MISSIONE 4
ISTRUZIONE
RICERCA

BRIEF BIOROBOTICS RESEARCH AND INNOVATION ENGINEERING FACILITIES

D.7.6 REPORT ON POLICY DESIGN AND ADVICE

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Quadro riassuntivo rilasci documento

Data	*Stato documento*	*Realizzato da*	*Note*	*Supervisione*
01-12-23	First draft	Arianna Rossi	First draft of the complete document v.0.1.	Giovanni Comandé
04-12-23	Draft review	Giovanni Comandé	Content review v.0.1.	Giovanni Comandé
04-12-23	Content update	Arianna Rossi	Reviewed draft v.0.2	Giovanni Comandé
21-12-23	Final draft	Arianna Rossi	Final draft addressing comments of reviewers	Giovanni Comandé

ACKNOWLEDGEMENT

This report is the outcome of collaborative work. The main authors are Arianna Rossi and Prof. Giovanni Comandé. However, other researchers of the Law and Policy Hub have contributed with their policy recommendations and best practices, namely in alphabetical order: Vittoria Caponecchia, Irina Carnat, Federica Casarosa, Georgios Christou, Tommaso Crepax, Maria Gagliardi, Francesca Gennari, Robert Lee Poe, and Stefano Tramacere. The cross-field regulatory analysis (reported in D7.3) on which the report builds was authored by Denise Amram and Francesca Gennari. We thank Irene Aprile, Annalisa Calabrò, Domenico Camboni, Federica Casarosa, Simona Crea, Francesca Gennari, Andrea Parziale, Sabrina Grigolo, Gianclaudio Malgieri and Maria Mauro for their valuable comments on the draft version of this report.

DISCLAIMER

This project has received funding by the Ministero dell'Università e della Ricerca (MUR), Direzione generale dell'internazionalizzazione e della comunicazione within the framework of the National Recovery and Resilience Plan (NRRP) within the call for proposals framework REFORMS AND INVESTMENTS UNDER THE RECOVERY AND RESILIENCE PLAN – Next Generation EU , Intervention field 6: Investment in digital capacities and deployment of advanced technologies DESI dimension 4: Integration of digital technologies + ad hoc data collections 055 - Other types of ICT infrastructure (including large-scale computer resources/equipment, data centres, sensors and other wireless equipment). Mission 4 – "Education and Research" Component 2: from research to business Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures Action 3.1.1 "Creation of new research infrastructures strengthening of existing ones and their networking for Scientific Excellence under Horizon Europe.

EXECUTIVE SUMMARY

This report contains a set of policy recommendations for European policymakers and best practices for researchers working in the biorobotics field, often with biomedical applications, within the BRIEF project. The previous cross-field regulatory analysis (published in deliverable D7.3) has identified the relevant regulatory frameworks that govern such a multidisciplinary area where technological advancements outpace the development of regulations. Such rapidly evolving frameworks concern personal and non-personal data management (i.e., the General Data Protection Regulation, the Data Governance Act, the Regulation on the Free Flow of Data, the European Health Data Space proposal, the Data Act proposal, and their national implementations), health law (e.g., the Clinical Trials Regulation, the Medical Devices Regulation and their national implementations), artificial intelligence (i.e., the AI Act proposal and the AI liability Directive proposal), liability (e.g., the Product Liability Directive Update), cybersecurity (i.e., the NIS Directive, NIS2 and the Cyber resilience Act, and their national implementations) and machinery (e.g., the Machinery Directive and the General Product Safety Regulation). Even in the absence of enforceable regulations, three main principles underpin the trustworthy-by-design development of technologies, namely fairness, accountability, and transparency.

Based on this analysis, the report provides an initial set of guidelines that are meant to equip researchers with hands-on best practices to be implemented in their R&I activities; and policy recommendations that identify regulatory gaps that need to be overcome to ensure legal certainty and support technological advancements. These are two of the possible interventions that we propose to facilitate the compliant design of new biorobotic technologies. Additional ones include e.g. educational and training interventions such as workshops, awareness panels and policy briefs. All these interventions are illustrated in this report throughout the coherent framework of behavior change.

Future work will complement the present policy recommendations and best practices, based on the ongoing cross-field regulatory analysis (for instance, concerning intellectual property aspects) and on the close collaboration with the researchers and technologists of the project that will elicit the challenges they encounter in the other working packages of BRIEF. The results will be published at the end of the project in an updated version of this deliverable.

1. INTRODUCTION

The legal-ethical framework that governs the multi-faceted biorobotics research activities of the BRIEF project is highly complex as it encompasses interconnected domain areas that can be organized coherently as: (Personal and non-personal) data management and data governance (see 1.5), Artificial intelligence law and governance (see 1.6), Regulation of medical devices and health law (see 1.7), Liability and insurance (see1.8), and Cybersecurity compliance and policy design (see 1.9). An initial mapping of the regulatory framework that highlights the complexity and interplay of the relevant legal provisions has been illustrated in the report dedicated to the Cross-field regulatory analysis (D7.3) and has emerged from the results of the survey on the stakeholders' needs (D7.2).

All these domains are characterized by intense lawmaking efforts both at the European and at the national level, which raise the necessity of comprehensively identifying and systematizing this growing body of rules. They also call for the provision of easy-to-follow practical instructions for researchers in biorobotics that need to navigate and apply such rules. Moreover, the considerable variation in terminology used to refer to the same concept across different regulations (e.g., the concept of interoperability) and the potential contrasts arising from the interplay between the provisions of applicable regulations governing similar aspects (e.g., on the grounds for admissible reuse of personal data) can give raise to legal uncertainty. An additional challenge is represented by the fact that many EU legislative proposals regarding technological aspects are still under negotiation within the EU Trilogue, while other approved regulations still need to be implemented into national laws or be adapted to the national legal system. As a consequence, it is difficult to anticipate the outcomes of such developments and put in place the necessary safeguards to engage in compliant-by-design research and innovation (R&I) activities. However, a proactive approach to legal compliance is necessary to carry out BRIEF's manifold experimental research tasks: since the early setting of any research and innovation activity, researchers need to have a clear understanding of the legal requirements that they need to respect and need to have the tools to address them efficiently, because such requirements may influence the very design of biorobotic devices and the exploitation of the results. A paramount example in this regard is the principle of privacy by design, which is a common practice of privacy engineers that has been formalized in international standards first (e.g., ISO 31700) and then included in the General Data Protection Regulation (Article 25) as one of the main overarching principles for lawfully developing applications and processes where personal data is involved.

It is for these reasons that the present deliverable offers two complementary types of contribution that have been developed by the Law and Policy Hub, i.e., a cohort of experts in relevant domains, that was set up as a first step of WP7 (see D7.1. "Set up of LaPoH"). On the one hand, the present deliverable provides policy recommendations that are mainly addressed to European law-makers and focus on specific, well-defined issues of the contemporary legal framework related to regulatory bottlenecks that hamper trustworthy R&I, for instance in terms of proposing how to redraft articles of legislative proposals that are under negotiation between the relevant European bodies. A plan for their dissemination is being currently drafted to increase their efficacy by enhancing the chances that such recommendations are considered by the relevant decision-makers (see also D7.7 "Report on Research Dissemination and Awareness").

On the other hand, this deliverable contains best practices for researchers that offer guidance and translate into actionable instructions the high-level requirements of relevant regulations. The dissemination plan contained in D7.7 also includes a strategy to ensure that the best practices and the policy briefs geared towards BRIEF researchers are communicated in a way that positively affects their activities. Lastly, this deliverable also systematizes a wider set of interventions that encompass policy recommendations and best practices with the goal of enabling the development of compliant-by-design outcomes of biorobotic research.

This deliverable must be understood as a living document as it is the first iteration of the final report on policy recommendations and best practices that will be published at the end of the project in March 2025. The final version of the report will contain a more exhaustive mapping of the relevant topics, policy areas and best practices that are relevant to the various BRIEF's R&I activities which will be generated by the ongoing cross-field regulatory analyses and by the close collaboration with the researchers and technologists pertaining to the other WPs.

This report is organized as follows. Section 2 presents a brief overview of the relevant regulations, illustrates the framework of interventions that can be applied to BRIEF and explains the methodology that has been adopted to produce the policy recommendations and the best practices that are illustrated in Section 3. Section 4 provides the roadmap for delivering an updated set of policy recommendations and best practices in the next version of the report, planned for March 2025.

2. METHODOLOGY

1.1 BRIEF's relevant regulatory frameworks

BRIEF foresees the creation of a comprehensive decentralized infrastructure with innovative laboratories and machineries for carrying out cutting-edge research in the fields of robotics and biorobotics on a wide range of projects. Even though at the time of writing the spaces and technologies that will be part of the infrastructure are still under construction, we describe hereby a selection of the research projects that are already under development and that have been illustrated by the BRIEF's technologists of WP3 (BioRobotics Science to Engineering Translation), WP4 (BioRobotics Platforms), WP5 (BioRobotics & Health) and WP6 (BioRobotics & Sustainability) to the technologists of WP7 during a collaborative meeting that had place at the Biorobotics Institute in Pontedera in November 2023.

The project is highly ambitious and spans across foundational components, platforms and applications of biorobotic research, thereby including a wide range of devices, machineries, and applications at the forefront of science. For instance, the Neuro-Robotic Touch Laboratory[1] studies the neuronal processes underlying the human sense of touch and engineers the artificial tactile sense. The Soft Mechatronics for BioRobotics Laboratory[2] develops soft, elastic and deformable systems, such as artificial organs (e.g., hearts, larynxes) and soft sensors for biomedical applications. The Healthcare Mechatronics Laboratory[3] designs and validates computer-integrated/assisted robotic systems that can assist surgical procedures. The Regenerative Technologies Laboratory[4] merges mechatronics, materials science and molecular

[1] https://www.santannapisa.it/en/institute/biorobotics/neuro-robotic-touch-laboratory
[2] https://www.santannapisa.it/en/institute/biorobotics/soft-mechatronics-biorobotics-laboratory
[3] https://www.santannapisa.it/en/institute/biorobotics/healthcare-mechatronics-laboratory
[4] https://www.santannapisa.it/en/institute/biorobotics/regenerative-technologies-laboratory

biology to develop new therapeutic systems for tissue and organ healing and regeneration. Even though many laboratories and projects concern medical robotics, wearable technologies, collaborative robotics, bioinspired robotics, neuroscience robotics, rehabilitation robotics and implantable technologies, BRIEF foresees the creation of laboratories that enable research beyond these disciplines, including underwater robotics, additive manufacturing, High Performance Computing systems and autonomous vehicles.

By establishing the Law and Policy Hub, WP7 provides a framework of support on the legal-ethical challenges that need to be addressed to enable trustworthy-by-design R&I in such fields (see Figure 1).

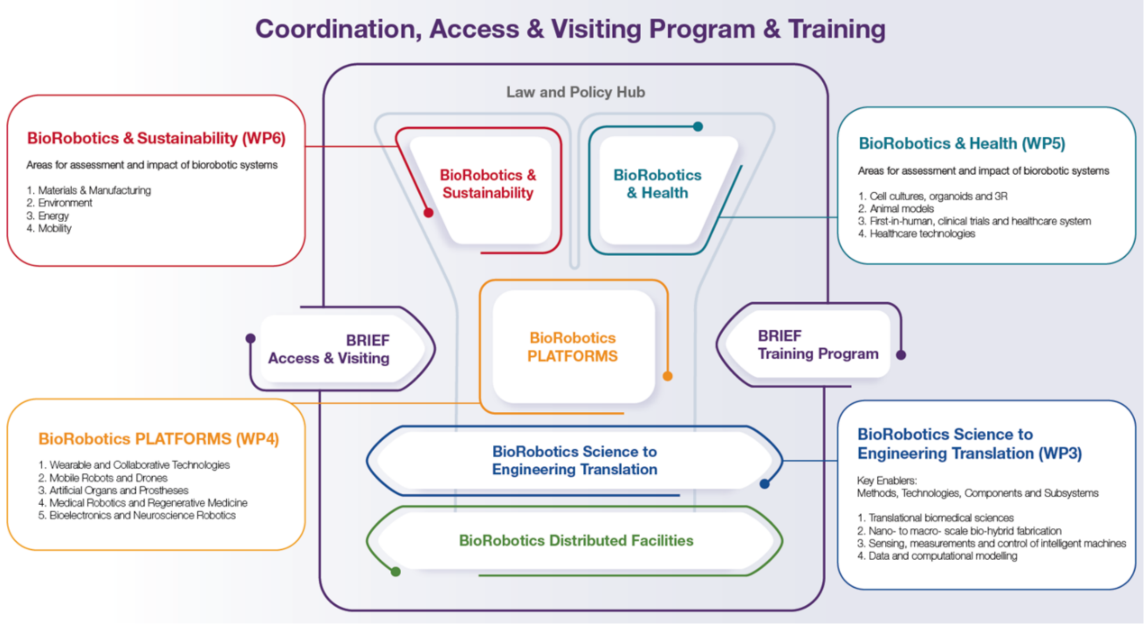

Coordination, Access & Visiting Program & Training
Law and Policy Hub
BioRobotics & Sustainability (WP6)
Areas for assessment and impact of biorobotic systems
1. Materials & Manufacturing
2. Environment
3. Energy
4. Mobility
BioRobotics & Sustainability
BioRobotics & Health
BioRobotics & Health (WP5)
Areas for assessment and impact of biorobotic systems
1. Cell cultures, organoids and 3R
2. Animal models
3. First-in-human, clinical trials and healthcare system
4. Healthcare technologies
BRIEF Access & Visiting
BioRobotics PLATFORMS
BRIEF Training Program
BioRobotics PLATFORMS (WP4)
1. Wearable and Collaborative Technologies
2. Mobile Robots and Drones
3. Artificial Organs and Prostheses
4. Medical Robotics and Regenerative Medicine
5. Bioelectronics and Neuroscience Robotics
BioRobotics Science to Engineering Translation
BioRobotics Distributed Facilities
BioRobotics Science to Engineering Translation (WP3)
Key Enablers:
Methods, Technologies, Components and Subsystems
1. Translational biomedical sciences
2. Nano- to macro- scale bio-hybrid fabrication
3. Sensing, measurements and control of intelligent machines
4. Data and computational modelling

Figure 1. The schema illustrates the interplay between the various WPs and the framework set by the Law and Policy Hub. Source: "Annex B – Part 2: BRIEF – Biorobotics and Innovation Engineering Facilties" of the grant application (p. 21). Available at: https://www.santannapisa.it/it/pnrr-santanna/brief

The broader regulatory framework

The first cross-field regulatory analysis that was reported in D7.3 has provided an initial mapping of the relevant legal requirements by examining EU regulations, their national implementations and the legislative initiatives that are currently under examination within the European trialogue. These legislative endeavours are part of the recent EU Commission's initiatives concerning digitalisation, datafication and innovation, such as the EU Digital Strategy[5] and the EU Data Strategy.[6] An additional aspect that was highlighted concerns the complex ethical values that govern biorobotics research and that are transposed into general or sectorial administrative procedures (e.g., ethical committees' authorization processes). Within the BRIEF' context, special emphasis must be placed on the secondary use of health data and on data-informed biomedical applications that are based on AI (e.g., machine-learning based diagnostics), for which there is the need to establish a common framework that facilitates and regulates the performance of clinical trials and the development of safe-by-design medical

[5] https://digital-strategy.ec.europa.eu/en

[6] Communication from the Commission to the European Parliament, the *Council, the European Economic and Social* Committee and the Committee of the Regions, "*A european strategy for data*", COM(2020) 66 final. For a general overview, see also https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

devices. Research and innovation need to uphold fundamental rights and protect research participants, including vulnerable populations, for instance by ensuring the confidentiality of data, while striving to adhere to the principle of openness that foregrounds the need for replicable experiments.

Whereas we refer the reader to the in-depth analysis reported in D7.3, we summarize here the main outcomes in terms of applicable laws that need to be examined to understand enablers and challenges to be addressed.

