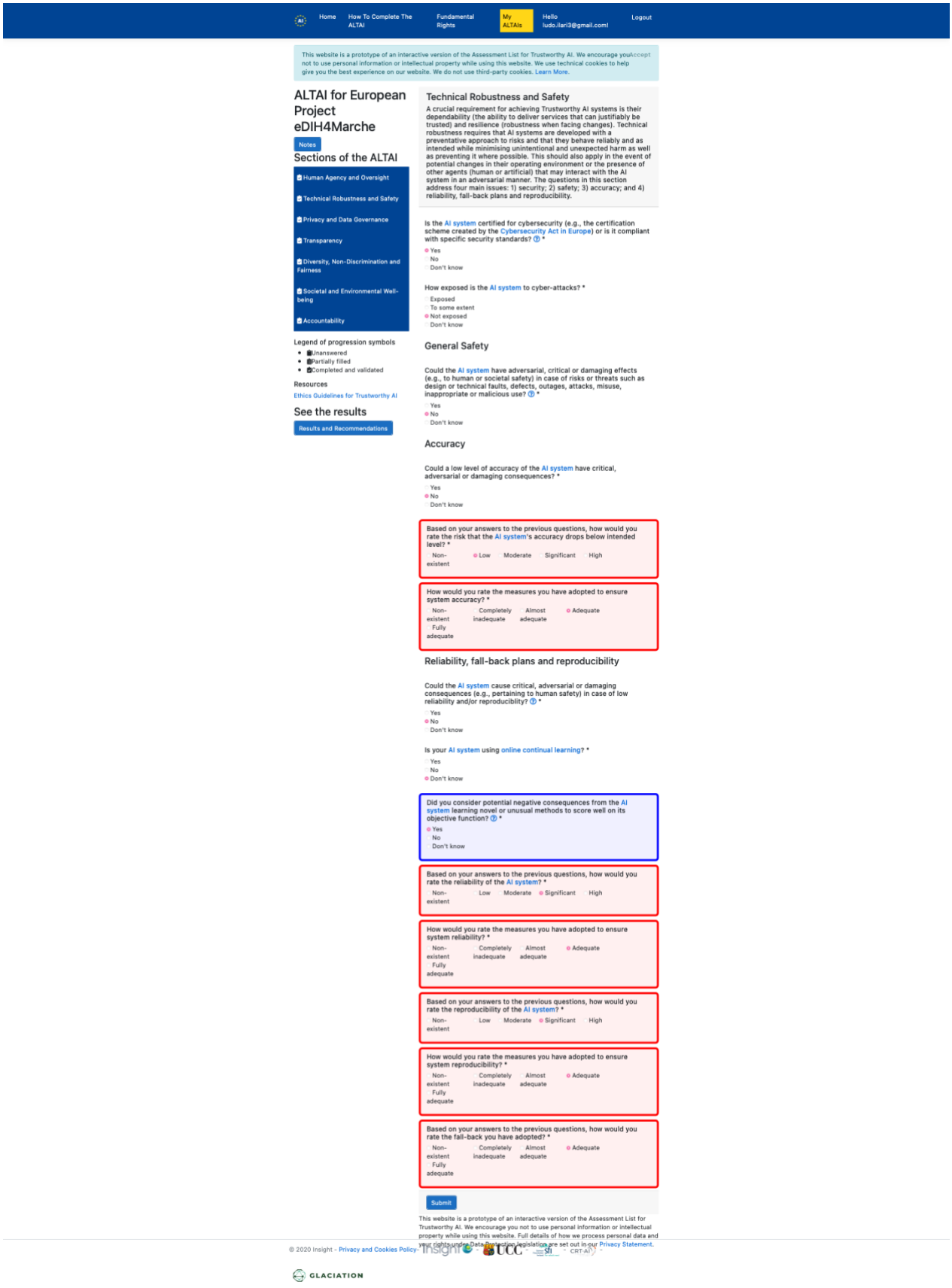

Figure A-9. Example of qualitative method. Self-assessment results for trustworthy artificial intelligence

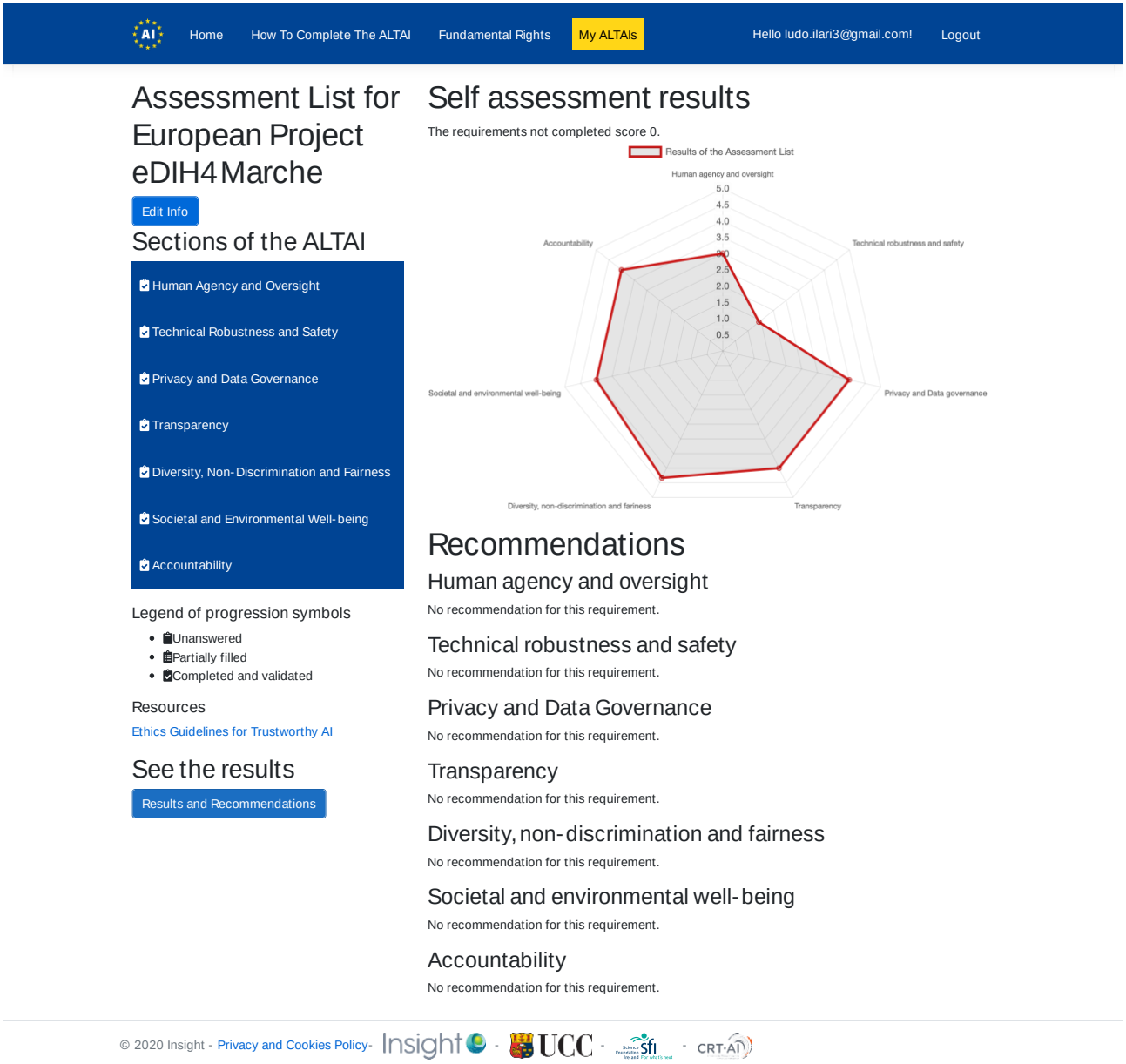

Figure A-10.. Example of semi-quantitative method. Self-assessment results for trustworthy artificial intelligence

Table A-24. Ethics issues table in the proposal of european project

ETHICS ISSUES TABLE

(To be filled in and uploaded as part of the application, until the Ethics Issues Table is available directly in the Submission System.)

Ethics issues

This table should be completed as part of your proposal. Please go through the table and indicate which elements concern your proposal by answering YES or NO.

If you answer YES to any of the questions:

- *indicate in the adjacent box at which page in your full proposal further information relating to that ethics issue can be found, and*
- *provide additional information on this ethics issue in the Ethics self-assessment section below.*

For more information on each of the ethics issues and how to address them, including detailed legal references,see the guidelines How to Complete your Ethics Self-Assessment.

1. Human embryonic stem cells and human embryos		Yes/No	Page
Does this activity involve human embryonic stem cells (hESCs)?		No	
If **YES**:	- Will they be directly derived from embryos within this project?		
	- Are they previously established cells lines? Are the cell lines registered in the European registry for human embryonicstem cell lines?		
Does this activity involve the use of human embryos?		No	
If **YES**:	- Will the activity lead to their destruction?		
2. Humans		Yes/No	Page
Does this activity involve human participants?		Yes	
If **YES**:	- Are they volunteers?	No	
	- Are they healthy volunteers for medical studies?	No	
	- Are they patients for medical studies?	No	
	- Are they potentially vulnerable individuals or groups?	No	
	- Are they children / minors?	No	
	- Are there other persons unable to give informed consent?	No	
Does this activity involve interventions (physical also including imaging technology, behavioural treatments, tracking and tracing, etc) on the study participants?		Yes	
If **YES**:	- Does it involve invasive techniques?	No	
	- Does it involve collection of biological samples?	No	
3. Human cells / tissues		Yes/No	Page
Does this activity involve the use of human cells or tissues (not covered by section 1) **?**		No	

If **YES**:	- Are they human embryonic or foetal cells or tissues?	No	
	- Are they available commercially?	No	
	- Are they obtained within this project?	No	
	- Are they obtained from another project, laboratory or institution?	No	
	- Are they obtained from a biobank?	No	

4. Personal data			Yes/No	Page
Does this activity involve processing of personal data?			Yes	
If **YES:**	- Does it involve the processing of special categories of personal data *(e.g. sexual lifestyle, ethnicity, genetic, biometric and health data, political opinion, religious or philosophical beliefs)*?		No	
	If **YES**:	- Does it involve processing of genetic, biometric or healthdata?	No	
		- Does it involve profiling, systematic monitoring ofindividuals, or processing of large-scale of special categories of data or intrusive methods of data processing (such as, surveillance, geolocation, tracking etc.)?	Yes	
Does this activity involve further processing of previously collected personal data (including use of preexisting data sets or sources, merging existing data sets)?			No	
Is it planned to export personal data from the EU to non-EU countries?			No	
If **YES**:	Specify the type of personal data and countries involved			
Is it planned to import personal data from non-EU countries into the EU or from a non-EU country to another non-EU country?			No	
If **YES**:	Specify the type of personal data and countries involved			
Does this activity involve the processing of personal data related to criminal convictions or offences?			No	

5. Animals	Yes/No	Page
Does this activity involve animals?	No	

6. Non-EU countries		Yes/No	Page
Will some of the activities be carried out in non-EU countries?		No	
If **YES**:	Specify the countries:		
In case non-EU countries are involved, do the activities undertaken in these countries raise potential ethics issues?			

If **YES**:	Specify the countries:		
Is it planned to use local resources (e.g. animal and/or human tissue samples, genetic material, live animals, human remains, materials of historical value, endangered fauna or flora samples, etc.)?		No	
Is it planned to import any material (other than data) from non-EU countries into the EU or from a non-EU country to another non-EU country? *For data imports, see section 4*		No	
If **Yes**:	Specify material and countries involved:		
Is it planned to export any material (other than data) from the EU to non-EU countries? *For data imports, see section 4*		No	
If **Yes**:	Specify material and countries involved:		
7. Environment, health and safety		Yes/No	Page
Does this activity involve the use of substances or processes that may cause harm to the environment, to animals or plants (during the implementation of the activity or further to the use of the results, as a possible impact)?		Yes	

8. Artificial intelligence	Yes/No	Page
Does this activity involve the development, deployment and/or use of Artificial Intelligence-based systems? *If yes, detail in the self-assessment whether that could raise ethical concernsrelated to human rights and values and detail how this will be addressed.*	Yes	
9. Other ethics issues	Yes/No	Page
Are there any other ethics issues that should be taken into consideration?	No	
Please specify:		

Table A-25. The different methods in the ethics self-assessment

Ethics self-assessment *If you have entered any issues in the ethics issue table, you must perform an ethics self-assessment inaccordance with the guidelines How to Complete your Ethics Self-Assessment and complete the table below.*
Ethical dimension of the objectives, methodology and likely impact *Explain in detail the identified issues in relation to:* – *objectives of the activities (e.g. study of vulnerable populations, etc.)* – *methodology (e.g. clinical trials, involvement of children, protection of personal data, etc.)*

– *the potential impact of the activities (e.g. environmental damage, stigmatisation of particular social groups, political or financial adverse consequences, misuse, etc.)*

Objective

The objective of the project concerning activities to support the digitisation, innovation, and technology transfer of enterprises in the fields of Artificial Intelligence (AI) and Cybersecurity (CS) did not encounter significant ethical problems. The individuals involved in the project are not vulnerable, elderly or children and all participants will receive informed consent. Nevertheless, personal data may belong to special categories and be processed on a large scale.

Methodology

The methodology of the project involves the collection, use and management of personal data and the deployment and use of AI technologies in manufacturing, agri-food, tourism and creative, trade and business, people and community services. Ethical issues related to those activities are reported in the “Ethics issues table” and analysed in the following sections.