Data laws. The General Data Protection Regulation[7] which sets harmonized rules for the collection, use and reuse of personal data, including special categories of data such as health data. The Regulation on the Free Flow of Non-Personal Data[8] represents the counterpart of the GDPR, that intends to encourage and govern the free movement of non-personal data across borders by abiding to cybersecurity requirements. The Data Governance Act[9] sets up novel mechanisms meant to enhance trust in data sharing and overcoming technical barriers to the reuse of data, for instance the secondary use of publicly held data such as health data. This is why it sets up common data spaces that consist in protected, interoperable data storage and exchange infrastructures in strategic domains, including health. In this respect, the European Health Data Space proposal[10] lays down rules, standards, and practices for the primary use of data, as well as secondary use of data. The Data Act proposal[11] establishes requirements addressing how private subjects can access IoT-generated personal and non-personal data and business data, with one of its pillars being interoperability.

Public health framework. Regulation (EU) 2017/745 on Medical Devices[12] and Regulation (EU) 2017/746 on In Vitro Diagnostic Medical Devices[13]recently entered into force after a postponement due to the Covid pandemics. The Medical Devices Regulation organizes medical devices in different classes of risk which determine whether and how such devices need to undergo certification and audits procedures before their entry into market. The Clinical Trials Regulation has the main objectives of enhancing the efficiency of conducting multinational trials and providing transparency to clinical trials data and processes. The regulation establishes that an authorization to proceed with the trial is required stemming from a thorough scientific and ethical review with the involvement of an Ethics Committee. The procedure to obtain

[7] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)
OJ L 119, 4.5.2016.

[8] Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union.

[9] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance) PE/85/2021/REV/1 OJ L 152, 3.6.2022, p. 1–44.

[10] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space. COM/2022/197 final.

[11] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act) COM/2022/68 final.

[12] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance.) OJ L 117, 5.5.2017, p. 1–175.

[13] Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU (Text with EEA relevance.) OJ L 117, 5.5.2017, p. 176–332.

authorization is complex and encompasses aspects related to risks and benefits for public health and research participants, informed consent, recruitment etc.

The emerging regulatory framework on AI. The AI Act proposal[14] is a risk-based regulation that strives to lay down harmonized rules for the development and deploying of AI systems. It mandates the creation of various risk categories for AI systems: depending on the level of risk that they pose, such applications will be governed by more or less stringent rules or banned altogether. Complementary to the AI Act, the AI Liability Directive[15] institutes uniform requirements for non-contractual civil liability concerning damages caused with the involvement of AI systems.

Cybersecurity. The legal framework encompasses the NIS Directive,[16] the NIS2 Directive,[17] and the Cyber Resilience Act proposal.[18]

Liability. The Product Liability Directive proposal[19] concerns the liability of defective products and revises the existing Product Liability Directive 85/374/EEC.

Safety. The Machinery Regulation[20] establishes health and safety requirements for the design and construction of machinery. The General Product Safety Regulation[21] modernises the EU general product safety framework and addresses the challenges posed to product safety by the digital economy.

In the future, this framework will be complemented with other relevant legislations, for examples those concerning copyright, patents and design rights.

From compliance as a duty to compliance as an ethos and good practice. The cross-field regulatory analysis reported in D7.3 identified three common tenets that underpin most of the cited regulations and that can act as general guiding principles of trustworthy R&I in biorobotics: accountability, fairness and transparency. These three principles are indeed the subject of many of the best practices and policy recommendations of this report, also because the development of technologies can implement them in various manners, without necessarily

[14] Proposal for a Regulation of the European Parliament and of the Council laying down Harmonized Rules on Artificial Intelligence (Artificial Antelligence Act) and amending centrain Union Legislative Acts, COM(2021) 206 final.

[15] European Commission, 'Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive), COM(2022) 496 Final'.

[16] Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union.

[17] Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive) .

[18] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on horizontal cybersecurity requirements for products with digital elements and amending Regulation (EU) 2019/1020. COM/2022/454 final.

[19] Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on liability for defective products COM/2022/495 final.

[20] Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC.

[21] Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council and Directive (EU) 2020/1828 of the European Parliament and the Council, and repealing Directive 2001/95/EC of the European Parliament and of the Council and Council Directive 87/357/EEC.

converging. Even though the concrete application of such principles in scenarios at the forefront of science and practice provokes lively debates, they can serve as guidance for researchers to overcome the regulatory uncertainty that was illustrated earlier and the legislative pace that is often slower than technological advancements.

For instance, even though there is general agreement on the discrimination risks provoked by biased automated decision-making systems, there are ongoing discussions on what, vice versa, constitutes **fairness** in AI applications and how such concept might be implemented in the metrics and techniques that are employed in contexts where such applications are increasingly used to take decisions that have serious implications on human lives, such as medical diagnoses and treatment. One of the policy recommendations deals exactly with the challenges of lawfully translating principles of justice into machine learning pipelines. Another concern that we address related to the fairness of AI-based applications regards the prohibition of those systems that employ deceptive and manipulative techniques. Academic literature and practice are recently unveiling the many ways in which AI systems can manipulate users. However, the punctual definition of such techniques is problematic as it risks being overinclusive or underinclusive, thereby hampering the legal certainty that underpins innovation.

Closely related to the concept of fairness is that of **accountability**. As recalled in one of the policy recommendations concerning this principle, accountability is concerned with fair and equitable governance and is thus an underlying notion of responsible innovation, as it serves diverse the regulatory goals of compliance, reporting, oversight, and enforcement. Accountability needs, however, a clear allocation of roles and responsibilities which is still undergoing, especially when it comes to AI systems.

Lastly, there is no accountability without **transparency** about practices, processes, and outcomes. This is why we also offer practical guidance for researchers on how to concretely provide transparent information to research participants about the management of their personal data. Other guidelines illustrate how to make AI applications explainable, with the goal of enabling their users (such as the medical personnel) to understand, and question, if necessary, the underlying functioning of automated decision making so that, for instance, algorithmic discrimination can be more easily avoided.

1.2 Challenges and interventions to encourage compliant behavior

In such a complex scenario, ensuring compliance of the BRIEF's biorobotic research activities with all the applicable laws, as well as their conformity with relevant research ethics principles, constitutes a great (research) challenge. Developing a unified, coherent understanding of the interplay of the various legal provisions in an ever-evolving national and international legal framework and their applicability to concrete cutting-edge biorobotic use cases is a complex exercise. This task is even more challenging considering that the legal framework of the European digital strategy is still being defined and many regulations are still being drafted at the moment of writing this report. This is why boiling down such complexity to lean, simple, coherent instructions and best practices for researchers is not a mundane task. Moreover, to ensure compliance, it is paramount to understand how legal norms apply in practice in the specific context at hand: in the end, norms do not only regulate research activities, but also the behavior of research scientists. In other words, the question on how to make compliance tasks easier practically concerns **people, their behaviors, and the organizational structures** they work in. It is by enabling people to accomplish certain tasks with certain purposes in a feasible manner that the many research and innovation activities and the various devices, software, data and products that are therein used and developed can become compliant with applicable laws.

This is why we find it useful to describe the underlying process for supporting this goal in the terms adopted by the **framework of behavior change**. In particular, the behavior change wheel[22] offers a systematization of useful concepts and affordances that have been applied to many domains where target behaviors need to be encouraged, for example in terms of compliance of medical personnel's practices with the hospital policies to enhance the wellbeing of patients and of patience's adherence to medication;[23] similarly, it has been applied to enable employees to more easily follow the cybersecurity policies of their organization[24] and thereby decrease the cyber-risk to which it is exposed.

The success of this model probably stems from the fact that is simple while being exhaustive, and so versatile that it can explain how human behavior works, while planning a set of possible **interventions with various functions that can encourage (or discourage) a certain target behavior**. In their seminal work based on a literature review of other major behavior models, Michie, van Stralen and West[25] identify 3 main components of behavior, summarized in what they called the Capability, Opportunity, Motivation Behavior model (the COM-B model). In a nutshell, behavior is influenced by:

1. **Capability** (individuals' capacity):
    a. physical capability (skills)
    b. psychological capability (knowledge, skills)
2. **Motivation** (broadly defined as all the brain processes that direct behavior):
    a. Reflective motivation (plans - intentions; evaluation - beliefs)
    b. Automatic motivation (emotions; desires; impulses)
3. **Opportunity** (factors that lie outside the individual):
    a. social opportunities (intrapersonal influences, socio-cultural norms)
    b. physical opportunities (environmental affordances; time; resources; location).

All components are necessary to achieve a target behavior, apart from reflective thinking.[26]

If we apply this model to the challenges posed by the compliance of BRIEF researchers with applicable laws, it becomes clear how all these components are necessary to ensure that certain requirements are respected, and rules applied correctly. Let us illustrate this with a concrete example that fits within this context. Research scientists need to have the *knowledge* that the personal data they gather in their experimental studies must be protected through appropriate organizational and technical safeguards to be able to apply such safeguards, such as encryption. They also need to have the right *skills* to do so e.g., to perform the technical operation of encrypting the data in a specific manner that ensures their confidentiality. If researchers do not

[22] Susan Michie, Maartje M van Stralen and Robert West, 'The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions' (2011) 6 Implementation Science 42 https://doi.org/10.1186/1748-5908-6-42.
[23] See for instance, Nicole Chiang and others, 'Interactive Two-Way mHealth Interventions for Improving Medication Adherence: An Evaluation Using The Behaviour Change Wheel Framework' (2018) 6 JMIR mHealth and uHealth e9187 <https://mhealth.jmir.org/2018/4/e87> accessed 1 December 2023.
[24] See for instance, Moneer Alshaikh and others, 'Toward Sustainable Behaviour Change: An Approach for Cyber Security Education Training and Awareness', 27th European Conference on Information Systems: Information Systems for a Sharing Society, ECIS 2019 (Association for Information Systems 2020) <https://ksascholar.dri.sa/en/publications/toward-sustainable-behaviour-change-an-approach-for-cyber-securit-2> accessed 1 December 2023.
[25] Michie, van Stralen and West (n21) 4.
[26] ibid 4–5.

have those skills within their team, then appropriate *resources* should be dedicated to acquiring those skills (e.g., through the acquisition of an encryption software) or requiring others (such as a person or a company with the required expertise) to encrypt the data. Further, in an organization where there is the *socio-cultural norm* to encrypt personal data for their storage and such norm is taught by senior researchers to early career ones as part of their tasks, it is going to be easier to conform to such norm and enact it, when compared to an organization where such norm is not established, and senior researchers disregard it. In other words, although the protection of personal data should theoretically be implemented based on legal norms that are applicable in a certain jurisdiction, the social reality is even more influential in the effective application of such norms in a certain context.

However, researchers need to be *motivated* to engage in such behaviors, as compliance constitutes an additional effort that is not necessarily perceived as pertaining to their usual (research and administrative) tasks. Motivation is also key: without it, even if there is the material capacity to do so, researchers would not adopt any behavior to be compliant. Such motivation can be *reflective* when researchers are persuaded of the benefits of protecting data and thus intend to do so, for instance because they can consequently avoid risk of e.g., bad publicity and public mistrust; it can become *automatic* whenever such motivation is internalized and routines are formed, for example by institutionalizing processes for compliance checks.

Interventions that aim to promote or deter a target behavior can be of various nature:[27]

1. **Education**: increasing knowledge and understanding
2. **Persuasion**: using communication to induce positive or negative feelings to stimulate actions
3. **Incentivization**: creating an expectation of reward
4. **Training**: imparting skills
5. **Enablement**: increasing means / reduce barriers to increase capability (beyond education) or opportunity (beyond environmental restructuring);
6. **Coercion**: creating an expectation of punishment or cost
7. **Restriction**: using rules to reduce the opportunity to engage in the target behavior
8. **Environmental restructuring**: changing the physical or social context
9. **Modelling**: provide examples to aspire or to imitate

The first five intervention typology places the emphasis on personal agency, while the other four focus on external resources. Each intervention can be implemented through specific fine-grained techniques that address one or more specific components of behavior and may serve various intervention functions.[28] The techniques that implement the general interventions pertain to the broader family of policies that can be summarized as follows:[29]

1. **Communication**: using media
2. **Guidelines**: creating documents that recommend or mandate practice
3. **Fiscal:** using the tax system to increase or decrease the financial costs
4. **Regulation**: establishing rules or principles of behavior or practice
5. **Legislation**: making or changing laws

[27] ibid 7.

[28] ibid 8.

[29] ibid 7.

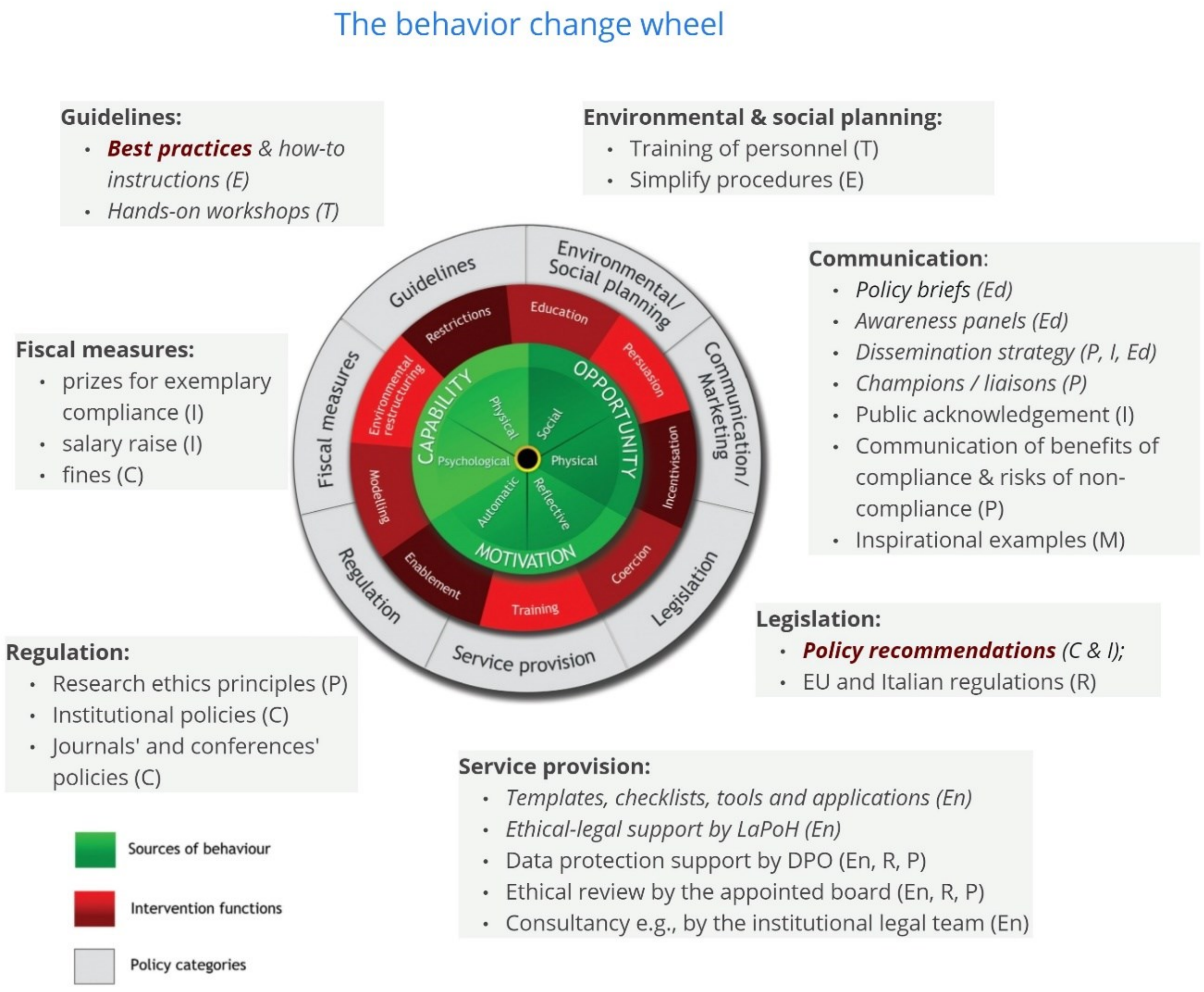
6. **Environmental / social planning**: designing and / or controlling the physical or social environment (including *nudges*)
7. **Service provision**: delivering a service.