Personal Data

The processing of personal data has been evaluated using tools, in particular the "**Ethics and data protection decision tree**" tool issued by the European Commission for the analysis of potential ethical aspects related to processing activities and the "Online tool for the security of personal data processing" developed by the European Union Agency for Cybersecurity (ENISA) for the assessment of risk in personal data processing. The risk was assessed "low", as visible in the annex (Figure 1-2). From the risk analysis, all technical and organizational measures were considered to further increase the level of security.

Artificial Intelligence

The deployment and use of AI were assessed using the **Assessment List for Trustworthy Artificial Intelligence tool (ALTAI)**. From this assessment, the project did not find any critical issues; in fact, no recommendations emerged for the seven principles of human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity non-discrimination and fairness, social and environmental well-being, accountability, as can be seen in the annex (Figure 3).

Impact

The adoption of AI and CS could impact on:

- Violation of fundamental rights
- Growing social inequality
- Identifying and tracking individuals
- Possible misuse
- Potential longer-term concerns
- More harmful attacks on the company’s system in case of very limited tools at a CS team’s hands
- Environmental pollution

Compliance with ethical principles and relevant legislation

Describe how the issue(s) identified in the ethics issues table above will be addressed in order to adhere to the ethical principles and what will be done to ensure that the activities are compliant with the E U / national legal and ethical requirements of the country or countries where the

tasks are to be carried out. It is reminded that foractivities performed in a non-EU country, they should also be allowed in at least one EU Member State.

Methodology

Personal Data

The project proposes the following strategy to mitigate the possible risks arising from the processing of personal data. Anonymization and pseudonymization techniques will be used to protect privacy, as well as policies and procedures to ensure the confidentiality, integrity, availability, and resilience of processing systems. The data will be stored securely and deleted after use and in addition, the limited use of the data will be guaranteed. The principle of data minimization as well as the principle of data protection by design and by default will be respected, according to the General Data Protection Regulation (GDPR).

A higher level of security will be ensured for data processing operations that may entail higher ethical risks, especially on the large-scale processing of personal data, on data belonging to special categories and data processed with the use of techniques that are vulnerable to misuse. Moreover, the use of big data, the use of AI techniques and in particular, automated decision-making processes that have a significant impact on the data subject will be treated properly, also considering the guidelines on the protection of individuals with regard to the processing of personal data in a world of big data (Big Data Guidelines).

Therefore, we will encrypt the search data and/or the devices on which it is stored and ensure that the keys/passwords are adequately protected; in addition, the data protection officer (DPO) or an appropriately qualified expert will be consulted for advice on how to achieve a level of data security that is commensurate with the risks to data subjects and for an assessment of the impact on data protection (DPIA). In addition, mathematical and statistical methods based on that type of techniques will be proven and reliable, the data will be as accurate as possible, and the risk of errors or discriminatory effects will be minimized. The transparency and accountability of the decision taken by these algorithms will also be guaranteed, and information will be provided to the interested parties on automated processing, profiling, and evaluation results, as well as on the appeal for the persons concerned to obtain an explanation and contest the decision reached, according to the GDPR.

Finally, technical and organizational measures will be put in place to ensure a level of data security commensurate with the risks to which the interested parties are exposed in the event of unauthorized access or disclosure of their data or their cancellation or accidental destruction (art.32 GDPR).

Artificial Intelligence

Considering the following policies “Ethical guidelines for trustworthy AI”, “Policy and Investment Recommendations for Trustworthy Artificial Intelligence”, and “Artificial Intelligence Act”, technical and non-technical measures will be considered in the implementation aimed at ensuring:

- *Fundamental rights* through an assessment of fundamental rights.
- *Human agency* by providing the knowledge and tools to understand and interact with the AI system to a sufficient extent.
- *Human oversight* through governance mechanisms such as a human-in-the-loop (HITL), human-on-the-loop (HOTL) or human-in-command (HIC) approach.
- *Resilience to attacks and security* bringing together the AI community and the security community and considering the European Union's Coordinated Plan on Artificial Intelligence and possible measures outlined in scientific articles (for example in "Malicious Use of AI, Avin S., Brundage M. et. al., 2018).
- *Fallback plan and general safety* through safety measures that will be developed and tested proactively also with the involvement of a human operator who can interact.
- *Accuracy* through an explicit and well-structured development and evaluation process that can support, mitigate, and correct unwanted risks arising from inaccurate forecasts. When occasional inaccurate predictions cannot be avoided, the system will indicate the likelihood of such errors.
- *Reliability and reproducibility* through the arrangement of a range of inputs involved in a series of situations and the use of "replication files" that will replicate every step of the AI system development process, from research and initial data collection to results.
- *Privacy and data protection* ensured for the entire life cycle through techniques such as k-anonymity and blurring, encryption, scrambling, masking, tokenization for anonymise and pseudonymize personal data, respectively.

- *Quality and integrity of data* using primitives such as cryptographic hash functions (SHA-3).
- *Access to data* with data protocols governing access to data and to qualified personnel.
- *Traceability* of the dataset as well as of the processes of data collection, data labelling and algorithmic selection, by documentation according to the best possible standards.
- *Explainability* of the technical process and related decisions through timely measures adapted to the expertise of stakeholders.
- *Communication* of the level of accuracy, limits, and capabilities of the AI system through information for AI operators or end-users.
- *Avoidance of unfair bias* through processes to analyse and address the system's purpose, constraints, requirements, data collection and decisions.
- *Accessibility and universal* design through user-centred systems designed in a way that allows all people to use AI products or services. Relevant accessibility standards will be considered (EN 301 549).
- *Stakeholder participation* through the collection of regular and long-term feedback from subjects with different backgrounds, cultures, and disciplines.
- *Sustainable and environmentally friendly AI* following relevant standards, guidelines, and sustainable development goals. Moreover, a regular environmental impact assessment will be carried out.
- *Social impact* through regular evaluation of the effects of these systems on the physical and mental well-being of people.
- *Society and democracy* using impact assessments.
- *Auditability* through evaluation by internal and external auditors and availability of such assessment reports.
- *Minimisation and reporting of negative impacts* through impact assessments proportionate to the risk posed by AI systems and consequent protection measures.
- *Trade-offs* by the release of relevant documentation in which trade-offs will be motivated and continuously reviewed. If ethically acceptable compromises cannot be identified, the development, deployment and use of the AI system will not proceed.
- *Redress* through accessible mechanisms ensuring adequate compensation, especially for vulnerable persons.

Impact

The impacts of the activities on the environment, especially for the potential energy consumption and excess CO2 emissions, could be monitored with the environmental impact assessment issued by the European Commission. Regarding the impact of AI and CS systems and the possible misuse, continuous monitoring will be guaranteed, by means of the assessment, to ensure respect for human rights. For CS systems, Codes of conduct will be considered, in particular "ACM code of ethics and professional conduct" and "IEEE Code of Ethics" (IEEE Policies, Section 7 - Professional Activities).

N.	Campo	Descrizione	
1	**ID_REGISTRO**	Identificativo del registro GDPR	INTEGER
2	**DES_REGISTRO**	Descrizione sintetica trattamento	
3	**TIPO_REGISTRO**	Tipo registro	COMBO
	Val.Ammessi	*0: ---* *1: Titolare* *2: Responsabile* *3: Progetti*	
4	**ID_REFERENTE_REGISTRO**	*** Identificativo referente registro *fk su MAN_CLI_FOR *** NON SI USA PIU', SOSTITUITO DA RESPONSABILE ESTERNO SU CONTRATTO*	INTEGER
5	**ID_AZIENDA**	Identificativo Azienda *fk su MTA_AZIENDA*	INTEGER
6	**UTE_MOD**	UTENTE CHE HA EFFETTUATO LA MODIFICA	CHAR(20)
7	**DATA_MOD**	DATA DELLA MODIFICA DEL RECORD	DATE
8	**TIME_MOD**	ORA E MINUTI DELLA MODIFICA DEL RECORD	CHAR(8)
9	**DATA_INIZIO**	Data Inizio	DATE
10	**DATA_FINE**	Data Fine	DATE
11	**STORICIZZATO**	FLAG GENERICO	CHECKBOX
12	**ASSESSMENT_DEFINITIVO**	Flag assessment o definitivo *0: Assessment* *1: Consolidato* *2: Chiuso*	CHECKBOX
13	**FINALITA_TRATTAMENTO**	*** Finalità del trattamento *Non si utilizza più, l'informazione viene salvata in GDP_REGISTRI_NOTE_ESTESE*	
14	**DES_DEL_TRATTAMENTO**	*** Descrizione del trattamento *Non si utilizza più, l'informazione viene salvata in GDP_REGISTRI_NOTE_ESTESE*	
15	**MODELLO_INFORMATIVA**	*** Modello Informativa *Non si utilizza più, aggiunto salvataggio documenti su LTA_DOCUMENTI*	
16	**MODELLO_CONSENSO**	*** Modello Consenso *Non si utilizza più, aggiunto salvataggio documenti su LTA_DOCUMENTI*	
17	**PERIODO_CONSERV**	Periodo Conservazione	INTEGER
18	**UM_TEMPO**	unità di misura tempo *0=Non specificato* *1=Mesi* *2=Anni* *3=Giorni* *4=Settimane*	COMBO
19	**DATA_RIFERIMENTO**	Data riferimento *Attualmente non gestito*	DATE
20	**DOC_GAR_TRASF**	*** Garanzie ai sensi dell'art. 46 (garanzie adeguate) *Non si utilizza più, sostituito da elenco movimenti in GDP_REGISTRI_VALUTAZ (tipologia 21)*	
21	**NOTE_PATH**	Path o directory *Attualmente non gestito*	
22	**RESPONSABILE_TECNICO**	Responsabile tecnico *Attualmente non gestito* *fk su MAN_CLI_FOR*	INTEGER
23	**NOTE**	Note	
24	**NOTE_TRATTAMENTO**	Note trattamento	
25	**NOTE_INFORMATIVA**	Note informativa	

Figure A-11. GDP_REGISTRI: GDP_data processing registers (first part)

N.	Campo	Descrizione	
26	**NOTE_CONSENSO**	Note Consenso	
27	**NOTE_REPERIM_DATI**	Note reperimento dati	
28	**NOTE_CONSERVAZIONE**	Note conservazione dati	
29	**LISTA_DEST_COMUNIC**	Lista destinatari comunicazione	
30	**LISTA_DEST_TRASF**	Lista destinatari del trasferimento	
31	**DEROGHE_TRASF**	*** Deroghe trasferimento *Non si utilizza più, sostituito da elenco movimenti in GDP_REGISTRI_VALUTAZ (tipologia 22)*	
32	**FINALITA_TRASF**	Finalità del trasferimento	
33	**ID_TIPO_CONTESTO**	Id tipo contesto *fk su GDP_TIPI_DI_CONTESTO*	INTEGER
34	**PERC_GEST_AUTO**	Percentuale impatto gestione automatizzata *La percentuale indica l'impatto della gestione automatizzata, se non specificata si assume che la gestione automatizzata ha lo stesso impatto della gestione manuale.*	PERCENTUALE
35	**ID_REGISTRO_INIZ**	Codice registro inziale *Codice registro nella versione iniziale di inserimento, ovvero il primo inserito.* *Viene valorizzato nei movimenti generati a seguito di storicizzazione o di DPIA*	INTEGER
36	**ID_REGISTRO_PADRE**	Codice registro padre *Codice del trattamento padre (versione precedente), consente di ricostruire la catena delle operazioni fatte a partire dalla prima versione del trattamento* *Viene valorizzato nei movimenti generati a seguito di storicizzazione o di DPIA*	INTEGER
37	**TIPO_RECORD**	Tipo record *0=Trattamento* *1=Copia movimento prima di rettifiche da DPIA* *Attualmente non gestito*	CHECKBOX
38	**STATO_RISCHIO_FINALE**	Stato indice di rischio finale *0=Accettabile* *1=Non Accettabile* *Viene calcolato a partire dal più alto dei coefficienti di rischio finale.* *La relativa descrizione da visualizzare/stampare viene ereditata dal corrispondente campo in GDP_VALORI_SOGLIA*	CHECKBOX
39	**VALORE_RISCHIO_FINALE_R**	Valore rischio finale Riservatezza	
40	**ID_RISCHIO_FINALE_R**	Id descrizione rischio finale Riservatezza *fk su GDP_VALORI_SOGLIA*	INTEGER
41	**VALORE_RISCHIO_FINALE_I**	Valore rischio finale integrità	
42	**ID_RISCHIO_FINALE_I**	Id descrizione rischio finale integrità *fk su GDP_VALORI_SOGLIA*	INTEGER
43	**VALORE_RISCHIO_FINALE_D**	Valore rischio finale integrità	
44	**ID_RISCHIO_FINALE_D**	Id descrizione rischio finale Disponibilità *fk su GDP_VALORI_SOGLIA*	INTEGER
45	**PRESENZA_ELEMENTI_RISCHIO**	Flag presenza elementi ad altro rischio *0=No* *1=Si* *Viene impostato ad 1 se almeno 2 dei criteri di valutazione sintetica di necessità di DPIA è uguale a SI*	CHECKBOX
46	**FLAG_POSIZIONE_DPIA**	Flag Posizione DPIA *Il flag assume i seguenti valori:* *0=Non necessaria/Da eseguire (Se STATO_RISCHIO_FINALE= 1 o PRESENZA_ELEMENTI_RISCHIO = 1)* *1=Eseguita (valore associato alla versione originale del trattamento associato a DPIA che in questo modo viene bloccato, se generata anche la nuova versione del trattamento, quella originale viene anche storicizzata)* *2=Trattamento di rettifica generato dalla DPIA (per le successive modifiche il presente trattamento verrà sempre storicizzato e andrà creata una nuova versione)*	CHECKBOX