There is no fixed formula for facilitating the compliance of R&I activities: rather, we should aim for a thoughtful mix of intervention techniques that achieve various objectives.

1.3 A framework of interventions for BRIEF's specific compliance challenges

Given the general methodological framework provided by the behavior change wheel, we have devised specific techniques of interventions that cover various functions and variously address the goal of facilitating the compliance of biorobotic engineering researchers with the normative framework that has been briefly reported in Section 1.1. As mentioned, that initial regulatory analysis constitutes a living document that will be continuously updated throughout the rest of the project. Its outputs will be included in D7.4 and D7.5 Cross-field Regulatory Analyses. Although this report only contains two types of interventions (i.e., policy recommendations and best practices), we find it useful to delineate in these pages the overall strategy that WP7's members are devising and putting in place. Such a strategy comprises additional types of interventions that we are designing and is complemented by other actions that are outside of our remit. Those interventions cover a broad range of functions, including incentives and disincentives, and can enhance the capability, the opportunity, or the motivation of researchers to comply with relevant norms.

Figure 2 provides an overview of the intervention techniques that can be applied to the context at hand and that are detailed in the following sections. In a nutshell, the **policy recommendations** that the LaPoH is developing on a broad range of relevant topics at the forefront of technological innovation are meant to influence the ongoing process of legislation, and therefore the final legislative texts that will enter into application, or to highlight critical points that call for legislative reform. They can have both a coercive and incentivizing function on behavior. The **best practices** under development aim at providing relevant, practical instructions that are designed for specific audiences that have specific needs. This is why best practices have the goal of enabling certain behaviors. Whereas policy recommendations address the abstract, general level of rules, the best practices instantiate those rules in specific contexts for specific people that need to comply, i.e., to behave in a desired manner. In addition, there are *services* that the LaPoH can establish (e.g., checklists), as well as **communication strategies** that use various media to raise awareness on e.g., the project-generated knowledge and the existence of the services (reported in italics in Figure 2 and discussed below). There are several other complementary solutions that already exist or can be implemented by actors other than the LaPoH. Examples of such solutions are shown on the Image but will not be illustrated in this report because the members of the project do not have a direct influence on such incentives.



Guidelines:
- ***Best practices*** *& how-to instructions (E)*
- *Hands-on workshops (T)*

Environmental & social planning:
- Training of personnel (T)
- Simplify procedures (E)

Communication:
- *Policy briefs (Ed)*
- *Awareness panels (Ed)*
- *Dissemination strategy (P, I, Ed)*
- *Champions / liaisons (P)*
- Public acknowledgement (I)
- Communication of benefits of compliance & risks of non-compliance (P)
- Inspirational examples (M)

Fiscal measures:
- prizes for exemplary compliance (I)
- salary raise (I)
- fines (C)

Legislation:
- ***Policy recommendations*** *(C & I);*
- EU and Italian regulations (R)

Regulation:
- Research ethics principles (P)
- Institutional policies (C)
- Journals' and conferences' policies (C)

Service provision:
- *Templates, checklists, tools and applications (En)*
- *Ethical-legal support by LaPoH (En)*
- Data protection support by DPO (En, R, P)
- Ethical review by the appointed board (En, R, P)
- Consultancy e.g., by the institutional legal team (En)

Figure 2. A non-exhaustive list of interventions that can facilitate compliance of researchers with legal and ethical norms governing biorobotics R&I, organized in the categories proposed by Michie, van Stralen and West[30] In italics, the intervention techniques that WP7 is putting in place to this end, whereas the other ones are techniques that may have an influence but are outside of our remit. In dark red bold characters, the interventions that are reported in this deliverable (i.e., policy recommendations and best practices). In brackets, the letters refer to the intervention functions that each technique can cover. Modified image from Michie, van Stralen and West[31]

Table 1 summarizes the range of interventions that are planned within BRIEF's WP7 and that adhere to the categories identified by Michie, van Stralen and West[32] and illustrated in the previous section. As it can be noticed, the interventions we laid down mainly aim at providing the means (*enablement*), the incentives (*incentivization*) or the restrictions (*coercion*) to promote a target behavior. Some interventions are meant to increase the knowledge and understanding of various stakeholders (*education*), while others are about the development of skills (*training*). The specifics are explained in the following paragraphs.

[30] Ibid 7.

[31] Ibid 7.

[32] Ibid 7.

Table 1. An overview of the techniques of interventions that are planned in WP7. In italics, the intervention techniques that are the object of this report (i.e., policy recommendations and best practices).

Intervention type	*Specific technique(s) adopted in BRIEF*	*Intervention function*
Legislation	*Policy recommendations*	Incentivization or Restriction
Guidelines	*Best practices* & how-to instructions Hands-on workshops	Enablement Training
Service provision	Templates, checklists, tools and applications Ethical-legal support by LaPoH	Enablement Enablement
Communication	Policy briefs Awareness panels Dissemination strategy Champions / liaisons	Education Education Persuasion, incentivization & Education Persuasion

Policy recommendations

The policy recommendations that have been developed by various members of the LaPoH (see Section 3) aim to uphold legal certainty and thus enhance compliance of the interested parties by identifying those aspects of the regulatory framework that necessitate modification to foster the development of trustworthy research and innovation activities, for example because there is lack of terminological clarity in the provisions, because there are contradictions between the provisions of different regulations concerning similar aspects or technologies, or because the implementation of certain provisions appears limited by practical constraints. Policy recommendations are hence understood as a type of legislative intervention that modifies the environment of action for biorobotic researchers with the objective of making it easier for them to implement practices that adhere to the appropriate rules and requirements, while making it harder for them to violate the relevant obligations.

In this respect, the contributions of LaPoH's members span across various topics of relevance, including: the definition of specific requirements for data portability that are meant to solve the terminological confusion adopted by many legislations and legislative proposals within the European Digital Strategy (Policy Recommendation 1 – PR1); a clarification of the roles and responsibilities of the actors that are involved in the accountability measures established for AI (PR2); the redefinition of the concept of justice that underlies that of fairness in machine learning so that it the metrics and techniques that are employed in this regard are compliant with EU anti-discrimination laws (PR3); a proposal for increasing the terminological clarity about subliminal, manipulative and deceptive techniques of the AI Act to overcome potential under- or over-encompassing definitions (PR4); a solution to the issues of uncertainty and slowdown that is caused by the Medical Device Regulation's regulatory process and the lack of notified bodies (PR5); a proposal for extending the liability of manufacturers of defective components to importers and authorized representatives to ease the process of consumers' compensation (PR6); a revisitation of the concept of personal injury compensation within the robotic context (PR7); a recommendation for a clearer involvement of the ENISA (European Union Agency for Cybersecurity) in the official definition of emerging cybersecurity issues in AI (PR8); the introduction in the AI Act proposal of a deadline for the reconsideration of the adopted standards and common specifications to account for technical developments and emerging cybersecurity threats (PR9). These policy recommendations are timely and relevant, since they mostly address legislation that is currently being negotiated within the European trialogue or that is yet to be implemented into national laws, and there is therefore space for influencing the legislative process.

Best practices

The best practices that are being drafted aim at providing practical guidance to BRIEF's members by helping them navigate and interpret relevant legal provisions in their application to their day-to-day R&I tasks. This is particularly challenging whenever what constitutes a good practice is being defined in a novel field of practice: before being able to recommend best practices, standards of practice need to be conceived, applied, tested, discussed, agreed upon and disseminated. As a consequence, the current version of the report only contains two best practices concerning the transparent-by-design information disclosure about data practices (Best Practice 1 – BP1) and the implementation of explainability requirements in automated decision-making applications deployed in the biomedical domain (BP2). Such best practices will be complemented in the next iteration of the deliverable at the end of the project. They are meant to be hands-on, relevant and designed for the needs and capacities of their intended audience.

In order to succeed, an iterative process of design of such best practices has been put in place: starting from the results of the survey carried out over spring 2023 and reported in D7.2 "Engagement strategy", a list of prioritized legal-ethical needs of the researchers in the other WPs were elicited. Briefly, the results show that researchers have doubts and seek help mainly about issues related to intellectual property, Clinical Trials Regulation, Medical Devices Regulation, health data management, contractual matters and CE certification. These findings need to be complemented with the punctual observations that arise from the close collaboration between the technologists who have legal-ethical expertise pertaining to WP7 and the technologists who have technical expertise pertaining to the other WPs. Two meetings have been held so far (in October and November 2023) to start exploring the specific technological development requirements within the research projects carried out by the various laboratories that are involved in BRIEF. Even though more of these collaborative opportunities will likely be planned in the upcoming months to better clarify the specific needs and co-devise applicable solutions (e.g., in terms of hands-on workshops), the outcome of the first two meetings already highlights the additional necessity to explore the re-use of health data (e.g., CT scans; patients' audio data, etc.) for research purposes and the need to analyze the role and the risk level of the various AI applications deployed within these research projects. The list of needs is open-ended; however, through the close collaboration with the technologists, a finite list of priorities will be set to enable the efficient addressing of the raised issues.

Additional interventions planned in WP7

Policy recommendations addressed to national and international policy-makers, as well as best practices addressed to researchers, are accompanied by a set of additional techniques that are meant to encourage compliance. First, there are a number of actors that are internal to the Scuola Superiore Sant'Anna, its institutes and the other organizations involved in BRIEF that can support the compliance tasks by providing the necessary support and consultancy services.

The LaPoH is one of such actors that through the elaboration of best practices stemming from actual research needs of the specific projects and the relative domain knoweldge seeks to enable researchers to perform their tasks in conformance with relevant norms. Other actors that can support compliance in the performance of the research activities are the Data Protection Officer of the institution, the joint ethical review board and the institutional legal team. An additional way to provide support is through the provision or novel elaboration of checklists (e.g., the

ALTAI checklist[33] for the development of trustworthy AI; a checklist for the submission of all necessary documents to ask the ethical review board's authorization of research studies on animals or vulnerable populations; etc.), the design of templates (e.g., consent forms for participation to research studies; information sheets about data protection management), and the development of tools (e.g., an online data protection impact assessment tool).

Further, the outputs resulting from the research work carried out in WP7, for instance in terms of policy briefs and best practices, need to be disseminated strategically to ensure that the addressees know that they exist and where to find them. We may also want to increase the impact of the generated knowledge and material by devising complementary measures that address other relevant stakeholders. This is where the communication and dissemination strategy plays an essential role (for further details, see "D7.7 Report on Research Dissemination and Awareness activities"). Therefore, for instance, the policy briefs are sent to the technologists of the other WPs who act as informal ambassadors (or champions/liaisons) and drag their colleagues' attention to them; the policy briefs are also available on demand on the shared Teams folder so that they can be easily consulted whenever necessary; moreover, to increase their visibility, they are publicly disseminated through awareness panels and the LIDER Lab's website[34].

Timeliness of the communication is key for its effectiveness; this is why this material is proactively brought to the attention of those who may need it, but also available on demand on the shared repository. Complementary strategies can also be devised. Even though, as mentioned before, the regulatory framework around biorobotics research is under construction and subject to modification, thus the generated knowledge is under constant evolution, a similar procedure should be adopted for disseminating the best practices and the policy recommendations. As outlined in the dissemination plan, for example, the authors of the policy recommendations have been encouraged to submit them as op-eds in relevant venues where they can exert a timely influence on the ongoing scholarly and policy debate. Some of the policy recommendations may also be further developed in the chapters that will be part of the book on "Personalized Smart Medicine", as outlined in "D7.7 Report on Research Dissemination and Awareness activities".

Finally, there may be other interventions that can be useful to bioengineering researchers, even though they are not listed in Table 1 and will not necessarily be provided within the activities of WP7. For instance, in addition to the relevant national and international laws, there are internal procedures that researchers need to follow, for instance when it comes to the ethical approval for research studies that should abide by the internal policies established by their institution of affiliation. Such policies are in line with general research ethics policies that apply to disciplinary fields (e.g., computer science) or research contexts (e.g., internet research data) that should also be respected by researchers in the view of their accountability. It would be thus important to point out to researchers the sectorial and institutional policies that are in place and the specific actors that can support the carrying out of their activities within the boundaries established by such policies. For example, a summary of the steps to follow to respect such procedures may also be drafted (e.g., in the form of checklists), if the necessity arises from the discussion with the technologists and researchers of the other WPs. Given the number of

[33] https://altai.insight-centre.org/

[34] See the policy briefs already published on https://www.lider-lab.it/news/

institutions participating in the project and the foreseeable diversity of their internal policies, providing this kind of help may be challenging, though.

There may be additional measures that need to be taken by other relevant actors to strengthen the chances that researchers comply with relevant regulations. For example, it may become clear that certain internal procedures need to be simplified or that financial resources for obtaining *ad hoc* external consultancy need to be planned by the institution to which the research laboratories are affiliated. If these measures prove necessary, they may become part of policy recommendations included in the last iteration of this deliverable at the end of the BRIEF project.

1.4 Drafting and review process of the report

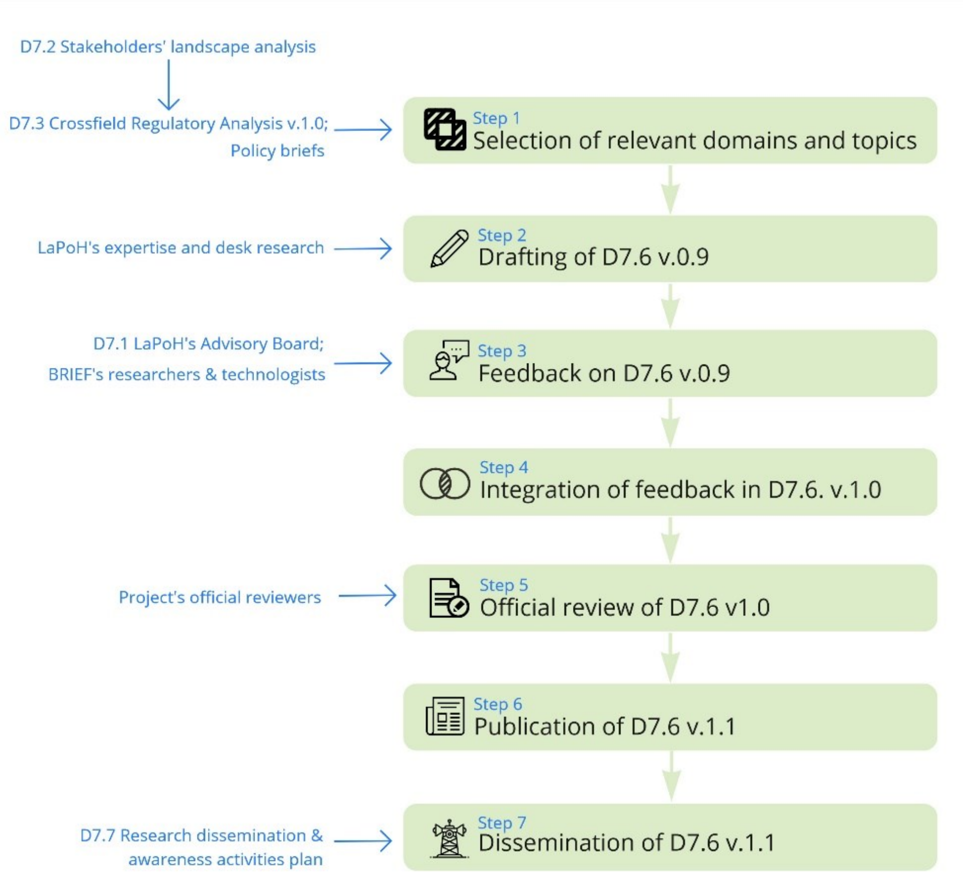

D7.2 Stakeholders' landscape analysis
D7.3 Crossfield Regulatory Analysis v.1.0; Policy briefs
Step 1
Selection of relevant domains and topics
LaPoH's expertise and desk research
Step 2
Drafting of D7.6 v.0.9
D7.1 LaPoH's Advisory Board; BRIEF's researchers & technologists
Step 3
Feedback on D7.6 v.0.9
Step 4
Integration of feedback in D7.6. v.1.0
Project's official reviewers
Step 5
Official review of D7.6 v1.0
Step 6
Publication of D7.6 v.1.1
D7.7 Research dissemination & awareness activities plan
Step 7
Dissemination of D7.6 v.1.1

Figure 3. Diagram representing the steps of the methodology that has been followed for preparing this report, as well as the next envisioned steps. In blue on the lefthand side, the relevant input sources

This report has been created thanks to a collective effort and the participatory input of the relevant stakeholders, as **Error! Reference source not found.** shows. Applicable domains and topics were selected based on the Crossfield regulatory analysis published as D7.3 that created a preliminary mapping of the national and EU regulations that may impact the R&I activities undertaken in the other WPs of BRIEF. Relevant input for the analysis was generated from the results of the survey investigating stakeholders and their needs carried out in D7.2.