Figure A-12. GDP_REGISTRI: GDP_data processing registers (second part)

N.	Campo	Descrizione	
47	ID_REGISTRO_CONFRONTO	*** Id Registro per confronto *Il campo non viene mai valorizzato, serve solo per evitare che il framework, all'apertura della form di dettaglio, esegua una lettura che va in errore*	INTEGER
48	ID_STRUT_TRATT_RISCHIO	*** Id profilo di rischio *Il campo non viene mai valorizzato, serve solo per evitare che il framework, all'apertura della form di dettaglio, esegua una lettura che va in errore*	INTEGER
49	ID_REGISTRO_STORICO	*** Id Registro da storico *Il campo non viene mai valorizzato, serve solo per evitare che il framework, all'apertura della form di dettaglio, esegua una lettura che va in errore*	INTEGER
50	ID_SERVIZIO_TREE_RICERCA	*** ID tassonomia *Il campo non viene mai valorizzato, serve solo per evitare che il framework, all'apertura della form di dettaglio, esegua una lettura che va in errore.* *Viene utilizzato per la selezione della tassonomia se STRUTTURA_TRATTAMENTO=1 in quanto rispetto al codice selezionato vanno aggiunti alla tabella le tassonomie di livello successivo*	INTEGER
51	FLAG_PRESENZA_RESP_ESTERNO	Flag presenza responsabile esterno	CHECKBOX
52	FLAG_DATI_DA_COMPLETARE_1	Flag presenza dati da completare art.30	CHECKBOX
53	FLAG_DATI_DA_COMPLETARE_2	Flag presenza dati da completare rischio	CHECKBOX
54	STRUTTURA_TRATTAMENTO	Struttura trattamento	COMBO
	Val.Ammessi	*0=Standard* *1=Trasversale (lo stesso trattamento viene associato a tassonomie diverse generando quindi una duplicazione virtuale). Sono quei trattamenti svolti con le stesse caratteristiche da settori diversi e che per semplicità vengono accorpati in un unico movimento* *2=Flusso (La gestione è identica a quella dei trattamenti trasversali, ma in questo caso il trattamento è unico e viene svolto in quota parte da 2 o più tassonomie. Cambia la modalità di stampa)*	
55	ID_TIPI_REGISTRO	Id Tipo registro *fk su S_GDP_TIPI_REGISTRI* *gestito solo per i trattamenti degli Asset*	INTEGER
56	ID_BASE_DATI	*** Id Base dati *fk su GDP_BASE_DATI* *Il campo non viene mai valorizzato, serve solo per evitare che il framework, all'apertura della form di dettaglio, esegua una lettura che va in errore*	INTEGER
57	VAL_RISCHIO_INERENTE_R	Valore rischio inerente Riservatezza	
58	ID_RISCHIO_INERENTE_R	Id descrizione rischio inerente Riservatezza *fk su GDP_VALORI_SOGLIA*	INTEGER
59	VAL_RISCHIO_INERENTE_I	Valore rischio inerente Integrità	
60	ID_RISCHIO_INERENTE_I	Id descrizione rischio inerente Integrità *fk su GDP_VALORI_SOGLIA*	INTEGER
61	VAL_RISCHIO_INERENTE_D	Valore rischio inerente Disponibilità	
62	ID_RISCHIO_INERENTE_D	Id descrizione rischio inerente Disponibilità *fk su GDP_VALORI_SOGLIA*	INTEGER
63	NUMERO_VERSIONE	Numero Versione del trattamento	INTEGER
64	DATA_PROSSIMA_REVISIONE	Data Prevista prossima revisione	DATE
65	FLAG_COMUNICAZIONE_ALTRI_DEST	Flag comunicazione altri destinatari *Se uguale ad 1 vengono abilitati i campi "Categorie di destinatari" e "Lista Sintetica dei Destinatari della Comunicazione"*	CHECKBOX

Figure A-13. GDP_REGISTRI: GDP_data processing registers (third part)

N.	Campo	Descrizione	
1	**ID_REGISTRO_VALUTAZ**	Identificativo	INTEGER
2	**ID_REGISTRO**	fk su GDP_REGISTRI	INTEGER
3	**TIPOLOGIA**	Tipologia	COMBO
	Val.Ammessi	*1 - Categorie di trattamento* *4 - Frequenza di trattamento* *5 - Quantità di dati trattati* *6 - Condizione di liceità del trattamento* *7 - Categorie dati personali* *8 - Categorie di interessati* *9 - Fascia d'età degli interessati* *12 - Categorie di destinatari a cui i dati personali saranno sono stati o saranno comunicati* *13 - Trasferimento verso paesi terzi o organizzazioni internazionali* *18 - Probabilità di concorrenza delle minacce* *19 - Numerosità di Interessati* *21 - Trasferimenti - Garanzie ai sensi dell'art. 46 del RGPD* *22 - Trasferimenti - deroghe art. 46* *24 - Finalità del Trattamento*	
4	**R**	R = Riservatezza	
5	**I**	I = Integrità	
6	**D**	D = Disponibilità	
7	**ID_STRUT_TRATT_RISCHIO**	fk su GDP_STRUTTRA_TRATT_RISCHIO	INTEGER
8	**ID_BASE_DATI_PROV**	id base dati di provenienza *fk su GDP_BASE_DATI.*	INTEGER
9	**FLAG_DISABILITATO**	Disabilitato *Gestito per i soli movimenti ereditati da una base dati*	CHECKBOX
10	**UTE_MOD**	UTENTE CHE HA EFFETTUATO LA MODIFICA	CHAR(20)
11	**DATA_MOD**	DATA DELLA MODIFICA DEL RECORD	DATE
12	**TIME_MOD**	ORA E MINUTI DELLA MODIFICA DEL RECORD	CHAR(8)

Figure A-14. Table 0 4. GDP_REGISTRI_VALUTAZ: evaluation of data processing registers

N.	Campo	Descrizione	
1	**ID_STRUT_TRATT_RISCHIO**	identificativo del doc del servizio	INTEGER
2	**DESCRIZIONE**	Descrizione	
3	**TIPOLOGIA**	Tipologia	COMBO
	Val.Ammessi	*1 - Categorie di trattamento* *4 - Frequenza di trattamento* *5 - Quantità di dati trattati* *6 - Condizione di liceità del trattamento* *7 - Categorie dati personali* *8 - Categorie di interessati* *9 - Fascia d'età degli interessati* *12 - Categorie di destinatari a cui i dati personali saranno sono stati o saranno comunicati* *13 - Trasferimento verso paesi terzi o organizzazioni internazionali* *18 - Probabilità di concorrenza delle minacce* *19 - Numerosità di Interessati* *21 - Trasferimenti - Garanzie ai sensi dell'art. 46 del RGPD* *22 - Trasferimenti - deroghe art. 46* *24 - Finalità del Trattamento*	
4	**R**	R = Riservatezza	
5	**I**	I = Integrità	
6	**D**	D = Disponibilità	
7	**TIPO_IMPATTO**	Tipo Impatto	COMBO
	Val.Ammessi	*1 - Gravità Intrinseca* *2 - Amplificazione della Gravità Intrinseca (Da 1 a 10)* *3 - Amplificazione della Probabilità (%)*	
8	**NOTE**	Note	CHAR(1000)
9	**ID_TIPO_CONTESTO**	Id tipo contesto *Fk su GDP_TIPI_DI_CONTESTO*	INTEGER
10	**ID_RIFERIMENTO_NORM**	id riferimento normativo	INTEGER
11	**FLAG_FATT_AMPL**	Previsto fattore di amplificazione *Gestito solo se TIPOLOGIA 19=Numerosità di interessati* *** Per ora non gestito*	CHECKBOX
12	**FATTORE_AMPLIFICAZIONE_DB**	Presenza fattore amplificazione data Breach *Gestita per le tipologie:* *5 - Quantità di dati trattati* *0=Non presente* *1=Alta* *2=Molto Alta* *8 - Categorie di interessati* *1=Categoria Potenzialmente Vulnerabile* *9 - Fascia d'età degli interessati* *1=Categoria Potenzialmente Vulnerabile* *19 - Numerosità di Interessati* *1=Alta Numerosità* *2=Larga Scala*	COMBO
13	**CODICE_GARANTE_DATA_BREACH**	Codifica del garante per comunicazione data breach *Gestita per le tipologie:*	COMBO

Figure A-15. GDP_STRUTTRA_TRATT_RISCHIO: treatment structure and risk (first part)

N.	Campo	Descrizione	
		7 - Categorie dati personali *8 - Categorie di interessati*	
14	**FLAG_TIPOLOGIA13_SI**	*Codifica del garante indicata nelle istruzioni per la comunicazione di data breach* Flag tipologia 13 impostato a SI *Gestito solo per la TIPOLOGIA 13=Trasferimento verso paesi terzi o organizzazioni internazionali.* *Viene utilizzato nei trattamenti per stabilire se è obbligatorio specificare i movimenti con tipologia:* *21 - Trasferimenti - Garanzie ai sensi dell'art. 46 del RGPD* *22 - Trasferimenti - deroghe art. 46*	CHECKBOX
15	**UTE_MOD**	UTENTE CHE HA EFFETTUATO LA MODIFICA	CHAR(20)
16	**DATA_MOD**	DATA DELLA MODIFICA DEL RECORD	DATE
17	**TIME_MOD**	ORA E MINUTI DELLA MODIFICA DEL RECORD	CHAR(8)
18	**ORDINAMENTO**	Ordinamento	INTEGER

Figure A-16. GDP_STRUTTRA_TRATT_RISCHIO: treatment structure and risk (second part)

N.	Campo	Descrizione	
1	**ID_REGISTRO_VALUTAZ**	Identificativo	INTEGER
2	**ID_REGISTRO**	fk su GDP_REGISTRI	INTEGER
3	**TIPOLOGIA**	Tipologia	COMBO
	Val.Ammessi	*1 - Categorie di trattamento* *4 - Frequenza di trattamento* *5 - Quantità di dati trattati* *6 - Condizione di liceità del trattamento* *7 - Categorie dati personali* *8 - Categorie di interessati* *9 - Fascia d'età degli interessati* *12 - Categorie di destinatari a cui i dati personali saranno sono stati o saranno comunicati* *13 - Trasferimento verso paesi terzi o organizzazioni internazionali* *18 - Probabilità di concorrenza delle minacce* *19 - Numerosità di Interessati* *21 - Trasferimenti - Garanzie ai sensi dell'art. 46 del RGPD* *22 - Trasferimenti - deroghe art. 46* *24 - Finalità del Trattamento*	
4	**R**	R = Riservatezza	
5	**I**	I = Integrità	
6	**D**	D = Disponibilità	
7	**ID_STRUT_TRATT_RISCHIO**	fk su GDP_STRUTTRA_TRATT_RISCHIO	INTEGER
8	**ID_BASE_DATI_PROV**	id base dati di provenienza *fk su GDP_BASE_DATI.*	INTEGER
9	**FLAG_DISABILITATO**	Disabilitato *Gestito per i soli movimenti ereditati da una base dati*	CHECKBOX
10	**UTE_MOD**	UTENTE CHE HA EFFETTUATO LA MODIFICA	CHAR(20)
11	**DATA_MOD**	DATA DELLA MODIFICA DEL RECORD	DATE
12	**TIME_MOD**	ORA E MINUTI DELLA MODIFICA DEL RECORD	CHAR(8)

Figure A-17. GDP_REGISTRI_VAL_SINT_DPIA: summary assessment of the need for DPIA

N.	Campo	Descrizione	
1	**ID_SERVIZIO_TREE**	identificativo del servizio tree	INTEGER
2	**COD_NODO**	codice del nodo	
3	**NOME_NODO**	nome o descrizione del nodo *La lunghezza effettiva del campo rimane sempre 255, la parte restante viene utilizzata per accodare alla descrizione:* *- Cognome e nome Designato* *- Descrizione Tag (nei record dei trattamenti. TIPO_RECORD=1)*	
4	**FLAG_IS_FOGLIA**	1=indica che il nodo dell'albero è una foglia	CHECKBOX
	Val.Ammessi	*0: il nodo dell'albero non è una foglia* *1: è una foglia*	
5	**ID_PADRE**	Id padre	INTEGER
6	**LIVELLO**	Livello nella treelist	INTEGER
7	**ORDINAMENTO**	Ordinamento	INTEGER
	Val.Ammessi	*Dafault: ascendente*	
8	**ID_CLI_FOR_RESPONS**	*** Identificativo persona responsabile *Indica la persona fisica responsabile del trattamento.* *Non si utilizza più, sostituito da GDP_SERVIZI_DESIGNATI*	INTEGER
9	**ID_ARTICOLO_SOFTWARE**	*** Identicativo Articolo software *fk su MAN_ARTICOLI - Non si usa più, sostituito da ID_ASSET*	INTEGER
10	**ID_AZIENDA**	Identificativo Azienda *fk su MTA_AZIENDA*	INTEGER
11	**TIPO_SERVIZIO**	Tipo Servizio	COMBO
	Val.Ammessi	*0: ---* *1: Business Service Catalog* *2: Internal Service Catalog*	
12	**UTE_MOD**	UTENTE CHE HA EFFETTUATO LA MODIFICA	CHAR(20)
13	**DATA_MOD**	DATA DELLA MODIFICA DEL RECORD	DATE
14	**TIME_MOD**	ORA E MINUTI DELLA MODIFICA DEL RECORD	CHAR(8)
15	**DATA_INIZIO**	Data Inizio	DATE
16	**DATA_FINE**	Data Fine	DATE
17	**DES_SERVIZIO**	Note Servizio	
18	**ASSESSMENT_DEFINITIVO**	*** Flag assessment o definitivo *0: Assessment* *1: Definitivo*	CHECKBOX
19	**ID_ASSET**	*** Id software *fk su GDP_ASSET (tipo_asset 1=software).* *Non si utilizza più*	INTEGER
20	**TIPO_RECORD**	Tipo Record *0=Tassonomia* *1=Trattamento* *2=Tag*	COMBO
21	**ID_REGISTRO**	Id trattamento *fk su GDP_REGISTRI* *Valorizzato solo se TIPO_RECORD=1*	INTEGER
22	**ID_TAG_TREE**	Id Tag *fk su GDP_TAG*	INTEGER