Based on the multifaceted legal and ethical expertise of the members of the LaPoH spanning the key legal domains identified in the Crossfield Regulatory Analysis, a set of policy recommendations and best practices was collected by the authors of the report (D7.6 v.0.9). These contributions do not aim to cover all the needs that have been identified. Rather, they represent hot topics and/or under-researched topics on which the members of the LaPoH have a specific expertise on and can propose original contributions at the forefront of the international academic and policy discussion on the regulation of technologies that are relevant for BRIEF. Two different templates, reported in Appendix I and Appendix II, were created on purpose to elicit the specific problems that need to be addressed and provide a coherent structure to the proposed solutions (i.e., a policy recommendation or a best practice). The best practices and policy recommendations that were proposed underwent (at least) a double round of internal reviews carried out by the authors of the report who requested to the authors of the contributions to enhance the clarity and relevance of their contributions.

The draft version of the report (D7.6 v.0.9) was then subjected to three rounds of reviews. First, feedback was sought from the researchers and technologists with bioengineering background that work on the experimental WPs of BRIEF and who were asked to evaluate the content of the deliverable, and in particular the best practices, in terms of clarity and usefulness for their work. Another round of review was requested from the members of the LaPoH's Advisory Board since their expertise covers data protection law, health law, biomedical entrepreneurship and practice, and patient-centered views. A third round of review was requested from experienced members of the LaPoH covering various domains of expertise. The suggested revisions were integrated into version 1.0 of D7.6 that was then submitted for review. Any suggested edits coming from the official review will be integrated in the next iteration of the deliverable before publication and dissemination.

3. POLICY RECOMMENDATIONS AND BEST PRACTICES

As mentioned earlier, this is a working document. The policy recommendations (PR) and best practices (BP) that are reported in this section will be complemented with additional ones over the course of the project. In Section 4, important topics that will be included in the next iteration of this report are briefly summarized.

1.5 (Personal and non-personal) data management and data governance

(PR1) Rights to data portability: Define "portability levels" to clarify portability rights and obligations, especially for providers of digital products and services

Main author: Tommaso Crepax

Addressees:
The European Commission, through implementing acts or delegated acts; The European Commission, in its role as enforcer of competition rules; National Regulatory Authorities, (Market and Competition, Data Protection and Privacy, Communications, etc.); The European Parliament and Council.

Context / history of the problem:

Data portability is a fundamental concept of the European Commission's Data Strategy.[35] It empowers individuals by enabling them to control their personal data and to switch services at will. Data portability liberates both end-users and business users of digital services from the previously uncomfortable shackles of vendor lock-ins. Furthermore, it fosters innovation, allowing new entrants to venture into markets previously dominated by *de facto* monopolists with a stranglehold on data and related services. Data portability also facilitates the development of technical solutions that enhance interoperability between systems, even among data spaces of different sectors, and allows all interested stakeholders, including individuals, businesses, and public bodies, to extract value from ported data. Failing to implement data portability effectively would signify failing to realize the overarching Data Strategy. Therefore, realizing data portability is of paramount importance.

Definition of the problem:

Numerous regulations have attempted to activate data portability, but their results have been notably limited. In its initial form within the General Data Protection Regulation ("GDPR")[36], data portability lacked strength.[37] The challenges to its realization included:

(1) unclarities on textual interpretations of Article 20 GDPR,[38] such as what data is considered "provided by the data subject", what formats are structured, commonly used, and machine readable,

(2) conflicting rights related to data protected by various legal means, like personal dataset encumbered by personal data as well as intellectual property rights of others,

(3) limited awareness among individuals regarding their right to personal data portability,

(4) a shortage of alternative digital services (outside of those offered by the major tech giants) for data transfer, and

(5) a dearth of portability-ready information systems encompassing software, platforms, IoTs, hardware, and operating systems.

Consequently, a lack of portability requests further led to a lack of enforcement, as well as jurisprudence and scholarly attention. On their side, big tech players lacked economic incentives to open their monopolies and, due to the absence of penalties and potential competitors on the market, they enjoyed and benefited from the *status quo*. These historical problems, appreciable since 2016, continued to resurface in subsequent definitions of data

[35] EU Data Strategy (n6).

[36] GDPR (n7).

[37] Oscar Borgogno & Giuseppe Colangelo, *Data Sharing and Interoperability Through APIs: Insights from European Regulatory Strategy*, SSRN JOURNAL (2018), https://www.ssrn.com/abstract=3288460 (last visited Oct 17, 2023).

[38] Paul De Hert et al., *The Right to Data Portability in the GDPR: Towards User-Centric Interoperability of Digital Services*, 34 COMPUTER LAW & SECURITY REVIEW 193 (2018), https://www.sciencedirect.com/science/article/pii/S0267364917303333 (last visited Dec 17, 2021).

portability in newer regulations, such as those found in Article 6 of the Free Flow of Non-Personal Data Regulation of 2018 [39] and onwards.

Nevertheless, the legal framework surrounding data portability, as delineated by the evolving Data Strategy implementing regulations, remains dynamic. Some of the most recent regulations such as the Digital Markets Act[40] (DMA) and the Data Act[41] have yet to produce their effects, others, like the European Health Data Space[42], are pending publication or have yet to be drafted, like the upcoming Common European Data Spaces regulations. Moreover, some of these new legal acts empower the European Commission to adopt delegated and implementing acts in collaboration with relevant expert authorities, groups, businesses, NGOs, and other stakeholders, that specify and establish uniform conditions for the realization of data portability. This means that, as of now, no such specifications or uniform conditions exist.

Schweitzer and Metzger[43] have summarized that, although a general right to access data, which is an enabler and a precondition to data portability, generated by a user should be granted, there is no such right yet. However, there is a variety of access regimes, such as those outlined in the GDPR article 20, or –under certain conditions--competition law, sector-specific regulations,[44] the DMA, and the Data Act. This combination of access regimes is legitimately referred to as a "patchwork" that creates a conflicting interplay of rules, roles, and responsibilities, hampering legal certainty and, with it, the growth of economic investments. Such legal confusion around rules on portability affects every player in the digital economy, be it a consumer, a small business, a research facility, or a big tech giant.

Proposed policy recommendation aimed at solving the problem:

The EU legislative texts prescribing rights and obligations on data portability do not have a harmonized, commonly shared understanding of its layered concept. Each regulation seems to apply its own considerations as regards what it believes constitutes a “portable dataset”. For example, while the GDPR art. 20 deems portable a personal dataset that is made of data provided by the data subject and kept in a structured, commonly used and machine readable format, the Data Act art. 4, in turn, requires the data holder to “make available” to a third party any (personal and non-personal) data generated by a connected product or related service, without undue delay, easily, securely, in a comprehensive, structured (“s”), commonly-used

[39] Regulation on the free flow of non-personal data (n8):
“Art. 6, Porting of data
1. The Commission shall encourage and facilitate the development of self-regulatory codes of conduct at Union level (‘codes of conduct’), in order to contribute to a competitive data economy, based on the principles of transparency and interoperability and taking due account of open standards, covering, inter alia, the following aspects:
(a) best practices for facilitating the switching of service providers and the porting of data in a structured, commonly used and machine-readable format including open standard formats where required or requested by the service provider receiving the data; […].”

[40] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance) PE/17/2022/REV/1.

[41] Data Act (n11).

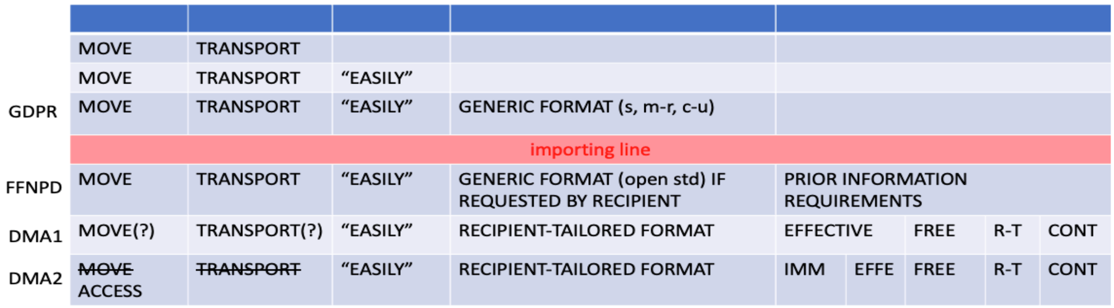
[42] European Health Data Space (n10).

[43] Heike Schweitzer & Axel Metzger, *Data Access under the Draft Data Act, Competition Law and the DMA: Opening the Data Treasures for Competition and Innovation?*, 72 GRUR INTERNATIONAL 337, 340 (2023), https://academic.oup.com/grurint/article/72/4/337/7072752 (last visited Oct 16, 2023).

[44] For example the Payment Service Directive 2, the EU Electricity Directive, and the Draft Access to Vehicle Data.

("c-u") and machine-readable ("m-r") format, as well as, if possible, in real time. The differences in the example --one of many (*see* table below)--show how data portability is defined differently in two regulations, a fact that heightens unclarity as regards to the rights of alleged rightsholders (who has right to what?), as well as to obligations of data holders (what technical implementations shall the information system have?).

The following table concisely summarizes the concept explained above. It shows how, thanks to a deconstruction of the concept of portability in its basic building blocks (*movability*, *transportability, ease of carry, …*), different regulations envision –sometimes defining--data portability diversely.



	MOVE	TRANSPORT							
	MOVE	TRANSPORT	"EASILY"						
GDPR	MOVE	TRANSPORT	"EASILY"	GENERIC FORMAT (s, m-r, c-u)					
	importing line								
FFNPD	MOVE	TRANSPORT	"EASILY"	GENERIC FORMAT (open std) IF REQUESTED BY RECIPIENT	PRIOR INFORMATION REQUIREMENTS				
DMA1	MOVE(?)	TRANSPORT(?)	"EASILY"	RECIPIENT-TAILORED FORMAT	EFFECTIVE		FREE	R-T	CONT
DMA2	~~MOVE~~ ACCESS	~~TRANSPORT~~	"EASILY"	RECIPIENT-TAILORED FORMAT	IMM	EFFE	FREE	R-T	CONT

Figure 4. Data portability spectrum

At its utmost basic level, the concept of data portability should embed the characteristics of data movability from one place and of transportability in a context dependent, sufficiently easy fashion. Keeping as starting points such foundational [45] blocks, regulations such as the GDPR, Free Flow of Non-Personal Data Regulation (FFNPD) and the DMA start going their separate ways. In fact, they each directly define or indirectly intend portability as a dataset to be treated differently, depending on, for instance, the need for awareness of the porting environment, the technical data format, the timing of service provision, and so on. For example, while in GDPR a controller could format porting datasets in a generic format while neglecting the receiving end, the FFNPD Regulation provides that the dataset should be formatted in a generic format, including open standard formats, but *if the recipient so requires* –therefore assuming the need of care for the receiving environment.[46] Moving further, the DMA cases of end user requests (DMA1 in the table above) and business user requests (DMA2 in the table above) bring altogether new issues: in the former, even though the text of article 6(9) refers explicitly to effective portability, what it describes in facts are means to access end users' generated data that, as such, do not necessarily require movability and transportability of the dataset; as for the latter, the reference to portability is not even explicit, and, again, the means described enable access to data, not portability. Hence, it can be argued that the DMA intends as portability something which is not such, as it lacks the definitional, foundational building blocks of movability and transportability.

In such a chaotic patchwork, what seems necessary is the deconstruction of the concept with a view to rebuilding it in a clearer, more streamlined, and organized fashion. Such a

[45] Definitional here means that, should an object such as a dataset not moveable and transportable to a contextually dependent, sufficient level of ease, it cannot be called portable.

[46] When moving from recipient agnostic to recipient aware portability requirements the "importing red line" is crossed.

deconstruction starts from the development of a toolset of conceptual building blocks to reconstruct and describe what each legislation understands as data portability. What follows is a blueprint of such toolset of concepts, specifically applied to descriptive levels that could be used to help answer the question: when is a dataset of one specific legislation considered portable?

- Level-0: The dataset is “movable” and “transportable” from one service to another.
- Level-1: the dataset is easily transferrable to a new environment to a sufficient degree.
- Level-2 (generic): the dataset is formatted in a fashion that is generically adaptable to a new environment (i.e., in commonly used, machine-readable, structured formats).
- Level-2 (specific): the dataset is extracted and managed in a format that is compatible with the specific new environment.
- Level-3: the dataset contains data that the porting environment can read with ease (syntactic-specific portability). The new environment should “read the sentence”, which, in machine readable terms means to be able to *read* the information (written in a similar or compatible programming language) and the logical structure of such information.
- Level-4: the dataset contains data that the receiving environment can understand and act upon (semantic-specific portability). The new environment should “understand the message”, meaning that not only it can read the information in their logical structure, but also understands the conveyed message.
- Level-5: the dataset is usable “upon request” and “in real time” by the new environment (real-time portability).

All this considered, the **policy recommendation** is the following*:*

Through delegated acts, the EC should acknowledge that data portability exists on a continuum or spectrum, which entails distinct levels (or types), and indicate as well as describe such levels. For each regulation, the EC should indicate what level of portability is required so that the portability rights are respected, and data holders know what is needed in their information systems to comply with portability requirements.

The policy recommendation has an historical parallel. In the realm of Autonomous Vehicles Regulation, it became necessary to highlight the existence of distinct levels of automation within self-driving cars and to establish specific rules for each level. Without tailored terminology to differentiate between levels, regulating all forms of automation uniformly would have yielded unreasonable consequences. A lack of distinction could have impacted safety at the societal level, hindered innovation within the market, and introduced various other complications. A similar approach should be considered in the regulation of data portability to ensure nuanced and context-appropriate guidelines.

The legislative acts should carry a clear indication that "Regulation/Directive [X] requires level X portability" and disclose a number of formatting options that are presumed compliant. It would be advisable for the aforementioned formatting options to be implemented through the mechanism of delegated acts. This approach leverages the fact that such acts can be subsequently modified by the Commission in response to technological advancements, while still preserving the general principles already established in the main text of the Regulation.