Figure A-18. GDP_REGISTRI_SERVIZI: list of services related to the GDPR nth register (first part)

N.	Campo	Descrizione	
23	**ID_TIPI_REGISTRO**	*Valorizzato solo se TIPO_RECORD=2* Id Tipo di registro Asset *fk su GDP_TIPI_REGISTRI*	INTEGER
24	**ID_TASSONOMIA**	*Gestito solo per TIPO_SERVIZIO = 3 (asset), il record viene creato/aggiornato automaticamente dalla tabella dei tipi registro* Id del servizio *Id della tassonomia a cui il trattamento (ed i relativi tag) è associato.* *Nel caso dei trattamenti è uguale a ID_PADRE, ma risolve il problema del filtro sulle autorizzazioni per i record dei tag.*	INTEGER
25	**ALIAS_TASSONOMIA**	*Valorizzato solo se TIPO_RECORD=1 o 2* alias Tassonomia	CHAR(5)

Figure A-19. GDP_REGISTRI_SERVIZI: list of services related to the GDPR nth register (second part)

N.	Campo	Descrizione	
1	**ID_REGISTRO_MINACCE**	Identificativo	INTEGER
2	**ID_REGISTRO**	fk su GDP_REGISTRI	INTEGER
3	**TIPO_PROCESSO**	Tipo processo di gestione dati	COMBO
	Val.Ammessi	*1=Automatizzato* *2=Manuale*	
4	**ID_MODELLO_VAL**	Id modello valutazione minacce *fk su GDP_MODELLO_VAL*	INTEGER
5	**R**	R = Riservatezza	
6	**I**	I = Integrità	
7	**D**	D = Disponibilità	
8	**ID_BASE_DATI**	Id base dati *fk su GDP_BASE_DATI*	INTEGER
9	**UTE_MOD**	UTENTE CHE HA EFFETTUATO LA MODIFICA	CHAR(20)
10	**DATA_MOD**	DATA DELLA MODIFICA DEL RECORD	DATE
11	**TIME_MOD**	ORA E MINUTI DELLA MODIFICA DEL RECORD	CHAR(8)

Figure A-20. GDP_REGISTRI_MINACCE: threat probability assessment

N.	Campo	Descrizione	
1	**ID_REGISTRO_MISURE_MIT**	Identificativo	INTEGER
2	**ID_REGISTRO**	fk su GDP_REGISTRI	INTEGER
3	**GRUPPO_MISURA**	Gruppo misure di mitigazione rischio	COMBO
	Val.Ammessi	*1=Generale* *2=Specifica*	
4	**ID_MISURE_MIT**	Id misura mitigazione *fk su GDP_MISURE_MIT*	INTEGER
5	**R**	R = Riservatezza	
6	**I**	I = Integrità	
7	**D**	D = Disponibilità	
8	**UTE_MOD**	UTENTE CHE HA EFFETTUATO LA MODIFICA	CHAR(20)
9	**DATA_MOD**	DATA DELLA MODIFICA DEL RECORD	DATE
10	**TIME_MOD**	ORA E MINUTI DELLA MODIFICA DEL RECORD	CHAR(8)

Figure A-21. GDP_REGISTRI_MISURE_MIT: risk mitigation measures

N.	Campo	Descrizione	
1	**ID_REGISTRO_CALCOLI**	Identificativo	INTEGER
2	**ID_REGISTRO**	fk su GDP_REGISTRI	INTEGER
3	**TIPO_TOTALE**	Tipo totale	INTEGER
4	**DES_TOTALE**	Descrizione tipo totale	
5	**R**	R = Riservatezza	
6	**I**	I = Integrità	
7	**D**	D = Disponibilità	
8	**NOTE_TOTALE**	Note	
9	**TIPOLOGIA**	Tipologia	COMBO
	Val.Ammessi	*1 - Categorie di trattamento* *4 - Frequenza di trattamento* *5 - Quantità di dati trattati* *6 - Condizione di liceità del trattamento* *7 - Categorie dati personali* *8 - Categorie di interessati* *9 - Fascia d'età degli interessati* *12 - Categorie di destinatari a cui i dati personali saranno sono stati o saranno comunicati* *13 - Trasferimento verso paesi terzi o organizzazioni internazionali* *18 - Probabilità di concorrenza delle minacce* *19 - Numerosità di Interessati* *21 - Trasferimenti - Garanzie ai sensi dell'art. 46 del RGPD* *22 - Trasferimenti - deroghe art. 46* *24 - Finalità del Trattamento*	
10	**UTE_MOD**	UTENTE CHE HA EFFETTUATO LA MODIFICA	CHAR(20)
11	**DATA_MOD**	DATA DELLA MODIFICA DEL RECORD	DATE
12	**TIME_MOD**	ORA E MINUTI DELLA MODIFICA DEL RECORD	CHAR(8)

Figure A-22. GDP_REGISTRI_CALCOLI: detail of calculations

B. Appendix: R Code for data pre-processing

This appendix includes the R code developed and used for the pre-processing step of the business intelligence process presented in this study. The code provides detailed scripts for the reproducibility and transparency of the results. Each script is organized according to the specific analytical tasks, enabling users to replicate or adapt the methodology for similar research purposes.

####################################**Data Collection**####################################

```
install.packages("pbixr")
install.packages("tidyverse")
install.packages("writexl")
install.packages("DBI")
install.packages("odbc")
install.packages("purrr")

library(pbixr)
library(dplyr)
library(tidyverse)

library(readxl)
library (writexl)

library(DBI)
library(odbc)
library (purrr)

conn <- DBI::dbConnect(
  odbc::odbc(),
  Driver = "SQL Server",
  Server = "XXXXXX",
  Database = "XXXXXXXXXXXX",
  uid = "XXX",
```

```
  pwd = "XX",
  #options(connectionObserver = NULL)
)

query1="SELECT * FROM GDP_REGISTRI"
GDP_REGISTRI=dbGetQuery(conn, query1)

query2="SELECT * FROM GDP_REGISTRI_VALUTAZ"
GDP_REGISTRI_VALUTAZ=dbGetQuery(conn, query2)

query3="SELECT * FROM GDP_STRUTTRA_TRATT_RISCHIO"
GDP_STRUTTRA_TRATT_RISCHIO=dbGetQuery(conn, query3)

query4="SELECT * FROM GDP_REGISTRI_VAL_SINT_DPIA"
GDP_REGISTRI_VAL_SINT_DPIA=dbGetQuery(conn, query4)

query5="SELECT * FROM GDP_REGISTRI_SERVIZI"
GDP_REGISTRI_SERVIZI=dbGetQuery(conn, query5)

query6="SELECT * FROM GDP_REGISTRI_MINACCE"
GDP_REGISTRI_MINACCE=dbGetQuery(conn, query6)

query7="SELECT * FROM GDP_REGISTRI_MISURE_MIT"
GDP_REGISTRI_MISURE_MIT=dbGetQuery(conn, query7)

query8="SELECT * FROM GDP_REGISTRI_CALCOLI"
GDP_REGISTRI_CALCOLI=dbGetQuery(conn, query8)

####################################Data Storage#####################################
write_xlsx(GDP_REGISTRI,"C:/Users/ludovica.ilari/OneDrive - Università degli Studi di
Macerata/dati jobsdp.Misti/temp.08.04.24.versione2/GDP_REGISTRI.xlsx")
```

```
write_xlsx(GDP_REGISTRI_VALUTAZ,"C:/Users/ludovica.ilari/OneDrive - Università degli Studi di Macerata/dati jobsdp.Misti/temp.08.04.24.versione2/GDP_REGISTRI_VALUTAZ.xlsx")
write_xlsx(GDP_STRUTTRA_TRATT_RISCHIO,"C:/Users/ludovica.ilari/OneDrive - Università degli Studi di Macerata/dati jobsdp.Misti/temp.08.04.24.versione2/GDP_STRUTTRA_TRATT_RISCHIO.xlsx")
write_xlsx(GDP_REGISTRI_VAL_SINT_DPIA,"C:/Users/ludovica.ilari/OneDrive - Università degli Studi di Macerata/dati jobsdp.Misti/temp.08.04.24.versione2/GDP_REGISTRI_VAL_SINT_DPIA.xlsx")
write_xlsx(GDP_REGISTRI_SERVIZI,"C:/Users/ludovica.ilari/OneDrive - Università degli Studi di Macerata/dati jobsdp.Misti/temp.08.04.24.versione2/GDP_REGISTRI_SERVIZI.xlsx")
write_xlsx(GDP_REGISTRI_MINACCE,"C:/Users/ludovica.ilari/OneDrive - Università degli Studi di Macerata/dati jobsdp.Misti/temp.08.04.24.versione2/GDP_REGISTRI_MINACCE.xlsx")
write_xlsx(GDP_REGISTRI_MISURE_MIT,"C:/Users/ludovica.ilari/OneDrive - Università degli Studi di Macerata/dati jobsdp.Misti/temp.08.04.24.versione2/GDP_REGISTRI_MISURE_MIT.xlsx")
write_xlsx(GDP_REGISTRI_CALCOLI,"C:/Users/ludovica.ilari/OneDrive - Università degli Studi di Macerata/dati jobsdp.Misti/temp.08.04.24.versione2/GDP_REGISTRI_CALCOLI.xlsx")

####################################Data Analisys####################################
head(GDP_REGISTRI)

GDP_REGISTRI_new = subset(GDP_REGISTRI, select = -c(ID_REFERENTE_REGISTRO,
ID_AZIENDA,UTE_MOD, DATA_MOD, FINALITA_TRATTAMENTO, FINALITA_TRASF,
TIME_MOD, DATA_RIFERIMENTO, DOC_GAR_TRASF, RESPONSABILE_TECNICO,
LISTA_DEST_COMUNIC, LISTA_DEST_TRASF, DEROGHE_TRASF,
FLAG_PRESENZA_RESP_ESTERNO, FLAG_DATI_DA_COMPLETARE_1,
FLAG_DATI_DA_COMPLETARE_2,FLAG_COMUNICAZIONE_ALTRI_DEST,
ID_STRUT_TRATT_RISCHIO))

GDP_REGISTRI_VALUTAZ_new = subset(GDP_REGISTRI_VALUTAZ, select = -c(TIME_MOD,
ID_STRUT_TRATT_RISCHIO_5))
```

```
GDP_STRUTTURA_TRATT_RISCHIO_new= subset(GDP_STRUTTRA_TRATT_RISCHIO, select = -
c(NOTE,UTE_MOD,TIME_MOD))

GDP_REGISTRI_VAL_SINT_DPIA_new= subset(GDP_REGISTRI_VAL_SINT_DPIA, select = -
c(NOTE,UTE_MOD, DATA_MOD, TIME_MOD))
GDP_REGISTRI_MINACCE_new = subset(GDP_REGISTRI_MINACCE, select = -c(UTE_MOD,
TIME_MOD))
GDP_REGISTRI_REGISTRI_CALCOLI_new= subset(GDP_REGISTRI_REGISTRI_CALCOLI, select = -
c(UTE_MOD, TIME_MOD))
GDP_REGISTRI_MISURE_MIT_new= subset(GDP_REGISTRI_MISURE_MIT, select = -c(UTE_MOD,
TIME_MOD))

################################################################################

numero.di.tipologie=c(1,4,5,6,7,8,9,12,13,18,19,21,22,24)
#CC=c(11:25)
CC=c(17:30)

V=c('CategorieDiTrattamento', 'FrequenzaDiTrattamento', 'QuantitàDiDatiTrattati',
'CondizioneDiLiceitàDelTrattamento', 'CategorieDiDatiPersonali', 'CategorieDiInteressati',
'FasciaDiEtàDEgliInteressati',
   'CategorieDiDestinatari', 'Trasferimento', 'ProbabilitàDiConcorrenzaDelleMinacce',
'NumerositàDiInteressati', 'TrasferimentiGaranzie', 'TrasferimentiDeroghe',
'FinalitàDelTrattamento')

V[1]
GDP_STRUTTURA_TRATT_RISCHIO_new[c('CategorieDiTrattamento', 'FrequenzaDiTrattamento',
'QuantitàDiDatiTrattati',
'CondizioneDiLiceitàDelTrattamento', 'CategorieDiDatiPersonali', 'CategorieDiInteressati',
'FasciaDiEtàDEgliInteressati','CategorieDiDestinatari', 'Trasferimento',
'ProbabilitàDiConcorrenzaDelleMinacce', 'NumerositàDiInteressati', 'TrasferimentiGaranzie',
'TrasferimentiDeroghe', 'FinalitàDelTrattamento')] <- NA
```

```
for (u in 1:(nrow(GDP_STRUTTURA_TRATT_RISCHIO_new))) {
  for (C in 1:(length(V))) {
if (GDP_STRUTTURA_TRATT_RISCHIO_new$TIPOLOGIA[u]==numero.di.tipologie[C]) {
#GDP_REGISTRI_VALUTAZ_new$tipologia1[u]=1
#GDP_REGISTRI_VALUTAZ_new[u,CC[C]]=numero.di.tipologie[C]
GDP_STRUTTURA_TRATT_RISCHIO_new[u,CC[C]]=GDP_STRUTTURA_TRATT_RISCHIO_new$DESCRI
ZIONE[u]

   }}
}

################################################################################

# Combinazione delle 3 tabelle: la prima combinazione "Combinazione1" è tra "ID_REGISTRO." e
"ID_STRUT_TRATT_RISCHIO.". Combinazione2 è tra l'uscita della precedente combinazione
(combinazione 1) e la tabella "GDP_STRUTTURA_TRATT_RISCHIO_new".
Combinazione1 <-
full_join(GDP_REGISTRI_new,GDP_REGISTRI_VALUTAZ_new,by=c("ID_REGISTRO"))

for (t in 1:(max(Combinazione1$ID_REGISTRO))){
 for (i in 1:nrow(Combinazione1)){
if (Combinazione1$ID_REGISTRO[i]==t) {
Combinazione1$DES_REGISTRO[i]=GDP_REGISTRI_new$DES_REGISTRO[which(GDP_REGISTRI_ne
w$ID_REGISTRO==t)]

Combinazione1$VALORE_RISCHIO_FINALE_R[i]=GDP_REGISTRI_new$VALORE_RISCHIO_FINALE_R[
which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$VALORE_RISCHIO_FINALE_I[i]=GDP_REGISTRI_new$VALORE_RISCHIO_FINALE_I[
which(GDP_REGISTRI_new$ID_REGISTRO==t)]
```

Combinazione1$VALORE_RISCHIO_FINALE_D[i]=GDP_REGISTRI_new$VALORE_RISCHIO_FINALE_D[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$VAL_RISCHIO_INERENTE_R[i]=GDP_REGISTRI_new$VAL_RISCHIO_INERENTE_R[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$VAL_RISCHIO_INERENTE_I[i]=GDP_REGISTRI_new$VAL_RISCHIO_INERENTE_I[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$VAL_RISCHIO_INERENTE_D[i]=GDP_REGISTRI_new$VAL_RISCHIO_INERENTE_D[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$TIPO_REGISTRO[i]=GDP_REGISTRI_new$TIPO_REGISTRO[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$DATA_INIZIO[i]=GDP_REGISTRI_new$DATA_INIZIO[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$DATA_FINE[i]=GDP_REGISTRI_new$DATA_FINE[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$STORICIZZATO[i]=GDP_REGISTRI_new$STORICIZZATO[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$PRESENZA_ELEMENTI_RISCHIO[i]=GDP_REGISTRI_new$PRESENZA_ELEMENTI_RISCHIO[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$STRUTTURA_TRATTAMENTO[i]=GDP_REGISTRI_new$STRUTTURA_TRATTAMENTO[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

```
Combinazione1$NUMERO_VERSIONE[i]=GDP_REGISTRI_new$NUMERO_VERSIONE[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$ASSESSMENT_DEFINITIVO[i]=GDP_REGISTRI_new$ASSESSMENT_DEFINITIVO[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$ID_REGISTRO_INIZ[i]=GDP_REGISTRI_new$ID_REGISTRO_INIZ[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$ID_REGISTRO_PADRE[i]=GDP_REGISTRI_new$ID_REGISTRO_PADRE[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$STATO_RISCHIO_FINALE[i]=GDP_REGISTRI_new$STATO_RISCHIO_FINALE[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$FLAG_POSIZIONE_DPIA[i]=GDP_REGISTRI_new$FLAG_POSIZIONE_DPIA[which(GDP_REGISTRI_new$ID_REGISTRO==t)]

Combinazione1$DATA_PROSSIMA_REVISIONE[i]=GDP_REGISTRI_new$DATA_PROSSIMA_REVISIONE[which(GDP_REGISTRI_new$ID_REGISTRO==t)]
    #
Combinazione1$Numero_Criteri_DPIA_Reale[i]=GDP_REGISTRI_new$Numero_Criteri_DPIA_Reale[which(GDP_REGISTRI_new$ID_REGISTRO==t)]
  }}
}

G=duplicated(names (Combinazione1))

E= Combinazione1[which (G==FALSE)]

################################################################################
```

```
#togliere NA

E.noNA=E[complete.cases(E$ID_REGISTRO), ]
E.noNA=E.noNA[complete.cases(E.noNA$R), ]
E.noNA=E.noNA[complete.cases(E.noNA$TIPOLOGIA),]

################################################################################

#Combinazione2 <- merge(E, GDP_STRUTTURA_TRATT_RISCHIO_new,
by=c("ID_STRUT_TRATT_RISCHIO."), all=FALSE)
Combinazione2 <- full_join(E.noNA, GDP_STRUTTURA_TRATT_RISCHIO_new,
by=c("ID_STRUT_TRATT_RISCHIO"))

#togliere NA
Tabella.finale.noNA.Combinazione2=Combinazione2[complete.cases(Combinazione2$TIPOLOGIA.
y),]
Tabella.finale.noNA.Combinazione2=Tabella.finale.noNA.Combinazione2[complete.cases(Tabella.fi
nale.noNA.Combinazione2$ID_REGISTRO), ]
Tabella.finale.noNA.Combinazione2=Tabella.finale.noNA.Combinazione2[complete.cases(Tabella.fi
nale.noNA.Combinazione2$R.x), ]

################################################################################

for (t in 1:(max(Tabella.finale.noNA.Combinazione2$ID_REGISTRO))){
  for (i in 1:nrow(Tabella.finale.noNA.Combinazione2)){
    if (Tabella.finale.noNA.Combinazione2$ID_REGISTRO[i]==t) {

Tabella.finale.noNA.Combinazione2$DES_REGISTRO[i]=E.noNA$DES_REGISTRO[which(E.noNA$ID_
REGISTRO==t)]
```

Tabella.finale.noNA.Combinazione2$VALORE_RISCHIO_FINALE_R[i]=E.noNA$VALORE_RISCHIO_FINALE_R[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$VALORE_RISCHIO_FINALE_I[i]=E.noNA$VALORE_RISCHIO_FINALE_I[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$VALORE_RISCHIO_FINALE_D[i]=E.noNA$VALORE_RISCHIO_FINALE_D[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$VAL_RISCHIO_INERENTE_R[i]=E.noNA$VAL_RISCHIO_INERENTE_R[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$VAL_RISCHIO_INERENTE_I[i]=E.noNA$VAL_RISCHIO_INERENTE_I[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$VAL_RISCHIO_INERENTE_D[i]=E.noNA$VAL_RISCHIO_INERENTE_D[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$TIPO_REGISTRO[i]=E.noNA$TIPO_REGISTRO[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$DATA_INIZIO[i]=E.noNA$DATA_INIZIO[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$DATA_FINE[i]=E.noNA$DATA_FINE[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$STORICIZZATO[i]=E.noNA$STORICIZZATO[which(E.noNA$ID_REGISTRO==t)]

```
Tabella.finale.noNA.Combinazione2$PRESENZA_ELEMENTI_RISCHIO[i]=E.noNA$PRESENZA_ELEMENTI_RISCHIO[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$STRUTTURA_TRATTAMENTO[i]=E.noNA$STRUTTURA_TRATTAMENTO[which(E.noNA$ID_REGISTRO==t)]

Tabella.finale.noNA.Combinazione2$NUMERO_VERSIONE[i]=E.noNA$NUMERO_VERSIONE[which(E.noNA$ID_REGISTRO==t)]
    ##
Tabella.finale.noNA.Combinazione2$Numero_Criteri_DPIA_Reale[i]=E$Numero_Criteri_DPIA_Reale[which(E$ID_REGISTRO==t)]
  }}
}

#############################################################################

A=duplicated(names (Tabella.finale.noNA.Combinazione2))

Tabella.f= Tabella.finale.noNA.Combinazione2[which (A==FALSE)]

#colnames(D)
#############################################################################
#######

Combinazione3GDP_REGISTRI_VAL_SINT_DPIA_new=full_join(Tabella.f,
GDP_REGISTRI_VAL_SINT_DPIA_new, by=c("ID_REGISTRO"))

#togliere NA
Combinazione3GDP_REGISTRI_VAL_SINT_DPIA_new2=Combinazione3GDP_REGISTRI_VAL_SINT_DPIA_new[complete.cases(Combinazione3GDP_REGISTRI_VAL_SINT_DPIA_new$TIPOLOGIA.y),]
```

```
Combinazione3GDP_REGISTRI_VAL_SINT_DPIA_new2=Combinazione3GDP_REGISTRI_VAL_SINT_D
PIA_new2[complete.cases(Combinazione3GDP_REGISTRI_VAL_SINT_DPIA_new2$ID_REGISTRO), ]
Combinazione3GDP_REGISTRI_VAL_SINT_DPIA_new2=Combinazione3GDP_REGISTRI_VAL_SINT_D
PIA_new2[complete.cases(Combinazione3GDP_REGISTRI_VAL_SINT_DPIA_new2$R.x), ]

AD=duplicated(names (Combinazione3GDP_REGISTRI_VAL_SINT_DPIA_new2))

Tabella.f2= Combinazione3GDP_REGISTRI_VAL_SINT_DPIA_new2[which (AD==FALSE)]

################################################################################
#Combinazione 4

library (dplyr)
GDP_REGISTRI_MINACCE_new= GDP_REGISTRI_MINACCE_new %>%
  rename ("R.MINACCE" = "R")
GDP_REGISTRI_MINACCE_new= GDP_REGISTRI_MINACCE_new %>%
  rename ("I.MINACCE" = "I")
GDP_REGISTRI_MINACCE_new= GDP_REGISTRI_MINACCE_new %>%
  rename ("D.MINACCE" = "D")
GDP_REGISTRI_MINACCE_new= GDP_REGISTRI_MINACCE_new %>%
  rename ("DATA_MOD.MINACCE" = "DATA_MOD")

Combinazione4GDP_REGISTRI_MINACCE_new=full_join(Tabella.f2,
GDP_REGISTRI_MINACCE_new, by=c("ID_REGISTRO"))

#togliere NA
Combinazione4GDP_REGISTRI_MINACCE_new2=Combinazione4GDP_REGISTRI_MINACCE_new[co
mplete.cases(Combinazione4GDP_REGISTRI_MINACCE_new$TIPOLOGIA.y),]
Combinazione4GDP_REGISTRI_MINACCE_new2=Combinazione4GDP_REGISTRI_MINACCE_new2[c
omplete.cases(Combinazione4GDP_REGISTRI_MINACCE_new2$ID_REGISTRO), ]
Combinazione4GDP_REGISTRI_MINACCE_new2=Combinazione4GDP_REGISTRI_MINACCE_new2[c
omplete.cases(Combinazione4GDP_REGISTRI_MINACCE_new2$R.x), ]
```

```
#Combinazione4GDP_REGISTRI_MINACCE_new2=Combinazione4GDP_REGISTRI_MINACCE_new2[
complete.cases(Combinazione4GDP_REGISTRI_MINACCE_new2$R.MINACCE), ]

AD=duplicated(names (Combinazione4GDP_REGISTRI_MINACCE_new2))

Tabella.f3= Combinazione4GDP_REGISTRI_MINACCE_new2[which (AD==FALSE)]

################################################################################
#Combinazione 5
GDP_REGISTRI_REGISTRI_CALCOLI_new= GDP_REGISTRI_REGISTRI_CALCOLI_new %>%
  rename ("R.REGISTRI_CALCOLI" = "R")
GDP_REGISTRI_REGISTRI_CALCOLI_new= GDP_REGISTRI_REGISTRI_CALCOLI_new %>%
  rename ("I.REGISTRI_CALCOLIE" = "I")
GDP_REGISTRI_REGISTRI_CALCOLI_new= GDP_REGISTRI_REGISTRI_CALCOLI_new %>%
  rename ("D.REGISTRI_CALCOLI" = "D")
GDP_REGISTRI_REGISTRI_CALCOLI_new= GDP_REGISTRI_REGISTRI_CALCOLI_new %>%
  rename ("DATA_MOD.REGISTRI_CALCOLI" = "DATA_MOD")

Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new=full_join(Tabella.f3,
GDP_REGISTRI_REGISTRI_CALCOLI_new, by=c("ID_REGISTRO"))

#togliere NA
Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new2=Combinazione5GDP_REGISTRI_REGISTRI
_CALCOLI_new[complete.cases(Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new$TIPOLOGI
A.y),]
Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new2=Combinazione5GDP_REGISTRI_REGISTRI
_CALCOLI_new2[complete.cases(Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new2$ID_REG
ISTRO), ]
Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new2=Combinazione5GDP_REGISTRI_REGISTRI
_CALCOLI_new2[complete.cases(Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new2$R.x), ]
```

```
Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new2=Combinazione5GDP_REGISTRI_REGISTRI
_CALCOLI_new2[complete.cases(Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new2$R.REGI
STRI_CALCOLI), ]

AD=duplicated(names (Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new2))

Tabella.f4=Combinazione5GDP_REGISTRI_REGISTRI_CALCOLI_new2[which (AD==FALSE)]

################################################################################
#Combinazione 5
GDP_REGISTRI_MISURE_MIT_new= GDP_REGISTRI_MISURE_MIT_new %>%
  rename ("R.REGISTRI_MISURE_MIT" = "R")
GDP_REGISTRI_MISURE_MIT_new= GDP_REGISTRI_MISURE_MIT_new %>%
  rename ("I.REGISTRI_MISURE_MIT" = "I")
GDP_REGISTRI_MISURE_MIT_new= GDP_REGISTRI_MISURE_MIT_new %>%
  rename ("D.REGISTRI_MISURE_MIT" = "D")
GDP_REGISTRI_MISURE_MIT_new= GDP_REGISTRI_MISURE_MIT_new %>%
  rename ("DATA_MOD.REGISTRI_MISURE_MIT" = "DATA_MOD")

Combinazione6GDP_REGISTRI_MISURE_MIT_new=full_join(Tabella.f4,
GDP_REGISTRI_MISURE_MIT_new, by=c("ID_REGISTRO"))

#togliere NA
Combinazione6GDP_REGISTRI_MISURE_MIT_new2=Combinazione6GDP_REGISTRI_MISURE_MIT_
new[complete.cases(Combinazione6GDP_REGISTRI_MISURE_MIT_new$TIPOLOGIA.y),]
Combinazione6GDP_REGISTRI_MISURE_MIT_new2=Combinazione6GDP_REGISTRI_MISURE_MIT_
new2[complete.cases(Combinazione6GDP_REGISTRI_MISURE_MIT_new2$ID_REGISTRO), ]
Combinazione6GDP_REGISTRI_MISURE_MIT_new2=Combinazione6GDP_REGISTRI_MISURE_MIT_
new2[complete.cases(Combinazione6GDP_REGISTRI_MISURE_MIT_new2$R.x), ]
#Combinazione6GDP_REGISTRI_MISURE_MIT_new2=Combinazione6GDP_REGISTRI_MISURE_MIT
_new2[complete.cases(Combinazione6GDP_REGISTRI_MISURE_MIT_new2$R.REGISTRI_MISURE_
MIT), ]
```

```
for (t in 1:(max(Combinazione6GDP_REGISTRI_MISURE_MIT_new2$ID_REGISTRO))){
  for (i in 1:nrow(Combinazione6GDP_REGISTRI_MISURE_MIT_new2)){
    if (Combinazione6GDP_REGISTRI_MISURE_MIT_new2$ID_REGISTRO[i]==t) {

Combinazione6GDP_REGISTRI_MISURE_MIT_new2$GRUPPO_MISURA[i]=GDP_REGISTRI_MISURE_
MIT_new$GRUPPO_MISURA[which(GDP_REGISTRI_MISURE_MIT_new$ID_REGISTRO==t)]}}}

AD=duplicated(names (Combinazione6GDP_REGISTRI_MISURE_MIT_new2))

Tabella.f5=Combinazione6GDP_REGISTRI_MISURE_MIT_new2[which (AD==FALSE)]

################################################################################
Tabella.finale=Tabella.f5

################################################################################

# Valutare gli NA
is.na(Tabella.finale)

colSums(is.na(Tabella.finale))

which(colSums(is.na(Tabella.finale))>0)

names(which(colSums(is.na(Tabella.finale))>0))

################################################################################

#Eliminare le cononne non importanti dopo la combinazione delle 3 tabelle (modifica del 04.07.23
per il reintegro di ASSESSMENT_DEFINITIVO, ID_REGISTRO_INIZ, ID_REGISTRO_PADRE,
STATO_RISCHIO_FINALE, FLAG_POSIZIONE_DPIA, DATA_PROSSIMA_REVISIONE )
names(Tabella.finale)
```