Without clear and specified levels, there is a risk that each participant in the digital market could interpret regulations in their own manner. This lack of uniformity could undermine the fundamental concept of data portability, which is the seamless reuse of data within the EU digital market. Establishing precise levels helps create a standardized understanding and implementation of data portability, fostering consistency and reliability across diverse players in the digital landscape, as well as balancing the diverging interests at stake. Without consistency and harmonization of data ontologies, formats, syntax, semantics, and best practices, there is a significant risk of encountering either substantial costs for the actual reuse of existing data (due to the necessity to sanitize and adapt it for each porting environment) or, even more critically, the loss of valuable data that cannot be effectively reused. Nomenclature standardization, meaning the process of standardizing different ways in which a concept shows itself, is not just a matter of convenience; it is a crucial factor in ensuring the efficient and meaningful exchange of data within the EU digital market.

Constraints of the policy recommendation:

The policy recommendations outlined above serve as blueprint, but they are not infallible and require additional research. For instance, it is essential to delve deeper into the question of whether the indicated levels should be viewed not as escalating numbers but rather as layers of characteristics that can be combined in several ways. In the case of autonomous vehicles, as they become progressively more autonomous, the numerical ordering of levels makes sense. However, there might be scenarios where a specific regulation calls for real-time portability coupled with generic data formats, essentially combining aspects of Level 2 and Level 5. This highlights the need for a flexible and nuanced approach that considers the interplay of different characteristics in regulatory frameworks.

(BP1) How to effectively inform study participants about personal data protection practices

Main author: Arianna Rossi

Addressees:

Researchers, medical personnel, and other relevant actors that are called to inform the participants to their research studies about their data protection practices. This also concerns those studies where personal data is not gathered directly from individuals, such as when datasets containing personal data and data gathered from the internet (e.g., scraped data) are employed. In such cases, when the direct provision of information about data processing to the involved individuals would prove impossible or constitute a disproportionate effort, researchers need nevertheless to make the information publicly available, for instance on the website of the research project.

Context of the problem:

The disclosure of information about the personal data that is gathered during research studies and the measures to manage such data is mandated by the obligations on transparency of Article 12, 13, and 14 of the General Data Protection Regulation[47] (GDPR) that aim at "engendering trust in the processes which affect the citizens by enabling them to understand, and if necessary, challenge those practices".[48] Prior to the GDPR, the Directive 95/46/EC[49] also mandated the disclosure of specific informational items to the individuals concerned by the personal data processing, such as the purposes of use of such data and the rights of individuals in that respect. [50] However, the resulting disclosure has often resulted in lengthy, verbose, obscure privacy policies[51] that have traditionally failed to properly inform the addresses of the disclosure. This is why Article 12 of the GDPR introduces provisions about the *manner* how the information items mandated by Articles 13 and 14 should be provided, namely "in a concise, transparent, intelligible and easily accessible form, using clear and plain language". These are user-centered transparency requirements that encompass the "quality, accessibility and comprehensibility of the information"[52] related to the data processing practices and the individuals' rights about their data. Transparency is now understood as a "user-centric rather than legalistic"[53] concept. This means that communications, be it privacy policies, consent forms or instruments for exercising data rights, should be designed to address the specific informational needs and the abilities of the intended audience,[54] as well as be subject to empirical tests to demonstrate their effectiveness.[55]

Definition of the problem:

The transparency obligations of the GDPR have given rise to a newly found interest in experimenting with new ways of communicating data privacy information. However, what constitutes transparent language may depend on the context and the audience: for example, a privacy-savvy knowledge may prefer legal jargon to what may be felt as oversimplified expressions, while sensitive contexts where deliberation can have severe implications such as the medical one may require more in-depth information rather than other contexts where disclosing personal data may have minor consequences. Moreover, Article 12 GDPR also suggests that providing "in an easily visible, intelligible and clearly legible manner" an overview of the data processing practices can be realized through the combination of textual content and standardized icons. Thus, the use of visual means to communicate complex

[47] GDPR (n7).

[48] Article 29 Data Protection Working Party, 'Guidelines on Transparency under Regulation 2016/679, 17/EN WP260 Rev.01. Adopted on 29 November 2017. As Last Revised and Adopted on 11 April 2018' 4 <https://ec.europa.eu/newsroom/article29/redirection/document/51025>.

[49] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

[50] The provision of information about the management of personal data is also an established practice in research ethics and is thus somehow overlapping with the data-related disclosure mandated by the GDPR. However, the sectorial or institutional research ethics policies may contain varying indications about the content of such disclosures and the required level of detail. The analysis of such policies is outside the scope of this contribution.

[51] For a more detailed overview of the hurdles to effective privacy communication, see Arianna Rossi and others, 'When Design Met Law: Design Patterns for Information Transparency' [2019] Droit de la Consommation = Consumenterecht : DCCR 79.

[52] Article 29 Data Protection Working Party (n 48) 5.

[53] ibid.

[54] Arianna Rossi and Gabriele Lenzini, 'Transparency by Design in Data-Informed Research: A Collection of Information Design Patterns' (2020) 37 Computer Law \& Security Review 3.

[55] Article 29 Data Protection Working Party (n 48) 7.

information is officially and groundbreakingly acknowledged as a valuable legitimate manner to enhance the transparency of the processing.

Guidelines from relevant independent authorities, for example the Guidelines on Transparency[56] by the Article 29 Working Party,[57] aim to ease the implementation of those legal requirements. Such guidelines provide useful interpretations about the transparency obligations, offer practical examples, and further suggest that additional visual means such as comics, pictograms, and animations[58] may be employed. However, these guidelines do not necessarily reach a researchers' audience, nor are they usable and easily navigable by them since they rather represent a useful source for an audience with legal expertise. Moreover, amidst many other research-related tasks, not every scientist has the skills, resources, time and motivation to design novel communications, experiment with them and test their efficacy with the intended audience. Other Data Protection Authorities, such as the Italian one, have organized public contests to design privacy icon sets,[59] but there has been no standardization nor guidelines for their implementation exist. Such a situation has created uncertainty as to what is permissible in terms of privacy communication design, rather than clarity.

How transparency-enhancing design patterns can solve the problem:

Researchers need shared, easy-to-implement, tangible solutions to commonly found problems in privacy communications: design patterns. Design patterns are not document templates that can be simply copy-pasted: they rather are systematized solutions that can be reused and readily adapted to new contexts. They constitute best practices that do not need to be evaluated individually, as they are solutions that are known to work in specific contexts. In the last few years, the research work carried out by researchers and practitioners[60] in this respect has been welcome by some data protection authorities, such as the French one (i.e., the CNIL) that has published a freely accessible online library of transparency-enhancing design patterns.[61] We invite the reader to explore the resources that are reported at the end of this piece since they contain many practical, visual examples, though we provide here some information to introduce the key points of such practices.

Design patterns can take on various functions that help enhance the transparency of privacy communication. Such functions go beyond improving the clarity of language and concern the broader user-centered design of communication. Design patterns are often collected in libraries that are organized according to those functions with the goal of helping the user to e.g., find the patterns they need to achieve a specific goal or to avoid a certain problem. This is why various ways of structuring libraries exist. However, the CNIL has proposed the first hands-on online library exclusively dedicated to the fulfilment of the GDPR's obligations on transparency through design patterns. Given the prominent role that this Data Protection Authority has had

[56] Article 29 Data Protection Working Party (n 48).

[57] The Article 29 Data Protection Working Party was an independent advisory board on matters related to data protection. Since the entry into force of the GDPR, it has been replaced by the European Data Protection Board.

[58] Article 29 Data Protection Working Party (n 48) 12.

[59] Icon sets available at: https://www.garanteprivacy.it/temi/informativechiare#2

[60] Rossi and others (n 51); Rossi and Lenzini (n 54); Arianna Rossi and Helena Haapio, 'Proactive Legal Design for Health Data Sharing Based on Smart Contracts', *Smart Contracts: Technological, Business and Legal Perspectives* (Marcelo Corrales, Mark Fenwick and Stefan Wrbka, Hart Publishing 2021).

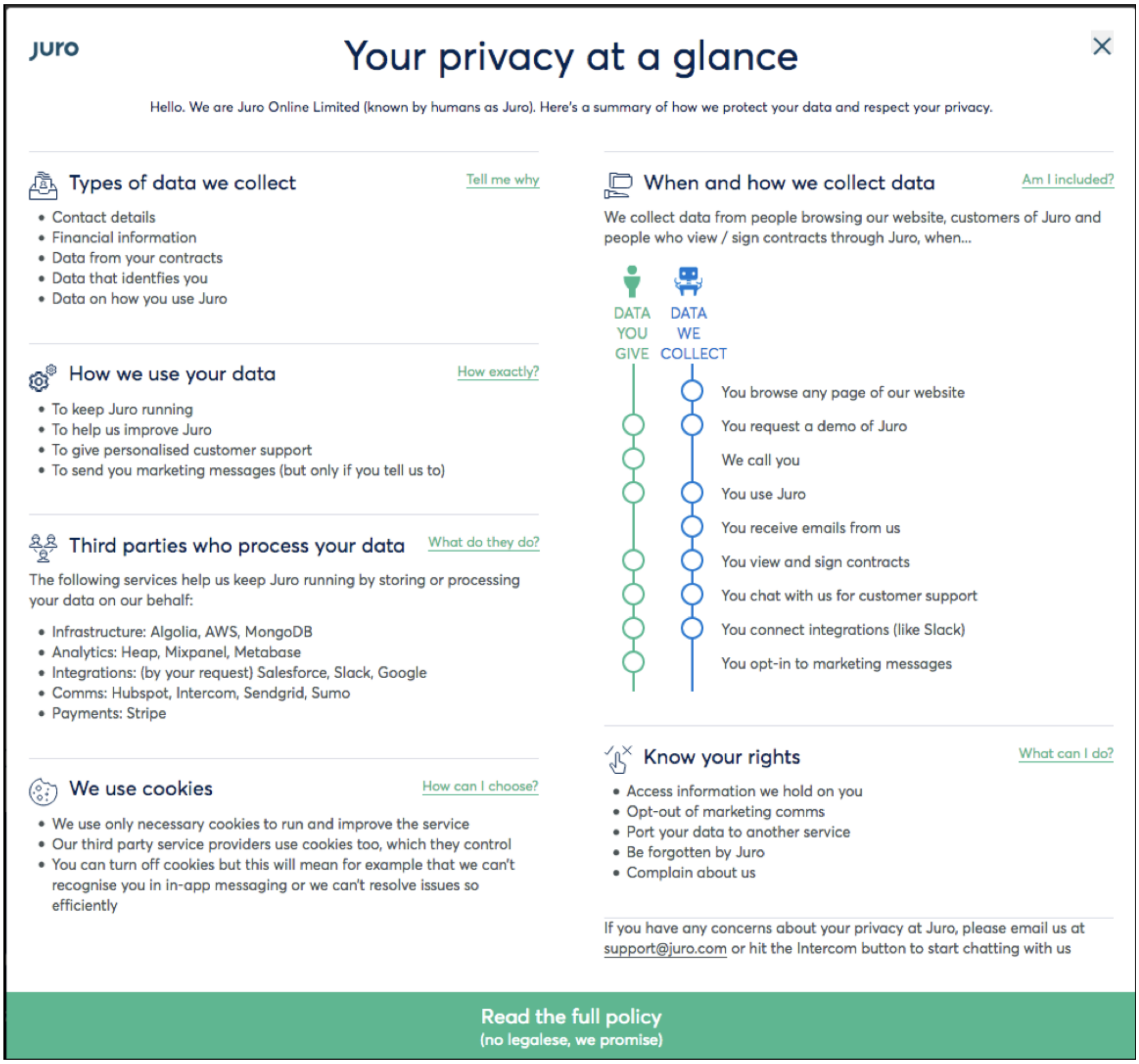
[61] Available at: https://design.cnil.fr/en/design-patterns/ (English) and https://design.cnil.fr/fr/design-patterns/ (French).

in addressing design issues in privacy[62] and the relatively simple arrangement of patterns in their library, we hereby provide a few functions and examples that follow the CNIL's categories and that can be viewed in Figure 5 (which is freely downloadable as template for online privacy policies):[63]

- **Structuring** (i.e., organizing information to facilitate skim reading): e.g., by structuring paragraphs logically by topic and introducing them with a short question as heading, as if they were FAQs.
- **Making it clear** (i.e., making information more understandable): e.g., by providing relevant examples that illustrate what legal or technical terms mean for the individual.
- **Summarising** (i.e., giving a brief account): e.g., by providing a short overview of the main content of a document as first layer, leaving the details to the second layer.
- **Drawing attention** (i.e., enabling people to quickly notice information): e.g., by using icons as information-markers that attract attention to the relevant section.
- **Browsing** (i.e., easing access to information and to the means to control one's data): e.g., by adding hyperlinks that support the navigation of a digital document.

[62] See e.g., the pioneering report dedicated to user-centered design in privacy: Régis Chatellier and others, 'Shaping Choices in the Digital World. From Dark Patterns to Data Protection: The Influence of UX/UI Design on User Empowerment' (CNIL-LINC 2019) <https://linc.cnil.fr/sites/default/files/atoms/files/cnil_ip_report_06_shaping_choices_in_the_digital_world.pdf> .

[63] Available at: https://github.com/juro-privacy/free-privacy-notice .



juro

Your privacy at a glance

Hello. We are Juro Online Limited (known by humans as Juro). Here's a summary of how we protect your data and respect your privacy.

Types of data we collect — Tell me why

- Contact details
- Financial information
- Data from your contracts
- Data that identfies you
- Data on how you use Juro

How we use your data — How exactly?

- To keep Juro running
- To help us improve Juro
- To give personalised customer support
- To send you marketing messages (but only if you tell us to)

Third parties who process your data — What do they do?

The following services help us keep Juro running by storing or processing your data on our behalf:

- Infrastructure: Algolia, AWS, MongoDB
- Analytics: Heap, Mixpanel, Metabase
- Integrations: (by your request) Salesforce, Slack, Google
- Comms: Hubspot, Intercom, Sendgrid, Sumo
- Payments: Stripe

We use cookies — How can I choose?

- We use only necessary cookies to run and improve the service
- Our third party service providers use cookies too, which they control
- You can turn off cookies but this will mean for example that we can't recognise you in in-app messaging or we can't resolve issues so efficiently

When and how we collect data — Am I included?

We collect data from people browsing our website, customers of Juro and people who view / sign contracts through Juro, when...

You browse any page of our website
You request a demo of Juro
We call you
You use Juro
You receive emails from us
You view and sign contracts
You chat with us for customer support
You connect integrations (like Slack)
You opt-in to marketing messages

Know your rights — What can I do?

- Access information we hold on you
- Opt-out of marketing comms
- Port your data to another service
- Be forgotten by Juro
- Complain about us

If you have any concerns about your privacy at Juro, please email us at support@juro.com or hit the Intercom button to start chatting with us

Read the full policy
(no legalese, we promise)

Figure 5. First layer of Juro's privacy policy, designed by Stefania Passera. Available at: https://stefaniapassera.com/portfolio/juro/

Constraints of the best practice:

There are two main constraints to the best practice of recurring to design patterns to enhance transparency of privacy communication. First, researchers need to devote resources (e.g., time) to their implementation within their specific context. However, there are free templates online and in commonly used software (e.g., PowerPoint, Keyword, etc.) that can be adapted to the specific needs, while online design pattern libraries as well as papers (see below) provide plenty of examples for inspiration. Researchers can also ask colleagues with the necessary skills to take care of such an aspect. Second, domain knowledge is needed to include accurate, reliable content about the data practices in the communication, for instance concerning the security measures that are adopted to protect the confidentiality of research data. Design patterns are containers for that kind of information, that should be developed together with domain experts, such as the Data Protection Officer of the institution.

In line with Article 25 GDPR that mandates data protection by design and by default, a transparency by design approach[64] implements transparency in the process of managing personal data. The transparent disclosure of such practices is simply the outcome of such an approach.