Tabella.finale$ORD

#D = D[,-c("STORICIZZATO.", "ASSESSMENT_DEFINITIVO.", "PERIODO_CONSERV.", "UM_TEMPO.", "ID_TIPO_CONTESTO..x", "ID_REGISTRO_INIZ." , "ID_REGISTRO_PADRE.", "TIPO_RECORD.", "STATO_RISCHIO_FINALE." ,"ID_RISCHIO_FINALE_R." , "ID_RISCHIO_FINALE_I." , "ID_RISCHIO_FINALE_D." , "FLAG_POSIZIONE_DPIA." , "ID_REGISTRO_CONFRONTO." , "ID_REGISTRO_STORICO." , "ID_SERVIZIO_TREE_RICERCA." , "ID_TIPI_REGISTRO." , "ID_BASE_DATI." , "ID_RISCHIO_INERENTE_R." ,"ID_RISCHIO_INERENTE_I." , "ID_RISCHIO_INERENTE_D.", "DATA_PROSSIMA_REVISIONE." , "ID_REGISTRO_VALUTAZ." , "ID_BASE_DATI_PROV." , "DATA_MOD..x" , "FLAG_DISABILITATO." , "DESCRIZIONE.", "TIPO_IMPATTO." , "ID_TIPO_CONTESTO..y" , "ID_RIFERIMENTO_NORM." , "FLAG_FATT_AMPL." , "FATTORE_AMPLIFICAZIONE_DB." , "CODICE_GARANTE_DATA_BREACH.", "FLAG_TIPOLOGIA13_SI." , "DATA_MOD..y" , "TIME_MOD.", "ORDINAMENTO.")]
#togliere "STORICIZZATO.", "ASSESSMENT_DEFINITIVO.", "PERIODO_CONSERV.", "UM_TEMPO.", "ID_TIPO_CONTESTO..x", "ID_REGISTRO_INIZ." , "ID_REGISTRO_PADRE.", "TIPO_RECORD.", "STATO_RISCHIO_FINALE." ,"ID_RISCHIO_FINALE_R." , "ID_RISCHIO_FINALE_I." , "ID_RISCHIO_FINALE_D." , "FLAG_POSIZIONE_DPIA." , "ID_REGISTRO_CONFRONTO." , "ID_REGISTRO_STORICO." , "ID_SERVIZIO_TREE_RICERCA." , "ID_TIPI_REGISTRO." , "ID_BASE_DATI." , "ID_RISCHIO_INERENTE_R." ,"ID_RISCHIO_INERENTE_I." , "ID_RISCHIO_INERENTE_D.", "DATA_PROSSIMA_REVISIONE." , "ID_REGISTRO_VALUTAZ." , "TIPOLOGIA..x", "ID_BASE_DATI_PROV." , "DATA_MOD..x" , "FLAG_DISABILITATO." , "DESCRIZIONE.", "TIPO_IMPATTO." , "ID_TIPO_CONTESTO..y" , "ID_RIFERIMENTO_NORM." , "FLAG_FATT_AMPL." , "FATTORE_AMPLIFICAZIONE_DB." , "CODICE_GARANTE_DATA_BREACH.", "FLAG_TIPOLOGIA13_SI." , "DATA_MOD..y" , "TIME_MOD.", "ORDINAMENTO.")
Tabella.finale = subset(Tabella.finale, select = -c(TIPOLOGIA.x, PERIODO_CONSERV, UM_TEMPO, ID_TIPO_CONTESTO.x, TIPO_RECORD, _RISCHIO_FINALE_R , ID_RISCHIO_FINALE_I , ID_RISCHIO_FINALE_D , ID_REGISTRO_CONFRONTO , ID_REGISTRO_STORICO , ID_SERVIZIO_TREE_RICERCA, ID_TIPI_REGISTRO, ID_RISCHIO_INERENTE_R, ID_RISCHIO_INERENTE_I, ID_RISCHIO_INERENTE_D, ID_REGISTRO_VALUTAZ , ID_BASE_DATI_PROV.x,ID_BASE_DATI_PROV.y, DATA_MOD.x, FLAG_DISABILITATO.x , FLAG_DISABILITATO.y, TIPO_IMPATTO , ID_TIPO_CONTESTO.y , ID_RIFERIMENTO_NORM, FLAG_FATT_AMPL, FATTORE_AMPLIFICAZIONE_DB, CODICE_GARANTE_DATA_BREACH,

```
FLAG_TIPOLOGIA13_SI,  DATA_MOD.y, ORDINAMENTO.y, MODELLO_INFORMATIVA,
MODELLO_CONSENSO, NOTE_PATH, NOTE,  NOTE_TRATTAMENTO, NOTE_INFORMATIVA,
NOTE_CONSENSO, NOTE_REPERIM_DATI,NOTE_CONSERVAZIONE))

#FINE PREPROCESSING 2 (Risultato del preprocessing 2 sono le tabelle secondo tipologia

################################################################################

#Aggiungere id_registro mancanti per le seguenti tipologie
B=c(12,8,7,1,9,24,4,19,5,18)
L=length(B)

for (r in 1:L){
H=sort(Tabella.finale$ID_REGISTRO[which(Tabella.finale$TIPOLOGIA.y==B[r])])

H1=sort(Tabella.finale$ID_REGISTRO[which(Tabella.finale$TIPOLOGIA.y==24)])

length(H)

GK=max(H1, na.rm=TRUE)

vector=c(1:GK)

for (t in 1:GK){
# FF=vector[t]!=H[t]

if (!(vector[t] %in% H)) {

 #U=c(59, 65,75)

 #inserire valori 1 in ID_REGISTRO=59, 65,75 per tipologia=24
```

```
tmp=rep(NA, ncol(Tabella.finale))
Tabella.finale=rbind(Tabella.finale,tmp)
Tabella.finale[nrow(Tabella.finale),"TIPOLOGIA.y"]=B[r]
Tabella.finale[nrow(Tabella.finale),"ID_REGISTRO"]=vector[t]

#Tabella.finale[nrow(Tabella.finale),"DES_REGISTRO"]=Tabella.finale$DES_REGISTRO[which(Tabell
a.finale$ID_REGISTRO==vector[t])]
Tabella.finale[nrow(Tabella.finale),"R.x"]=1
Tabella.finale[nrow(Tabella.finale),"D.x"]=1
Tabella.finale[nrow(Tabella.finale),"I.x"]=1
Tabella.finale[nrow(Tabella.finale),"R.MINACCE"]=1
Tabella.finale[nrow(Tabella.finale),"D.MINACCE"]=1
Tabella.finale[nrow(Tabella.finale),"I.MINACCE"]=1
Tabella.finale[nrow(Tabella.finale),"R.REGISTRI_MISURE_MIT"]=1
Tabella.finale[nrow(Tabella.finale),"D.REGISTRI_MISURE_MIT"]=1
Tabella.finale[nrow(Tabella.finale),"I.REGISTRI_MISURE_MIT"]=1
}}}

#for (t in 1:(max(Tabella.finale$ID_REGISTRO))){
#for (i in 1:nrow(Tabella.finale)){
#if (Tabella.finale$ID_REGISTRO[i]==t) {
#
Tabella.finale$DES_REGISTRO[i]=Tabella.finale$DES_REGISTRO[which(Tabella.finale$ID_REGISTRO
==t)]}}}

Tabella.finale$ID_REGISTRO[which(Tabella.finale$TIPOLOGIA.y==24)]

sort(Tabella.finale$ID_REGISTRO[which(Tabella.finale$TIPOLOGIA.y==7)])

TabellaFinale=Tabella.finale
```