To know more about transparency-enhancing design patterns

- Contract design pattern library: https://contract-design.worldcc.com/
- CNIL’s design pattern library: https://design.cnil.fr/en/design-patterns/
- Rossi A and others, ‘When Design Met Law: Design Patterns for Information Transparency’ [2019] Droit de la Consommation = Consumenterecht : DCCR 79. Available at: https://orbilu.uni.lu/bitstream/10993/40116/1/A.%20Rossi%2C%20R.%20Ducato%2C%20H.%20Haapio%20et%20S.%20Passera.pdf
- Rossi A and Haapio H, ‘Proactive Legal Design for Health Data Sharing Based on Smart Contracts’, *Smart Contracts: Technological, Business and Legal Perspectives* (Marcelo Corrales, Mark Fenwick and Stefan Wrbka, Hart Publishing 2021). Available at: https://orbilu.uni.lu/bitstream/10993/49595/1/Rossi_Haapio-Proactive_legal_design_health_data_sharing_smart_contracts.pdf
- Rossi A and Lenzini G, ‘Transparency by Design in Data-Informed Research: A Collection of Information Design Patterns’ (2020) 37 Computer Law \& Security Review. Available at: https://www.sciencedirect.com/science/article/pii/S0267364920300078
- The Behavioural Insights Team, 2019. Best practice guide. Improving consumer understanding of contractual terms and privacy policies: evidence-based actions for businesses. Department of Business, Energy and Industrial Strategy of the UK. Available at: https://www.bi.team/publications/improving-consumer-understanding-of-contractual-terms-and-privacy-policies-evidence-based-actions-for-businesses/

1.6 Artificial intelligence governance

(PR2) The principle of accountability for responsible innovation

Main author: Irina Carnat

Addressees:

European Parliament, European Commission, Member States Parliaments, Market supervision authorities

Context:

In the specific context of technological innovation, accountability emerged as a core tenet of responsible innovation as a response to the inadequacies of the traditional regulatory and liability regimes regarding the new risks posed by technologies such as Artificial Intelligence

[64] Rossi and Lenzini (n 7) 3.

(AI), robotics, autonomous vehicles, etc.[65]. In fact, the rapid development and deployment of AI systems in high-risk sectors like healthcare, transportation, and criminal justice has raised concerns about their accountability. As AI systems become more complex, opaque, and autonomous, it becomes difficult to attribute responsibility when harm occurs. However, although the regulatory challenge regarding such disruptive technologies may be new, accountability tools are well-known and already established in the EU regulatory landscape[66], thus constituting an important policy foundation.

Definition of the problem:

The core problem is a potential accountability gap, caused by the so-called 'black-box problem', since their complex and opaque decision-making processes make it difficult to pinpoint responsibility for harmful effects. When AI systems are deployed for decision-making in certain critical areas, such as medicine, law enforcement or access to services, and the algorithmic outcome is incorrect, biased, erroneous or otherwise unpredictable, it's not clear whether the developers, the data, or the algorithms are at fault because the internal functioning of such systems are often opaque and not interpretable by humans. Although research has been concerned with developing tools and means to make AI systems more explainable[67], there are currently no comprehensive legal or technical mechanisms to ensure AI systems are sufficiently transparent. In fact, the EU's regulatory landscape is still ongoing, pending the adoption and the entry into force of three important pieces of legislation in the field of AI and robotics, namely the Proposed Regulation laying down harmonized rules on Artificial Intelligence ('AI Act')[68], the revised Product Liability Directive and an *ad hoc* AI Liability Directive[69]. In this context, the lack of clear allocation of roles and responsibilities along the complex AI value chain creates legal uncertainty that deters investment, puts citizens at risk of harm from unsafe systems, and does not incentivize – neither legally nor from a perspective of reputation benefits - developers of AI systems to comply with ethical requirements, ultimately undermining the societal trust in the technology and leading to its abuse, misuse or disuse.

Proposed policy recommendation:

The proposed policy recommendation leverages on the principle of accountability to achieve the desired legal certainty in the context of rapid technological development. Accountability is a multifaceted principle usually associated with fair and equitable governance. However, since it can serve a wide range of regulatory goals, it can be well adapted and implemented in any context where the decisions taken by an individual or a group impact a wider pool of individuals. As such, accountability can be defined as "*a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences*"[70]. Thus,

[65] European Commission, 'Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics COM(2020) 64 Final' (2020) <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0064> accessed 26 April 2023.

[66] Paul de Hert and Guillermo Lazcoz, 'When GDPR-Principles Blind Each Other: Accountability, Not Transparency, at the Heart of Algorithmic Governance' (2022) 8 European Data Protection Law Review (EDPL) 31 <https://heinonline.org/HOL/P?h=hein.journals/edpl8&i=37> accessed 26 June 2023.

[67] https://ec.europa.eu/research-and-innovation/en/horizon-magazine/opening-black-box-artificial-intelligence

[68] AI Act Proposal (n14).

[69] AI Liability Directive Proposal (n15).

[70] Mark Bovens, 'Analysing and Assessing Accountability: A Conceptual Framework1' (2007) 13 European Law Journal 447 <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0386.2007.00378.x> accessed 12 August 2022.

being accountable is seen both as a virtue, due to the deriving obligation to provide justification for a conduct, and as a mechanism, which allows for such accounts to be practically rendered to the forum[71]. It serves diverse regulatory goals, such as compliance with either legal or ethical standards; reporting, concerning the explanation and justification of the actor's conduct; oversight, i.e. the evaluation of the actor's conduct; and finally enforcement, with reference to the consequences the actor must bear following the reporting and oversight processes. It is a contextual principle that can assume multiple forms and dimensions based on the normative logic, the power relation between the actor and the forum, or the adopted substantive conception. Such principle is already applied across many regulatory domains, among which data protection: the GDPR at Article 5(2) regards accountability as a meta-principle, ensuring that the data controller indeed complies and provides proof of compliance with the set principles relating to the processing of personal data. More specifically in the EU's regulatory strategy, accountability is regarded as a principle requiring organizations to put in place appropriate technical and organizational measures to ensure and to demonstrate compliance with legal requirements[72]. Based on the normative basis of accountability, the actors shall face consequences if accounts are not rendered or insufficiently rendered: such consequences may be political, disciplinary, or legal, either in terms of liability for damages or criminal responsibility.

The proposed accountability toolkit, briefly described as follows, aims at achieving the goals of compliance, report, oversight and enforcement[73].

- Algorithmic impact assessments[74]: a structured evaluation process that examines the potential risks and consequences of the AI system's development and deployment on various aspects such as the environment, society, and the economy.
- Algorithmic audits[75]: a systematic examination and evaluation of records, statements, or processes to ensure accuracy, compliance with regulations or norms, and transparency.
- Harmonized standardization: the development by standardization organizations of technical standards that are mutually agreed upon and recognized across different entities or jurisdictions, the compliance with which ensure consistency and compatibility in products, services, or processes.

Although some of the proposed accountability tools are already envisioned in the AI Act, for instance, it is recommended to further clarify the roles and responsibilities of the actors involved, including consequences for failure to comply with regulatory obligations. While stricter accountability requirements may be justified for AI systems that, following an impact

[71] Mark Bovens, 'Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism' (2010) 33 West European Politics 946 <https://doi.org/10.1080/01402382.2010.486119> accessed 2 February 2023.

[72] European Data Protection Board: Accountability, available at: https://edps.europa.eu/data-protection/our-work/subjects/accountability_en#:~:text=The%20General%20Data%20Protection%20Regulation,and%20its%20effectiveness%20when%20requested, accessed 8 November 2023.

[73] Jennifer Cobbe, Michelle Seng Ah Lee and Jatinder Singh, 'Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2021) <https://dl.acm.org/doi/10.1145/3442188.3445921> accessed 24 November 2022.

[74] https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf

[75] https://www.adalovelaceinstitute.org/wp-content/uploads/2021/12/ADA_Technical-methods-regulatory-inspection_report.pdf

assessment, are expected to have a higher impact on safety and fundamental rights, it is nonetheless recommended that a minimum set of accountability measures shall be implemented for all AI systems, regardless of their level of risk, so as to guarantee a minimum level of documentation of the system's safety, as well as ex post redress in case of harm.

Constraints of the policy recommendation:

While strict regulatory requirements could apply only to high-risk AI applications, avoiding over-regulation of low-risk systems, it is worth noting that accountability principles benefit all innovators. Even in the absence of binding compliance requirements, documenting design choices and assessing potential impacts enables businesses to fulfill the burden of proof more effectively in potential liability cases for damages. An example of such an approach is the proposed regulation of foundation models, which, by definition, are suitable for a wide range of downstream tasks, therefore it is not possible to establish ex ante the level of risk. The amendments to the original text of the AI Act proposed by the European Parliament in Article 4 a) aimed at regulating all AI systems, regardless of their level of risk, adopting a principle-based regulatory approach.[76] At the same time, *ad hoc* obligations for developers of foundation models were introduced in the proposed Article 28 b, which resembles rule-based regulation. This constitutes an example of how the principle of accountability may be overlooked or poorly implemented, leading to a proliferation of compliance obligations, while at the same time undermining the normative force of other regulatory principles. For this reason, the policymaker shall develop comprehensive accountability practices for any entity producing impactful technological products, regardless of perceived risk levels, for a truly future-proof and resilient regulation.[77]

(PR3) Redefining Algorithmic Fairness for High-Impact Automated Decision-Making

Main author: Robert Lee Poe

Addressee:

The policy recommendation is addressed to individuals seeking to implement fair machine learning metrics in pipelines that are responsible for the distribution of finite resources (e.g., hiring, emergency-care resource allocation, diagnosis, etc.).

Context / history of the problem:

What constitutes a just society is a question that has perennially occupied human thought, and the answers to that question have guided human action for millennia. At the core of this inquiry are, generally speaking, two competing notions of justice, offering conflicting perspectives on how to make sense of the boons and burdens that differentiate the lives of individuals in society. These are distributive and non-distributive justice, respectively. Distributive justice is concerned with the equitable allocation of resources among members of a society, asking

[76] Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) available at https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html accessed 25 November 2023.

[77] Irina Carnat, 'Ethics Lost in Translation: Trustworthy AI from governance to regulation' (pre-print 2023) 4 Opinio Juris in Comparatione 30-31.

questions about what should be distributed, to whom it should be distributed, and in what manner.[78] Many specific theories of justice, such as social justice, environmental justice, and health justice, involve considerations of distributive justice because they focus on how boons and burdens should be shared. Non-Distributive justice relates to aspects of justice that do not involve this sort of sharing out of boons and burdens. Instead, non-distributive justice is about the fair treatment of individuals regardless of the outcomes of the distribution, and it includes theories such as procedural justice, which focuses on the fairness of processes, and retributive and corrective justice, which are concerned with the response to both virtuous and unvirtuous behavior.

The conflict between these two concepts of justice can perhaps best be understood through a brief explanation of their most notable, contemporary advocates. In *A Theory of Justice*, John Rawls embeds his argument for distributive justice in a thought experiment known as the "Original Position," which asks decision-makers to operate under a veil that obscures their own (original) position in society, ensuring that the principles they choose would be fair to all. Rawls' two principles of justice—the liberty principle and the difference principle—prioritize basic liberties for all and allow social and economic inequalities only if they benefit the least advantaged members of society.[79] In contrast, Robert Nozick's *Anarchy, State, and Utopia* counters with a non-distributive conception of justice. Nozick emphasizes individual rights and entitlements, arguing that justice is not about the end-state distribution of goods but about the processes that lead to that distribution. He introduces the entitlement theory, which justifies distributions based on principles of just acquisition, transfer, and rectification.[80]

Definition of the problem:

The philosophical tensions between these kinds of conceptions of justice find a modern parallel in the developing field of "fair machine learning." As machine learning algorithms increasingly influence decisions that affect human lives—ranging from employment and loan approvals to medical diagnoses and treatment—scholars and practitioners are struggling with the challenge of integrating established principles of justice into these technologies. These principles extend beyond the ethical theories historically debated by philosophers; they encompass the concrete conceptions of justice that have been crystallized in legal statutes and case law over centuries. The conception of justice embodied in fair machine learning metrics and techniques is based on theories of distributive justice, characterized as egalitarian and equitable. [81]

Nevertheless, the equitable conception of justice that is central to fair machine learning frequently clashes with the norms and laws of historically liberal legal orders. This dichotomy poses a dual challenge, both legal and ethical. A cornerstone of AI ethics is the premise that

[78] Sven Ove Hansson, *Equity, Equality, And Egalitarianism*, 87 ARSP: ARCHIV FÜR RECHTS- UND SOZIALPHILOSOPHIE / ARCHIVES FOR PHILOSOPHY OF LAW AND SOCIAL PHILOSOPHY 529 (2001).

[79] JOHN RAWLS, A THEORY OF JUSTICE: ORIGINAL EDITION (1971), https://www.jstor.org/stable/j.ctvjf9z6v (last visited Nov 2, 2023).

[80] ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA (1974).

[81] Reuben Binns, *Fairness in Machine Learning: Lessons from Political Philosophy*, *in* PROCEEDINGS OF THE 1ST CONFERENCE ON FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY 149 (2018), https://proceedings.mlr.press/v81/binns18a.html (last visited Nov 2, 2023); Robert Lee Poe & Soumia Zohra El Mestari, *The Flawed Foundations of Fair Machine Learning*, (2023), http://arxiv.org/abs/2306.01417 (last visited Sep 1, 2023).

unlawfulness in AI systems inherently undermines their ethical standing.[82] Consequently, automated decision-making systems are obligated to adhere to legal standards—upholding the rule of law—while also accommodating the lawful, normative aims of individuals, businesses, and institutions operating within those boundaries. It is here that our first obstacle in applying algorithmic fairness emerges: adherence to the law.

The hiring example sheds light on how the use of fair machine learning techniques can be unlawful. According to the Court of Justice of the European Union, preferential treatment in hiring is only allowed in tie-breaking scenarios where two candidates are equally qualified, and the comparison of candidatures must be subject to an objective assessment (Marschall Test).[83] However, when a fairness metric is chosen that requires the elimination of group dissimilar outcomes based on a protected attribute while disregarding the base-rate differences between groups, the effect is to give systematic, preferential treatment to the individuals of one group at the expense of the other; and the severity of that systematic deviation from equal treatment (i.e., direct or positive discrimination) is dependent on the strength of the correlation between the sensitive attribute and the target variable in the original, unmodified sample.[84]

If a model is trained on a representative sample where group disparities are present in the target population, the outcomes will, of course, be group dissimilar. This realization leads us to the question about what to do with group dissimilar outcomes, which is the fundamental question of (un)fairness in machine learning. Should the base-rate differences between groups be disregarded through the curation of the sample or modification of the objective function—the playing-field tilted at the moment of competition—resulting in the preferential treatment of some and the disadvantageous treatment of others based on their protected attributes in order to arrive at an equitable distribution of goods (distributive justice); or should the decision stand, ensuring equal treatment and resulting in an impartial comparison in the particular competition—relying on institutions guided by substantive equality of opportunity and the corresponding policies of positive action[85] to achieve factual equality between groups in our societies (non-distributive justice)? Depending on the field of application (hiring, admissions, loan approval, etc.) and jurisdiction, the answer to this normative question may have already been decided.

[82] Luciano Floridi, *Soft Ethics and the Governance of the Digital*, 31 PHILOS. TECHNOL. 1 (2018) for the distinction between soft and hard ethics that was adopted by the High-Level Expert Group on AI and their "Trustworthy AI Guidelines" (p. 12.); Giovanni Comandé, *Unfolding the Legal Component of Trustworthy AI: A Must to Avoid Ethics Washing*, (2020), https://papers.ssrn.com/abstract=3690633 (last visited Feb 21, 2023) for an analysis of the relationship between law and AI ethics.

[83] *See* Case 450/93 *Kalanke v Bremen* [1995] ECR I-3051; Case 409/95 *Marschall v Land Nordrhein-Westfalen* [1997] ECR I-6363; Case 158/97 *Badeck v Hessischer Ministerpresident* [2000] ECR I-1875; Case 476/99 *Lommers v Minister van Landbouw, Natuurbeheer en Visserij* [2002] ECR I-02891; and Case 407/98 *Abrahamsson and Andersson v Fogelqvist* [2000] ECR I-5539.

[84] Robert Lee Poe, *Why Fair Automated Hiring Systems Breach EU Non-Discrimination Law*, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases - Workshop and Tutorial Track (2023), http://arxiv.org/abs/2311.03900 (last visited Nov 9, 2023) for an example of the conflict, specifically between fair automated hiring and EU non-discrimination law.

[85] For an exhaustive description of positive action doctrine in the EU, *see* Van Caeneghem, J.: Legal Aspects of Ethnic Data Collection and Positive Action: The Roma Minority in Europe. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-23668-7, http://link.springer.com/10.1007/978-3-030-23668-7; see also Directive 2006/54/ EC of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation.

The second challenge to algorithmic fairness, as currently defined, is less of an obstacle and more of an impasse. To understand this impasse, the relationship between statistically accurate outcomes and group similar outcomes should be understood.[86] Traditional machine learning tries to understand a description of reality encapsulated in a dataset that maps to the relevant features for a ranking and makes a prediction consistent with that description. It is a *descriptive* and *predictive* process. Fair machine learning enforces a given notion of fairness on the outcome of the decision. It is a *prescriptive* process. Where the objective of traditional machine learning is to understand what "is" so that a model can predict what is likely to be, fair machine learning asserts what "ought" to be instead.

Fair machine learning is an effort to transform societies by placing normative constraints on decision-makers, specifically by hardcoding equity (group similarity in outcome) in decision-making systems, in order to balance power imbalances and reverse historical effects of discrimination.[87] It is in this way that algorithmic fairness, as paradigmatically defined, is ahistorical; the more information a system has about a data setting filled with group disparities, the more group dissimilarities there will be in the outcomes. The *ahistorical* constraint placed on this *data-driven* process results in the tradeoff between statistically accurate and group similar outcomes. The relationship between statistically accurate and group similar outcomes entails that where group disparities are greatest, data-driven processes are the least useful—old-fashioned quotas would have the same effect. The good news is that the ahistorical nature of algorithmic fairness is simply a direct consequence of defining fairness in outcomes (i.e., through distributive justice).

Proposed policy recommendation aimed at solving the problem:

Thus, a critical examination reveals that the application of distributive justice in the domain of machine learning, while well-intentioned, is incompatible with a statistical approach and may result in conflicts with non-discrimination law, where the principle of equal treatment is systematically violated, and data protection law, where the sensitive attributes of individuals (religion, race, gender, etc.,) are needed in order to engage in the kind of positive discrimination required to achieve equitable outcomes. While the CJEU has clearly found such practices unlawful in the context of employment, the Court has found that reserving training positions for individuals based on sensitive attributes to be lawful, as well as making it mandatory for underrepresented groups to be called during the interviewing process. The guiding principle for when *special measures* go too far, becoming positively discriminatory, is the principle of substantive equality of opportunity which is distinguished from equality of outcome.[88] By understanding the difference between those two principles, practitioners can identify where the concept of distributive justice may be applied lawfully and where it may not. Practitioners should be especially careful when there is an "attempt to achieve a final result".[89] Regardless, non-distributive justice might offer a more robust and legally and ethically compliant framework, fostering trust and acceptance among the public. In the machine learning pipeline, non-distributive justice would require robust models trained on representative data samples,

[86] Poe and Mestari (n81).

[87] Alycia N. Carey & Xintao Wu, *The Statistical Fairness Field Guide: Perspectives from Social and Formal Sciences*, 3 AI ETHICS 1 (2023).

[88] Case 158/97 Badeck v Hessischer Ministerpresident [2000] ECR I-1875, § 19.

[89] Id. at §60.

and a feature selection process that satisfies the proportionality test required for the use of features that result in a disparate impact based on a sensitive attribute.[90]

(PR4) Subliminal, manipulative and deceptive techniques in the context of the AI Act: new definitions proposal

Main author: Vittoria Caponecchia

Addressee:

In a world pervaded by artificial intelligence (AI), it is necessary for the law to maintain a predominant position, guaranteeing the protection and preservation of human rights and interests, especially in terms of legal certainty. This is because, while AI undoubtedly brings benefits in any field, it also entails risks for both individuals and society.[91] It is proving increasingly problematic, however, to ensure that the law keeps pace with the development of new technologies, which run much faster and therefore become difficult to regulate. Precisely for this reason, several regulations have been proposed and even adopted at EU level, the most recent of which is the recently adopted Artificial Intelligence Act (AI Act), which fits perfectly within the European digital strategy[92], the aim of which is to create a single European data space (single market for data) while leaving a central position for humans[93].

The AI Act establishes harmonised rules for artificial intelligence, with the aim, among others, of meeting the requirements of a well-functioning internal market[94], ensuring a high level of data protection, digital rights and ethical standards95, and addressing the opacity and complexity of AI systems, as well as a certain degree of unpredictability and partially autonomous behaviour of certain AI systems, to ensure their compatibility with fundamental rights and to facilitate the enforcement of legal rules[96].

Nonetheless, although the specific objectives of the proposal include ensuring legal certainty and improving the effective application of existing legislation, the proposal itself emphasises, in recital 15, how artificial intelligence today "*can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices*". For this reason, article 5(1)(a) of the AI Act proposal needs to be changed in some of its points, in order to avoid uncertainty and misunderstandings, as well as to raise the awareness of the addressees of the proposal, namely the AI service providers and their users (and of those who will have to enforce

[90] Hacker, P.: Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law (Apr 2018), https://papers.ssrn.com/abstract=3164973

[91] "*Given the major impact that AI can have on our society and the need to build trust, it is vital that European AI is grounded in our values and fundamental rights such as human dignity and privacy protection. Furthermore, the impact of AI systems should be considered not only from an individual perspective, but also from the perspective of society as a whole*", White Paper On Artificial Intelligence- A European approach to excellence and trust, COM(2020) 65 final.

[92] https://eufordigital.eu/discover-eu/eu-digital-strategy/; EU Data Strategy (n6); Commission's Communication on "*Shaping Europe's digital future*", 2020.

[93] EU Data Strategy (n6) 4.

[94] AI Act proposal (n14).

[95] European Council, *European Council meeting (19 October 2017)* – Conclusion EUCO 14/17, 2017, p. 8.

[96] Council of the European Union, Presidency conclusions- The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change, 11481/20, 2020, p. 5.

the text of the regulation once it enters into force). Such could be resolved by the EU legislator, to whom this policy recommendation is addressed, as he could amend the text of the proposal by addressing these issues.

Context / history of the problem:

The first part of article 5(1)(a) of the AI Act, as last amended, prohibits "*the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective to or the effect of materially distorting a person's or a group of persons' behaviour by appreciably impairing the person's ability to make an informed decision, thereby causing the person to take a decision that that person would not have otherwise taken in a manner that causes or is likely to cause that person, another person or group of persons significant harm*"[97].

The problem arises from the lack of definitions of "*subliminal techniques*", "*manipulative techniques*" and "*deceptive techniques*", as well as of "*significant harm*". It is necessary to recall that it is very difficult to find a precise definition of such techniques in the legal sphere, since they are phenomena typical of other fields of science, such as psychology, philosophy, neurology and marketing (although some legal texts, e.g. the Unfair Commercial Practices Directive, provides some definitions, albeit referring to the commercial sphere[98]). However, since these techniques also have repercussions on people's rights and, therefore, their use is prohibited, it is good to clarify with certainty what they refer to and, therefore, what is prohibited, in order also to respond to the request of article 5(1)(a), already anticipated by recital 16 of the same proposal. In fact, as mentioned at the beginning, one of the main tasks of law is to guarantee the principle of certainty, according to which the law must have a predictable application. Otherwise, confusion and insecurity arise, making it pratically impossible to understand how to act within the limits of the law.

In order to prevent providers from developing, deploying or commercializing AI systems that may breach the obligations of the proposed AI Act, it is necessary to specify the meaning of the above-mentioned expressions (*subliminal*, *manipulative* and *deceptive techniques*). For the sake of cohesion and brevity, this recommendation will omit, however, an exploration of the meaning of "*significant harm*", which would require an in-depth discussion in its own right.

This policy recommendation was written after examining the most recent regulations that are applicable within the scope, and for the purposes, of the European digital strategy (i.e., Digital Services Act - DSA[99]; Digital Markets Act - DMA[100]; Data Act[101]). In addition to these, the most important consumer protection legislation was studied (Unfair Commercial Practices

[97] Amendments to the AI Act (n76).

[98] Article 5(b) of the Directive states that a practice is unfair if "*it materially distorts or is likely to materially distort the economic behaviour with regard to the product of the average consumer whom it reaches or to whom it is addressed, or of the average member of the group when a commercial practice is directed to a particular group of consumers*". That provision adds, moreover, that misleading (articles6 and 7, which will be commented on later) and aggressive commercial practices are considered unfair. Among the latest Italian case law on the subject, see Council of State, Sec. VI, Sent. no. 4498/2023; Council of State, Sec. VI, Sent. no. 203/2022; Council of State, Sec. VI, Sent. no. 2414/2020.

[99] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 october 2022 on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act).

[100] DMA (n40).

[101] Data Act (n11).

Directive - UCPD[102]; and, at Italian level, Legislative Decree No. 145/2007[103] and Legislative Decree No. 146/2007[104]), insofar as the aforementioned techniques can be classified as unfair commercial practices and therefore subject to the relevant discipline.

It was observed that none of these regulations contain express references to the notions of subliminal, manipulative and deceptive techniques, but how they may contain references in general to subliminality, manipulation and deception, terms that are united by the fact that they fall within (or, as the case may be, contain the) category of so-called dark patterns. The latter were coined in 2010 by Harry Brignull, U.S. researcher and user experience designer, who defined them as "*a user interface that has been carefully crafted to trick users into doing things, such as buying insurance with their purchase or signing up for recurring bill*"[105]. In order to find an unambiguous meaning of the expressions mentioned in article 5(1)(a) of the AI Act or, in any case, to try to better understand what they refer to, let us proceed to examine the above-mentioned regulations.

Definition of the problem:

Starting with the notion of "*subliminal technique*", we can see that none of the above-mentioned regulations contain such an expression. Since the BRIEF project concerns the Euro-Italian area, Italian legislation was also analysed. At a national level, the Italian Legislative Decree No. 145/2007, concerning misleading advertising, affirm, in article 5, the need for transparency in advertising and expressly prohibits subliminal advertising.
The same decree, in article 1, states that "*advertising must be clear, truthful and correct*"[106], while article 2 defines misleading advertising as "*any advertising which in any way, including its presentation, is likely to mislead the natural or legal persons to whom it is addressed or whom it reaches and which, by reason of its misleading character, is likely to prejudice their economic behaviour, or which, for that reason, is likely to harm a competitor*".
At this point, two questions spontaneously arise concerning the interpretation of the term "*subliminal"*:

- Does it refer to advertisement that is not "*clear, truthful and correct"*[107] (since, if an advertisement must be transparent in order not to be considered subliminal, then it must also be clear)?;

[102] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council (Unfair Commercial Practices Directive).

[103] Legislative Decree No. 145 of 2 August 2007 "Implementation of Article 14 of Directive 2005/29/EC amending Directive 84/450/EEC concerning misleading advertising", published in the Official Gazette No. 207 of 6 September 2007;

[104] Legislative Decree No. 146 of 2 August 2007 "Implementation of Directive 2005/29/EC concerning unfair business-to-consumer commercial practices in the internal market and amending Directives 84/450/EEC, 97/7/EC, 98/27/EC, 2002/65/EC, and Regulation (EC) No. 2006/2004", published in the Official Gazette No. 207 of 6 September 2007.

[105] Harry Brignull, *What are dark patterns?*, 2010, https://www.deceptive.design/types; Harry Brignull, *Deceptive patterns. Exposing the tricks tech companies use to control you*, Testimonium Ldt, 2023, p. 5.

[106] Personal translation of art. 1, Legislative Decree 145/2007, which states: "*La pubblicità deve essere palese, veritiera e corretta*".

[107] *Ibid*.

- Assuming that "*transparent*" is equivalent to "*clear*"[108], is an advertisement that is not transparent then misleading? If so, does "*subliminal"* then fall under the latter definition?

It should be noted, however, that these definitions are contained in a decree pertaining exclusively to advertising, so all areas in which AI deploys negative effects that do not concern advertising, such as, but not limited to, virtual assistants[109] (the design of whose interfaces is often designed in such a way as to push users to make unwanted choices, e.g. buying or engaging more, hijacking their decision-making capability[110]) and language models with strategic reasoning (e.g. CICERO by Meta[111], strategy game based on negotiation and persuasion of opponents[112]) would be left out. Moreover, this decree implement the Unfair Commercial Practices Directive at internal level, therefore only at Italian one. This implies that other EU member States may have regulated the matter differently, using other expressions or dictating other definitions, which contributes to legal uncertainty.

As far as "*manipulative techniques*" are concerned, this term is found in both the DSA and the Data Act, but with different nuances.

In the DSA, the most relevant references to manipulation are to be found in the following recitals, which do not provide a precise definition of the term in question, but allow us to understand what is meant:

- Recital 21, suggests that manipulation can be a technique that "*alter the integrity of the information transmitted or to which access is provided*";
- Recital 69, implies that manipulation can be a technique that "*can negatively impact entire groups and amplify societal harms, for example by contributing to disinformation campaigns or by discriminating against certain groups"*;
- Recital 83, which, pointing to the fourth category of systemic risks that undermine online security through certain "*design, functioning or use of very large online platforms and of very large online search engines"*, mentions manipulation as a means by which these risks could materialise. It further specifies that these could have "*actual or foreseeable negative effect on the protection of public health, minors and serious negative consequences to a person's physical and mental well-being, or on gender-based violence"* (the text of article 5(1)(a) prior to the June 2023 amendments mentioned more narrowly "*physical or psychological harm"*). Finally, it adds that "*such risks may also stem from coordinated disinformation campaigns [...] or from online interface design that may stimulate behavioural addictions of recipients of the service"* (probably referring to dark patterns, which we will discuss below);
- Recital 84, which, on the subject of systemic risk assessment of online platforms, also calls for an assessment of manipulation, which can occur, for example, through the misleading use of the service itself.

[108] *Ibid*.

[109] Silvia De Conca, *The present looks nothing like The Jetsons: deceptive design in virtual assistants and the protection of the rights of users*, https://www.sciencedirect.com/science/article/pii/S0267364923000766?ssrnid=4412646&dgcid=SSRN_redirect_SD.

[110] *Ivi*, p. 1.

[111] https://ai.meta.com/research/cicero/ ; https://ai.meta.com/research/cicero/diplomacy/

[112] Meta Fundamental AI Research Diplomacy Team (FAIR) *et al., Human-level play in the game of Diplomacy by combining language models with strategic reasoning, Science* 378,1067-1074(2022), DOI:10.1126/science.ade9097.

The main reference to this issue made by the Data Act, on the other hand, is contained in recital 34, which prohibits the third party from using coercive, deceptive "*or*" manipulative means (thus, implicitly differentiating them from each other, but without specifying why they differ) against the user, subverting or impairing the user's autonomy, decision-making or choices, including through a digital interface. With reference to the latter, the recital 34 states that, in this context, third parties should not even refer to dark patterns in their design, describing them as "*design techniques that push or deceive consumers into decisions that have negative consequences for them*". They can be used, indeed, as this recital also states, "*to persuade users, particularly vulnerable consumers, to engage in unwanted behaviours, and to deceive users by nudging them into decisions on data disclosure transactions or to unreasonably bias the decisionmaking of the users of the service, in a way that subverts and impairs their autonomy, decision-making and choice"*.