C. Appendix: DAX Code for Business Intelligence analysis

This appendix contains the DAX code applied in the analysis phase of the business intelligence workflow outlined in this study. The code offers structured formulas and calculations that enhance the reproducibility and clarity of the findings. Each section of the code is tailored to distinct analytical objectives, allowing others to replicate the process or modify it for comparable research applications.

```
Tabella.tipologia1 = CALCULATETABLE(FILTER('Sheet1', 'Sheet1'[TIPOLOGIA.y]=1))
Tabella.tipologia12 = CALCULATETABLE(FILTER('Sheet1', 'Sheet1'[TIPOLOGIA.y]=12))
Tabella.tipologia13 = CALCULATETABLE(FILTER('Sheet1', 'Sheet1'[TIPOLOGIA.y]=13))
Tabella.tipologia19 = CALCULATETABLE(FILTER('Sheet1', 'Sheet1'[TIPOLOGIA.y]=19))
Tabella.tipologia24 = CALCULATETABLE(FILTER('Sheet1', 'Sheet1'[TIPOLOGIA.y]=24))
Tabella.tipologia4 = CALCULATETABLE(FILTER('Sheet1', 'Sheet1'[TIPOLOGIA.y]=4))
Tabella.tipologia5 = CALCULATETABLE(FILTER('Sheet1', 'Sheet1'[TIPOLOGIA.y]=5))
Tabella.tipologia7 = CALCULATETABLE(FILTER('Sheet1', 'Sheet1'[TIPOLOGIA.y]=7))
Tabella.tipologia8 = CALCULATETABLE(FILTER('Sheet1', 'Sheet1'[TIPOLOGIA.y]=8))
Tabella.tipologia9 = CALCULATETABLE(FILTER('Sheet1', 'Sheet1'[TIPOLOGIA.y]=9))
```

```
Tabella.raggruppamento.tip.1 = SUMMARIZE('Tabella.tipologia1',
'Tabella.tipologia1'[ID_REGISTRO], "Massimo_R", CALCULATE (MAX('Tabella.tipologia1'[R.x])),
"Massimo_I", CALCULATE (MAX('Tabella.tipologia1'[I.x])), "Massimo_D", CALCULATE
(MAX('Tabella.tipologia1'[D.x])))
Tabella.raggruppamento.tip.12 = SUMMARIZE('Tabella.tipologia12',
'Tabella.tipologia12'[ID_REGISTRO], "Massimo_R", CALCULATE (MAX('Tabella.tipologia12'[R.x])),
"Massimo_I", CALCULATE (MAX('Tabella.tipologia12'[I.x])), "Massimo_D", CALCULATE
(MAX('Tabella.tipologia12'[D.x])))
Tabella.raggruppamento.tip.13 = SUMMARIZE('Tabella.tipologia13',
'Tabella.tipologia13'[ID_REGISTRO], "Massimo_R", CALCULATE (MAX('Tabella.tipologia13'[R.x])),
"Massimo_I", CALCULATE (MAX('Tabella.tipologia13'[I.x])), "Massimo_D", CALCULATE
(MAX('Tabella.tipologia13'[D.x])))
```

```
Tabella.raggruppamento.tip.19 = SUMMARIZE('Tabella.tipologia19',
'Tabella.tipologia19'[ID_REGISTRO], "Massimo_R", CALCULATE
(MAX('Tabella.tipologia19'[R.x])),"Massimo_I", CALCULATE (MAX('Tabella.tipologia19'[I.x])),
"Massimo_D", CALCULATE (MAX('Tabella.tipologia19'[D.x])))
Tabella.raggruppamento.tip.24 = SUMMARIZE('Tabella.tipologia24',
'Tabella.tipologia24'[ID_REGISTRO], "Massimo_R", CALCULATE (MAX('Tabella.tipologia24'[R.x])),
"Massimo_I", CALCULATE (MAX('Tabella.tipologia24'[I.x])), "Massimo_D", CALCULATE
(MAX('Tabella.tipologia24'[D.x])))
Tabella.raggruppamento.tip.4 = SUMMARIZE('Tabella.tipologia4',
'Tabella.tipologia4'[ID_REGISTRO], "Massimo_R", CALCULATE (MAX('Tabella.tipologia4'[R.x])),
"Massimo_I", CALCULATE (MAX('Tabella.tipologia4'[I.x])), "Massimo_D", CALCULATE
(MAX('Tabella.tipologia4'[D.x])))
Tabella.raggruppamento.tip.5 = SUMMARIZE('Tabella.tipologia5',
'Tabella.tipologia5'[ID_REGISTRO], "Massimo_R", CALCULATE (MAX('Tabella.tipologia5'[R.x])),
"Massimo_I", CALCULATE (MAX('Tabella.tipologia5'[I.x])), "Massimo_D", CALCULATE
(MAX('Tabella.tipologia5'[D.x])))
Tabella.raggruppamento.tip.7 = SUMMARIZE('Tabella.tipologia7',
'Tabella.tipologia7'[ID_REGISTRO], 'Tabella.tipologia7'[DES_REGISTRO] ,"Massimo_R", CALCULATE
(MAX('Tabella.tipologia7'[R.x])), "Massimo_I", CALCULATE (MAX('Tabella.tipologia7'[I.x])),
"Massimo_D", CALCULATE (MAX('Tabella.tipologia7'[D.x])))
Tabella.raggruppamento.tip.8 = SUMMARIZE('Tabella.tipologia8',
'Tabella.tipologia8'[ID_REGISTRO], "Massimo_R", CALCULATE (MAX('Tabella.tipologia8'[R.x])),
"Massimo_I", CALCULATE (MAX('Tabella.tipologia8'[I.x])), "Massimo_D", CALCULATE
(MAX('Tabella.tipologia8'[D.x])))
Tabella.raggruppamento.tip.9 = SUMMARIZE('Tabella.tipologia9',
'Tabella.tipologia9'[ID_REGISTRO], "Massimo_R", CALCULATE (MAX('Tabella.tipologia9'[R.x])),
"Massimo_I", CALCULATE (MAX('Tabella.tipologia9'[I.x])), "Massimo_D", CALCULATE
(MAX('Tabella.tipologia9'[D.x])))
```

```
table_min = SUMMARIZE('Sheet1', 'Sheet1'[ID_REGISTRO], 'Sheet1'[ID_REGISTRO_MISURE_MIT],
'Sheet1'[GRUPPO_MISURA] , 'Sheet1'[R.REGISTRI_MISURE_MIT],
'Sheet1'[I.REGISTRI_MISURE_MIT], 'Sheet1'[D.REGISTRI_MISURE_MIT])
table_min2.GM1 = SUMMARIZE(FILTER('Sheet1', 'Sheet1'[GRUPPO_MISURA]=1 &&
NOT(ISBLANK('Sheet1'[ID_REGISTRO_MISURE_MIT]))), 'Sheet1'[ID_REGISTRO],
'Sheet1'[ID_REGISTRO_MISURE_MIT], 'Sheet1'[GRUPPO_MISURA],
'Sheet1'[R.REGISTRI_MISURE_MIT], 'Sheet1'[I.REGISTRI_MISURE_MIT],
'Sheet1'[D.REGISTRI_MISURE_MIT])
table_min2.GM2 = SUMMARIZE(FILTER('Sheet1', 'Sheet1'[GRUPPO_MISURA]=2 &&
NOT(ISBLANK('Sheet1'[ID_REGISTRO_MISURE_MIT]))), 'Sheet1'[ID_REGISTRO],
'Sheet1'[ID_REGISTRO_MISURE_MIT], 'Sheet1'[GRUPPO_MISURA],
'Sheet1'[R.REGISTRI_MISURE_MIT], 'Sheet1'[I.REGISTRI_MISURE_MIT],
'Sheet1'[D.REGISTRI_MISURE_MIT])
table_min3.GM1 = SUMMARIZE('table_min2.GM1', 'table_min2.GM1'[ID_REGISTRO],
"prodottoRmin1", CALCULATE(PRODUCTX(
'table_min2.GM1','table_min2.GM1'[R.REGISTRI_MISURE_MIT])), "prodottoImin1",
CALCULATE(PRODUCTX(
'table_min2.GM1','table_min2.GM1'[I.REGISTRI_MISURE_MIT])),  "prodottoDmin1",
CALCULATE(PRODUCTX( 'table_min2.GM1','table_min2.GM1'[D.REGISTRI_MISURE_MIT])))
table_min3.GM2 = SUMMARIZE('table_min2.GM2',
'table_min2.GM2'[ID_REGISTRO],"prodottoRmin2", CALCULATE(PRODUCTX(
'table_min2.GM2','table_min2.GM2'[R.REGISTRI_MISURE_MIT])), "prodottoImin2",
CALCULATE(PRODUCTX(
'table_min2.GM2','table_min2.GM2'[I.REGISTRI_MISURE_MIT])),  "prodottoDmin2",
CALCULATE(PRODUCTX( 'table_min2.GM2','table_min2.GM2'[D.REGISTRI_MISURE_MIT])))
```

```
Tabella.amplificazione.tip.12 = SELECTCOLUMNS('Tabella.tipologia12', "ID_REG",
'Tabella.tipologia12'[ID_REGISTRO], "ID_strutt", 'Tabella.tipologia12'[ID_STRUT_TRATT_RISCHIO],
"R", 'Tabella.tipologia12'[R.x], "I", 'Tabella.tipologia12'[I.x], "D", 'Tabella.tipologia12'[D.x])
```

```
Tabella.amplificazione.tip.19 = SELECTCOLUMNS('Tabella.tipologia19', "ID_REG",
'Tabella.tipologia19'[ID_REGISTRO], "ID_strutt", 'Tabella.tipologia19'[ID_STRUT_TRATT_RISCHIO],
"R", 'Tabella.tipologia19'[R.x], "I", 'Tabella.tipologia19'[R.x], "D", 'Tabella.tipologia19'[R.x])
Tabella.amplificazione.tip.4 = SELECTCOLUMNS('Tabella.tipologia4', "ID_REG",
'Tabella.tipologia4'[ID_REGISTRO],"ID_strutt", 'Tabella.tipologia4'[ID_STRUT_TRATT_RISCHIO], "R",
'Tabella.tipologia4'[R.x],  "I", 'Tabella.tipologia4'[I.x],  "D", 'Tabella.tipologia4'[D.x])
Tabella.amplificazione.tip.5 = SELECTCOLUMNS('Tabella.tipologia5', "ID_REG",
'Tabella.tipologia5'[ID_REGISTRO],"ID_strutt", 'Tabella.tipologia5'[ID_STRUT_TRATT_RISCHIO], "R",
'Tabella.tipologia5'[R.x], "I", 'Tabella.tipologia5'[I.x], "D", 'Tabella.tipologia5'[D.x])
Tabella.amplificazione.tip.9 = SELECTCOLUMNS('Tabella.tipologia9', "ID_REG",
'Tabella.tipologia9'[ID_REGISTRO],"ID_strutt", 'Tabella.tipologia9'[ID_STRUT_TRATT_RISCHIO], "R",
'Tabella.tipologia9'[R.x], "I", 'Tabella.tipologia9'[I.x], "D", 'Tabella.tipologia9'[D.x])
```

```
dati_filtrati_12 =

   SUMMARIZECOLUMNS(

      'Tabella.amplificazione.tip.12'[ID_REG] ,

      'Tabella.amplificazione.tip.12'[ID_strutt],

      'Tabella.amplificazione.tip.12'[R], 'Tabella.amplificazione.tip.12'[I],
'Tabella.amplificazione.tip.12'[D]

   )
dati_filtrati_19 =

   SUMMARIZECOLUMNS(

      'Tabella.amplificazione.tip.19'[ID_REG] ,
```

```
    'Tabella.amplificazione.tip.19'[ID_strutt],

    'Tabella.amplificazione.tip.19'[R],  'Tabella.amplificazione.tip.19'[I],  'Tabella.amplificazione.tip
.19'[D]

  )
dati_filtrati_4 =

  SUMMARIZECOLUMNS(

    'Tabella.amplificazione.tip.4'[ID_REG] ,

    'Tabella.amplificazione.tip.4'[ID_strutt],

    'Tabella.amplificazione.tip.4'[R],  'Tabella.amplificazione.tip.4'[I],  'Tabella.amplificazione.tip.4'
[D]

  )
dati_filtrati_5 =

  SUMMARIZECOLUMNS(

    'Tabella.amplificazione.tip.5'[ID_REG] ,

    'Tabella.amplificazione.tip.5'[ID_strutt],

    'Tabella.amplificazione.tip.5'[R],  'Tabella.amplificazione.tip.5'[I],  'Tabella.amplificazione.tip.5'
[D]

  )
dati_filtrati_9 =
```

```
   SUMMARIZECOLUMNS(

      'Tabella.amplificazione.tip.9'[ID_REG] ,

      'Tabella.amplificazione.tip.9'[ID_strutt],

      'Tabella.amplificazione.tip.9'[R], 'Tabella.amplificazione.tip.9'[I],
'Tabella.amplificazione.tip.9'[D]

   )
```

```
dati_aggregati_12 =

   SUMMARIZE(

      dati_filtrati_12,

      dati_filtrati_12[ID_REG],

      "Prod_R", PRODUCTX(dati_filtrati_12, dati_filtrati_12[R]), "Prod_I",
PRODUCTX(dati_filtrati_12, dati_filtrati_12[I]), "Prod_D", PRODUCTX(dati_filtrati_12,
dati_filtrati_12[D])

   )
dati_aggregati_19 =

   SUMMARIZE(

      dati_filtrati_19,

      dati_filtrati_19[ID_REG],
```

```
    "Prod_R", PRODUCTX(dati_filtrati_19, dati_filtrati_19[R]),  "Prod_I",
PRODUCTX(dati_filtrati_19, dati_filtrati_19[I]),  "Prod_D", PRODUCTX(dati_filtrati_19,
dati_filtrati_19[D])


  )
dati_aggregati_4 =
  SUMMARIZE(
    dati_filtrati_4,
    dati_filtrati_4[ID_REG],
    "Prod_R", PRODUCTX(dati_filtrati_4, dati_filtrati_4[R]),  "Prod_I", PRODUCTX(dati_filtrati_4,
dati_filtrati_4[I]),  "Prod_D", PRODUCTX(dati_filtrati_4, dati_filtrati_4[D])
  )
dati_aggregati_5 =
  SUMMARIZE(
    dati_filtrati_5,
    dati_filtrati_5[ID_REG],
    "Prod_R", PRODUCTX(dati_filtrati_5, dati_filtrati_5[R]), "Prod_I", PRODUCTX(dati_filtrati_5,
dati_filtrati_5[I]), "Prod_D", PRODUCTX(dati_filtrati_5, dati_filtrati_5[D])
  )
dati_aggregati_9 =
```

```
  SUMMARIZE(

    dati_filtrati_9,

    dati_filtrati_9[ID_REG],

    "Prod_R", PRODUCTX(dati_filtrati_9, dati_filtrati_9[R]), "Prod_I", PRODUCTX(dati_filtrati_9, dati_filtrati_9[I]), "Prod_D", PRODUCTX(dati_filtrati_9, dati_filtrati_9[D])

  )
```

```
Tabella amplificazione gravità = SUMMARIZE('Tabella.raggruppamento.tip.12', 'Tabella.raggruppamento.tip.12'[ID_REGISTRO], 'Tabella.raggruppamento.tip.7'[DES_REGISTRO] ,"G_Rampl", CALCULATE(SUMX(FILTER('Tabella.raggruppamento.tip.12',NOT(ISBLANK([Massimo_R]))),[Massimo_R]* RELATED ('Tabella.raggruppamento.tip.8'[Massimo_R])* RELATED ('Tabella.raggruppamento.tip.1'[Massimo_R])* RELATED ('Tabella.raggruppamento.tip.9'[Massimo_R])* RELATED ('Tabella.raggruppamento.tip.4'[Massimo_R])* RELATED ('Tabella.raggruppamento.tip.19'[Massimo_R])* RELATED ('Tabella.raggruppamento.tip.5'[Massimo_R]))), "G_Iampl", CALCULATE(SUMX(FILTER('Tabella.raggruppamento.tip.12',NOT(ISBLANK([Massimo_I]))),[Massimo_I]* RELATED ('Tabella.raggruppamento.tip.8'[Massimo_I])* RELATED ('Tabella.raggruppamento.tip.1'[Massimo_I])* RELATED ('Tabella.raggruppamento.tip.9'[Massimo_I])* RELATED ('Tabella.raggruppamento.tip.4'[Massimo_I])* RELATED ('Tabella.raggruppamento.tip.19'[Massimo_I])* RELATED ('Tabella.raggruppamento.tip.5'[Massimo_I]))), "G_Dampl", CALCULATE(SUMX(FILTER('Tabella.raggruppamento.tip.12',NOT(ISBLANK([Massimo_D]))),[Massimo_D]* RELATED ('Tabella.raggruppamento.tip.8'[Massimo_D])* RELATED ('Tabella.raggruppamento.tip.1'[Massimo_D])* RELATED
```

('Tabella.raggruppamento.tip.9'[Massimo_D])* RELATED
('Tabella.raggruppamento.tip.4'[Massimo_D])* RELATED
('Tabella.raggruppamento.tip.19'[Massimo_D])* RELATED
('Tabella.raggruppamento.tip.5'[Massimo_D]))))

Tabella Gravità = SUMMARIZE('Tabella.raggruppamento.tip.7',
'Tabella.raggruppamento.tip.7'[ID_REGISTRO],
'Tabella.raggruppamento.tip.7'[DES_REGISTRO],"GI_R",
CALCULATE(SUMX(FILTER('Tabella.raggruppamento.tip.7',NOT(ISBLANK([Massimo_R]))),[Massimo_R]*RELATED('Tabella.raggruppamento.tip.24'[Massimo_R]))), "GI_I",
CALCULATE(SUMX(FILTER('Tabella.raggruppamento.tip.7',NOT(ISBLANK([Massimo_R]))),[Massimo_R]*RELATED('Tabella.raggruppamento.tip.24'[Massimo_I]))),"GI_D",
CALCULATE(SUMX(FILTER('Tabella.raggruppamento.tip.7',NOT(ISBLANK([Massimo_R]))),[Massimo_R]*RELATED('Tabella.raggruppamento.tip.24'[Massimo_D]))), "G_R",
CALCULATE(SUMX(FILTER('Tabella.raggruppamento.tip.7',NOT(ISBLANK([Massimo_R]))),[Massimo_R]*RELATED('Tabella.raggruppamento.tip.24'[Massimo_R]) * RELATED
('dati_aggregati_12'[Prod_R])* RELATED ('Tabella.raggruppamento.tip.8'[Massimo_R])* RELATED
('Tabella.raggruppamento.tip.1'[Massimo_R])* RELATED ('dati_aggregati_9'[Prod_R])* RELATED
('dati_aggregati_4'[Prod_R]) * RELATED ('dati_aggregati_19'[Prod_R])* RELATED
('dati_aggregati_5'[Prod_R]))), "G_D",
CALCULATE(SUMX(FILTER('Tabella.raggruppamento.tip.7',NOT(ISBLANK([Massimo_R]))),[Massimo_R]*RELATED('Tabella.raggruppamento.tip.24'[Massimo_D]) * RELATED
('dati_aggregati_12'[Prod_D])* RELATED ('Tabella.raggruppamento.tip.8'[Massimo_D])* RELATED
('Tabella.raggruppamento.tip.1'[Massimo_D])* RELATED ('dati_aggregati_9'[Prod_D])* RELATED
('dati_aggregati_4'[Prod_D])* RELATED ('dati_aggregati_19'[Prod_D])* RELATED
('dati_aggregati_5'[Prod_D]))), "G_I",
CALCULATE(SUMX(FILTER('Tabella.raggruppamento.tip.7',NOT(ISBLANK([Massimo_R]))),[Massimo_R]*RELATED('Tabella.raggruppamento.tip.24'[Massimo_I]) * RELATED
('dati_aggregati_12'[Prod_I])* RELATED ('Tabella.raggruppamento.tip.8'[Massimo_I])* RELATED
('Tabella.raggruppamento.tip.1'[Massimo_I])* RELATED ('dati_aggregati_9'[Prod_I])* RELATED

('dati_aggregati_4'[Prod_I])* RELATED ('dati_aggregati_19'[Prod_I])* RELATED ('dati_aggregati_5'[Prod_I])))))

Tabella Probabilità e Rischio R.I.D = SUMMARIZECOLUMNS('Sheet1'[ID_REGISTRO], 'Sheet1'[DES_REGISTRO], 'Data Refresh'[Data Time],"Media R.DIGITALE", CALCULATE(AVERAGE('Sheet1'[R.MINACCE]), 'Sheet1'[TIPO_PROCESSO]=1),"Media I.DIGITALE", CALCULATE(AVERAGE('Sheet1'[I.MINACCE]), 'Sheet1'[TIPO_PROCESSO]=1),"Media D.DIGITALE", CALCULATE(AVERAGE('Sheet1'[D.MINACCE]), 'Sheet1'[TIPO_PROCESSO]=1), "Probabilità amplificata R", MIN(1,(CALCULATE(SUMX(CALCULATETABLE(VALUES('Sheet1'[R.x]), 'Sheet1'[TIPOLOGIA.y]=18), [R.x]/100)+1)* (CALCULATE(AVERAGE('Sheet1'[R.MINACCE]), 'Sheet1'[TIPO_PROCESSO]=1)))),"Probabilità amplificata I", MIN(1,(CALCULATE(SUMX(CALCULATETABLE(VALUES('Sheet1'[I.x]), 'Sheet1'[TIPOLOGIA.y]=18), [I.x]/100)+1)* (CALCULATE(AVERAGE('Sheet1'[I.MINACCE]), 'Sheet1'[TIPO_PROCESSO]=1)))),"Probabilità amplificata D", MIN(1,(CALCULATE(SUMX(CALCULATETABLE(VALUES('Sheet1'[D.x]), 'Sheet1'[TIPOLOGIA.y]=18), [D.x]/100)+1)* (CALCULATE(AVERAGE('Sheet1'[D.MINACCE]), 'Sheet1'[TIPO_PROCESSO]=1)))),"Media R.CARTACEO", CALCULATE(AVERAGE('Sheet1'[R.MINACCE]), 'Sheet1'[TIPO_PROCESSO]=2),"Media I.CARTACEO", CALCULATE(AVERAGE('Sheet1'[I.MINACCE]), 'Sheet1'[TIPO_PROCESSO]=2),"Media D.CARTACEO", CALCULATE(AVERAGE('Sheet1'[D.MINACCE]), 'Sheet1'[TIPO_PROCESSO]=2), "PERC_GEST_AUTO", CALCULATE (MAX('Sheet1'[PERC_GEST_AUTO])))

MgD1 = 1*RELATED('table_min3.GM1'[prodottoDmin1])

MgD1_new1 = COALESCE('Tabella Probabilità e Rischio R.I.D'[MgD1],1)

MgI1 = 1*RELATED('table_min3.GM1'[prodottoImin1])

MgI1_new1 = COALESCE('Tabella Probabilità e Rischio R.I.D'[MgI1],1)

MgR1 = 1*RELATED('table_min3.GM1'[prodottoRmin1])

MgR1_new1 = COALESCE('Tabella Probabilità e Rischio R.I.D'[MgR1],1)

MsD2 = 1*RELATED('table_min3.GM2'[prodottoDmin2])

MsD2_new1 = COALESCE('Tabella Probabilità e Rischio R.I.D'[MsD2],1)

MsI2 = 1*RELATED('table_min3.GM2'[prodottoImin2])

MsI2_new1 = COALESCE('Tabella Probabilità e Rischio R.I.D'[MsI2],1)

MsR2 = 1*RELATED('table_min3.GM2'[prodottoRmin2])

MsR2_new1 = COALESCE('Tabella Probabilità e Rischio R.I.D'[MsR2],1)

Prob.D.complessiva.tip13 = 'Tabella Probabilità e Rischio R.I.D'[probababilità complessiva D]*RELATED('Tabella.raggruppamento.tip.13'[Massimo_D])

Prob.I.complessiva.tip13 = 'Tabella Probabilità e Rischio R.I.D'[probababilità complessiva I]*RELATED('Tabella.raggruppamento.tip.13'[Massimo_I])

Prob.R.complessiva.tip13 = 'Tabella Probabilità e Rischio R.I.D'[probababilità complessiva R]*RELATED('Tabella.raggruppamento.tip.13'[Massimo_R])

probababilità complessiva D = IF(ISBLANK([Risultato D]), IF(NOT (ISBLANK([Probabilità amplificata D])), 'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata D],'Tabella Probabilità e Rischio R.I.D'[Media D.CARTACEO]),[Risultato D])

probababilità complessiva I = IF(ISBLANK([Risultato I]), IF(NOT (ISBLANK([Probabilità amplificata I])), 'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata I],'Tabella Probabilità e Rischio R.I.D'[Media I.CARTACEO]),[Risultato I])

probababilità complessiva R = IF(ISBLANK([Risultato R]), IF(NOT (ISBLANK([Probabilità amplificata R])), 'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R],'Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO]),[Risultato R])

R2 =

```
IF(

   NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media R.DIGITALE])) && NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO])),

   IF(

      'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO] <> 0,

      (('Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R] * 'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO]) / 100) + ('Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO] * ((100 - 'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO]) / 100)),

      IF(
```

```
        'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO] = 0,

        'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R] + ('Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO] / 2),

        'Tabella Probabilità e Rischio R.I.D'[Media R.DIGITALE]

      )

   ),

   IF(

      NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media R.DIGITALE])),

      'Tabella Probabilità e Rischio R.I.D'[Media R.DIGITALE],

      IF(

         NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R])),

         'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R],

         BLANK()

      )

   )

)
```

Rischio D INERENTE = 'Tabella Probabilità e Rischio R.I.D'[Prob.D.complessiva.tip13]*RELATED('Tabella Gravità'[G_D])

Rischio finale D = 'Tabella Probabilità e Rischio R.I.D'[Rischio D INERENTE]*'Tabella Probabilità e Rischio R.I.D'[MgD1_new1]*'Tabella Probabilità e Rischio R.I.D'[MsD2_new1]

Rischio finale I = 'Tabella Probabilità e Rischio R.I.D'[Rischio I INERENTE]*'Tabella Probabilità e Rischio R.I.D'[MgI1_new1]*'Tabella Probabilità e Rischio R.I.D'[MsI2_new1]

Rischio finale R = 'Tabella Probabilità e Rischio R.I.D'[Rischio R INERENTE]*'Tabella Probabilità e Rischio R.I.D'[MgR1_new1]*'Tabella Probabilità e Rischio R.I.D'[MsR2_new1]

Rischio I INERENTE = 'Tabella Probabilità e Rischio R.I.D'[Prob.I.complessiva.tip13]*RELATED('Tabella Gravità'[G_I])

Rischio R INERENTE = 'Tabella Probabilità e Rischio R.I.D'[Prob.R.complessiva.tip13]*RELATED('Tabella Gravità'[G_R])

Risultato D =

```
IF(

   NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media D.DIGITALE])) && NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media D.CARTACEO])),

   IF(

      'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO] <> 0,

      (('Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata D] * 'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO]) / 100) + ('Tabella Probabilità e Rischio R.I.D'[Media D.CARTACEO] * ((100 - 'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO]) / 100)),

      IF(

         'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO] = 0,
```

('Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata D] + 'Tabella Probabilità e Rischio R.I.D'[Media D.CARTACEO]) / 2

)

), IF(

ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media D.CARTACEO]),

'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata D],

IF(

ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media D.DIGITALE]),

'Tabella Probabilità e Rischio R.I.D'[Media D.CARTACEO],

BLANK()

)

)

)

Risultato I =

IF(

```
   NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media I.DIGITALE])) && NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media I.CARTACEO])),

   IF(

      'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO] <> 0,

      (('Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata I] * 'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO]) / 100) + ('Tabella Probabilità e Rischio R.I.D'[Media I.CARTACEO] * ((100 - 'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO]) / 100)),

      IF(

         'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO] = 0,

        ( 'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata I] + 'Tabella Probabilità e Rischio R.I.D'[Media I.CARTACEO]) / 2



      )

   ),   IF(

     ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media I.CARTACEO]),

      'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata I],

      IF(

         ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media I.DIGITALE]),
```

```
        'Tabella Probabilità e Rischio R.I.D'[Media I.CARTACEO],

        BLANK()

      )

   )

)
```

Risultato R =

```
IF(

   NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media R.DIGITALE])) && NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO])),

   IF(

      'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO] <> 0,

      (('Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R] * 'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO]) / 100) + ('Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO] * ((100 - 'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO]) / 100)),

      IF(

         'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO] = 0,

        ( 'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R] + 'Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO]) / 2
```

```
        )
    ),   IF(
        ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO]),
        'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R],
        IF(
            ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media R.DIGITALE]),
            'Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO],
            BLANK()
        )
    )
)
```

Risultato2 = IF(ISBLANK([Risultato R]), IF(NOT (ISBLANK([Probabilità amplificata R])), 'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R],'Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO]),[Risultato R])

RR =

```
IF(
```

```
   NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media R.DIGITALE])) && NOT(ISBLANK('Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO])),

   IF(

      'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO] <> 0,

      (('Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R] * 'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO]) / 100) + ('Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO] * ((100 - 'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO]) / 100)),

      IF(

         'Tabella Probabilità e Rischio R.I.D'[PERC_GEST_AUTO] = 0,

        ( 'Tabella Probabilità e Rischio R.I.D'[Probabilità amplificata R] + 'Tabella Probabilità e Rischio R.I.D'[Media R.CARTACEO]) / 2,

         BLANK()

      )

   ),

   BLANK()

)
```