According to this recital, dark patterns do not correspond exactly to "*coercive, deceptive or manipulative means"*, but they are a subcategory of them and, in particular, of deceptive means. Moreover, the term "*persuasion*", used in this context, suggests that deception can be associated with persuasion itself. Nevertheless, the concepts of persuasion and manipulation could also be associated ("*these manipulative techniques can be used to persuade users*"), because the former can be seen as a subcategory of the second (some understand persuasion as the impulse that rationally convinces people to do something, thus never pushing them to do what they do not want to do – unwanted behaviour[113]). So, is deception also a subcategory of manipulation? And in what terms? And what does manipulation consist of?

With regard to dark patterns, specifically, recital 67 of the DSA also evokes them, defining them as "*practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions. Those practices can be used to persuade the recipients of the service to engage in unwanted behaviours or into undesired decisions which have negative consequences for them. Providers of online platforms should therefore be prohibited from deceiving or nudging recipients of the service and from distorting or impairing the autonomy, decision-making, or choice of the recipients of the service via the structure, design or functionalities of an online interface or a part thereof [...] presenting choices in a non-neutral manner*".

Similarly to the Data Act, the DSA mentions deception rather than manipulation and it additionally refers to "*non-neutrality*", which could be linked to the expression "*subliminal technique*". Indeed, non-neutrality consists of a partial or biased attitude, which can be held through subliminal techniques, in order to steer recipients in a certain direction, without explicitly stating a position. At the same time, the use of subliminal techniques may serve precisely to achieve a purpose, in a more subtle way.
And, in connection with what has been said above, in the analysis of the Italian Legislative Decree No. 145/2007, if subliminal technique were to be equated with a lack of transparency, the fact that the concepts of non-neutrality and subliminality can coexist would also include the concept of non-transparency: the subliminal (or non-transparent) technique can be the means by which non-neutrality is exercised or the very result of the experiment of a non-neutral action, thus the lack of transparency allows (or leads) to a non-neutral result.

[113] Daniel Susser, Beate Roessler, Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, Georgetown Law Technology Review, 2018.

Of course, these conclusions are hypothetical, since it is impossible to know the intention of the legislator with absolute certainty, such as why they distinguished these expressions that are often used interchangeably in everyday life.

Coming finally to the analysis of the term "*deceptive technique*", it could be argued that it is the least problematic since, as we have seen, it is much more widespread in the regulatory texts mentioned so far. However, there is no definition of this term, which we find in the form of "*misleading commercial practice*" in the Unfair Commercial Practices Directive (later incorporated by Italian Legislative Decree No. 146/2007).

In this context, it is assumed that the terms "*misleading*" and "*deceptive*" can be considered synonymous, since articles 6 and 7 expressly contain the statement "*a commercial practice shall be regarded as misleading if it contains false information and is therefore untruthful or in any way, including overall presentation, deceives or is likely to deceive the average consumer*". In particular, the Directive defines "*misleading commercial practices*" as:

- Art. 6(1): "*A commercial practice shall be regarded as misleading if it contains false information and is therefore untruthful or in any way, including overall presentation, deceives or is likely to deceive the average consumer, even if the information is factually correct [...] and in either case causes or is likely to cause him to take a transactional decision that he would not have taken otherwise*";
- Art. 6(2): "*A commercial practice shall also be regarded as misleading if, in its factual context, taking account of all its features and circumstances, it causes or is likely to cause the average consumer to take a transactional decision that he would not have taken otherwise [...]*";
- Art. 7(1): "*A commercial practice shall be regarded as misleading if, in its factual context, taking account of all its features and circumstances and the limitations of the communication medium, it omits material information that the average consumer needs, according to the context, to take an informed transactional decision and thereby causes or is likely to cause the average consumer to take a transactional decision that he would not have taken otherwise*";
- Art 7(2): "*It shall also be regarded as a misleading omission when [taking account of the matters described in paragraph 1], a trader hides or provides in an unclear, unintelligible, ambiguous or untimely manner such material information [as referred to in that paragraph] or fails to identify the commercial intent of the commercial practice if not already apparent from the context, and where, in either case, this causes or is likely to cause the average consumer to take a transactional decision that he would not have taken otherwise*".

It should be recalled, however, that the Directive covers "*commercial practices directly related to influencing consumers' transactional decisions in relation to products. It does not address commercial practices carried out primarily for other purposes*" (recital 7). This means that everything outside the commercial scope and unrelated to a product is excluded from such discipline.
Article 5 of the AI Act, on the other hand, concerns AI systems in general, so they could have negative implications both in commercial terms[114] and non-commercial terms (e.g. they could aim at obtaining consent and personal data, just think of online phishing).

[114] EU regulations on digital services and digital market also refer to misleading practices in commercial terms by prohibiting them (e.g. Recital 35 DMA aims at "*fight fraudulent and deceptive commercial practices*"), therefore recognising the existence of deceptive practices that have negative effects in commercial terms.

What is evident is the link between deceptive techniques, as defined in the commercial sphere, and dark patterns, which Harry Brignull actually prefers to call "*deceptive patterns"*, as he wrote in his recently published book[115].

Proposed policy recommendation aimed at solving the problem:

In the light of what has been examined so far, it is proposed, first of all, to remove the term "*subliminal technique"* from the text of the AI Act, since subliminality is a stimulus that is too weak to be perceived and recognised, but not so weak that it does not influence a person's behaviour or psyche[116]. Thus, it is very difficult, if not impossible, to detect it and often not even the perpetrator is aware of it. If this reference is not removed from the proposal, there is a risk of pursuing something unknown. The result would be either the uncertainty of classifying a certain behaviour as subliminal or not and, therefore, not knowing whether to sanction it or not, risking not punishing unlawful behaviour or, on the contrary, the sanctioning of behaviour that is not unlawful.

Irrespective of whether there is the willingness of the AI service provider to cause harm to one or more persons, it is hereby recommend to focus on "*deceptive techniques*" and define them as "*any active or passive behaviour - action or omission - that leads a person to make choices that he or she would not otherwise have made, because of incorrect, false, misleading or incomplete information or, conversely, the lack of information relevant to make an informed decision. The relevance of that information must be assessable ex post, making it possible to understand whether it could have enabled the subject to make a different choice, more favourable to her. This evaluation must be carried out considering the typical diligence of the average person, normally informed and reasonably observant and circumspect, taking into account social, cultural and linguistic factors, as interpreted by the Court of Justice*

The category of deceptive techniques also includes dark patterns, design techniques that deceive consumers into making decisions that have negative consequences for them".
Secondly, a distinction has to be made between the notions of "*deceptive technique*" and "*manipulative technique*", with the latter being defined as *"the concrete behaviour that alters the quality and integrity of the information or design and development processes of the AI system, in order to cause significant harm* [an expression also, as I mentioned at the outset, to be clarified and specified by the legislator] *to one or more persons*"[117].

Finally, in order to better guide the recipient of the proposal and to help the interpreter in the application of the text of the law, it is proposed that the above definitions of "*deceptive techniques*" and "*manipulative techniques*" be introduced in article 3 AI Act.

Constraints of the policy recommendation:

This policy recommendation has been formulated taking the above-mentioned legislation as a reference, as this is the most recent legislation applicable in the context of the European digital

[115] Harry Brignull, *Deceptive patterns. Exposing the tricks tech companies use to control you*, Testimonium Ldt, 2023, p. 241.

[116] Il nuovo Zingarelli minore, vocabolario della lingua italiana, Zanichelli, Milano, 2008, p. 1220.

[117] Personal formulation of "*manipulative technique*", reconstructed following the definitions currently found in the various legislative texts. In particular, reference is made to what is already contained in recital 21 DSA and art. 5 AI Act.

strategy. There may therefore be other sources, both normative and doctrinal, to support alternative or conflicting solutions to the one outlined in this recommendation. However, the latter could make a real change in terms of certainty.

Its main objective is to clarify and make the recipients of the AI Act (AI system providers and their users, as well as the interpreter) aware of the terminology and, consequently, the existence of certain phenomena (such as dark patterns), with the hope that, in this way, AI system providers will be able to recognise the "*limits of the lawful*" within which they must act, that those who feel they have exceeded them and claim to have suffered harm will be able to defend themselves, and that judges will have better defined parameters to ensure a consistent and safe application of the law.

To know more:

- Juan Pablo Bermúdez, Rune Nyrup, Sebastian Deterding, Laura Moradbakhti, Céline Mougenot, Fangzhou You, Rafael A. Calvo, What Is a Subliminal Technique? An Ethical Perspective on AI-Driven Influence?, IEEE Ethics-2023 Conference Proceedings (2023);
- Mark Leiser, lluminating Manipulative Design: From "Dark Patterns" to Information Asymmetry and the Repression of Free Choice Under the Unfair Commercial Practices Directive, Loyola Consumer Law Review, Volume 34, Issue 3 Symposium Issue 2022;
- Mark Leiser, Psychological Patterns and Article 5 of the AI Act Proposal. AI-Powered Deceptive Design in the System Architecture & the User Interface, available at SSRN: https://ssrn.com/abstract=4631535;
- Peter S. Park, Simon Goldstein, Aidan O'Gara, Michael Chen, Dan Hendrycks, AI Deception: A Survey of Examples, Risks, and Potential Solutions, https://arxiv.org/abs/2308.14752.

(BP2) Guidelines for researchers to ensure the transparency of AI systems used in bio-robotics context

Main author: Stefano Tramacere

Addressees:

The addressees of these best practices are mainly bioengineering researchers working in a public research center studying and testing new AI systems in the medical field, collecting health data, and training such automated systems on these datasets. An accountability framework is needed so that doctors and healthcare facilities have less liability if the AI tool, tested by researchers, causes harm to the end user, *i.e.*, the patient.

Context/history of the problem:

The use of AI systems in the healthcare sector raises significant ethical, societal, and legal concerns regarding the protection of fundamental rights[118]. One of the main problems is the

[118] For an in-depth examination read the Study of the Panel for the Future of Science and Technology (STOA), European Parliamentary Research Service, *Artificial Intelligence in healthcare – Applications, risks, and ethical and societal impacts*, June 2022; and J. Van De Hoven et al., *Toward a Digital Ecosystem of Trust: Ethical, Legal and Societal Implications,* in Opinio Juris in Comparatione, 2021, p. 131.

opacity of most state-of-the- art AI systems, *i.e., black box models*[119]. These models might have millions of parameters that capture the extreme non-linearities of the input features, making their internal decision-making process hard to understand and interpret by humans[120]. Hence, the opacity of these models makes it difficult to examine their reliability, to detect and prevent potential malfunctions and ensure a high level of protection to individuals[121]. From a technical point of view, some solutions to provide greater transparency are eXplainability techniques (*XAI*)[122]. One approach involves incorporating explicit explainability features into the design of AI models (*ex-ante*) to develop transparent-by-design or explainable-by-design models. A different approach focuses on creating tools and methods that generate *post-hoc* explanations from an output after the decision has been made[123], such as feature importance scores or counterfactuals[124].

Definition of the problem:

The lack of transparency of AI systems can generate several risks: (1) the *automation bias* which refers to the phenomenon where individuals place blind trust in the outcomes generated by automation, even when they possess knowledge or awareness that the automation may be fallible; (2) the *translational bias* which concerns the adverse consequences (*e.g.,* inaccurate prediction) of using an AI system that has been trained on certain categories of data in a specific context and then subsequently employed in an only apparently similar one[125]. Due to the opacity of the models used, these phenomena can lead to two opposing physicians' reactions: either *overreliance* or *distrust* in AI systems[126]. For example, doctors can make crucial decisions for the life of patients using medical AI applications that provide highly accurate diagnoses, without knowing that the decision was generated by an AI system and without having a clear and complete understanding of the logic behind them. In fact, the lack of transparency could

[119] G. Comandé, *Intelligenza artificiale e responsabilità tra liability e accountability – Il carattere trasformativo dell'IA e il problema della responsabilità*, in Analisi Giuridica dell'Economia (a cura di A. Nuzzo, G. Olivieri), il Mulino, n.1/2019, pp. 169-188.

[120] R. Guidotti, F. Giannotti, D. Pedreschi, *Explainability (30),* in Edgar Encyclopedia of Law and Data Science, edited by G. Comandé, 2022, pp. 160-168.

[121] B. Béviére-Boyer, *The French paradox of the Halftone Legislative Intervention on Artificial Intelligence in Health by the Bioethics Law of August 2, 2021,* in Artificial Intelligence Law – Between Sectoral Rules and Comprehensive Regime Comparative Law, edited by C. Castets-Renard and J. Eynard, Bruylant, 2023, pp. 277-282; and G. Maliha, et al., *Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation,* in The Milbank Quarterly, n.3/2021, pp. 629-647.

[122] R. Guidotti, et al., *A Survey Methods for Explaining Black Box Models,* in ACM Computing Surveys, n.5/2018.

[123] Regarding this central distinction, read B. Gyevnar, et al., *Bridging the Transparency Gap: What Can Explainable AI Learn from the AI Act?*, in Proceeding of ECAI 2023, the 26th European Conference on Artificial Intelligence, p.966, where the Authors write: *"Ante-hoc explanations are generated directly from the internal representations and processes of white box systems, while post-hoc explanations are inferred from an output after the decision was made. Thus, ante-hoc explanations are truthful to the decision process by design. Post-hoc explanation may distort the causality underlying the model's decision process and require more effort to generate but apply to both white and black box systems."*

[124] S. Cussat-Blanc, *Which artificial intelligence for augmented medicine*?, in Artificial Intelligence Law – Between Sectoral Rules and Comprehensive Regime Comparative Law, edited by C. Castets-Renard and J. Eynard, Bruylant, 2023, pp. 234-252; and R. Guidotti, F. Giannotti, D. Pedreschi, *Explainability (30)* (n120).

[125] On these profiles and their relation to civil liability, see G. Comandé*, Intelligenza artificiale e responsabilità tra liability e accountability* (n119) 176.

[126] On this topic, we recommend reading the interesting study conducted by C. Panigutti, et al., *Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems*, in CHI Conference on Human factors in Computing Systems, 2022.

hide incorrect inferences[127] and algorithmic discriminations[128] that could endanger the health and safety of patients[129], in violation of their fundamental rights.[130] Therefore, it is necessary to devise a transparent risk management system, that entails knowing when one is interacting with an AI system and understanding how opaque AI systems are trained, which datasets they use, how they process data and for which specific purposes[131]. From a legal perspective, opacity could interfere with the attribution of civil liability in case the AI system's output cause harm to the patient because it is more difficult to prove the causal link[132]. Thus, the use of *black box* medical AI systems could undermine the liability of healthcare professionals by leaving injured patients unprotected[133]. In this respect, it is necessary for both those who have trained and those

[127] A well-known case of erroneous inference is found in G. Comandé, *Intelligenza artificiale e responsabilità tra liability e accountability* (n119) 182, resuming R. Caruana, et al., *Intellegible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission,* in Proceeding of the 21st ACM SIGKDD, 2015, pp. 1721-1730, which presents an algorithm designed to predict the probability of death among hospital patients with pneumonia systematically classified asthmatic patients at low risk due to a spurious correlation: patients with asthmatic pneumonia were sent directly to the intensive care unit where they received continuous treatment which improved their prognosis so substantially that they appeared to have a better than average chance of survival.

[128] For example, if AI system to check for skin cancer is trained on data from only white people of Caucasian origin, and then subsequently used and tested on dark-skinned people of sub-Saharan origin, the AI system will not be accurate in its prediction and will consequently discriminate against the population not represented in the training data set. On the topic, read C.Y. Johnson, *Racial Bias in a medical algorithm favors white patients over sicker black patients*, in The Washington Post, 2019.

[129] The risk of *"blind"* medical practice if the algorithmic processing cannot be explained is presented by B. Béviére-Boyer, in *The French paradox of the Halftone Legislative Intervention* (n121) 278-280.

[130] See C. D'Elia, *Gli strumenti di intelligenza artificiale generativa nel contesto sanitario: problemi di ottimizzazione delle risorse e questioni di spiegabilità,* in Rivista Italiana di Medicina Legale, n.2/2023, pp. 357-360. Moreover, on the role of digital vulnerability in healthcare read D. Amram, *La transizione digitale delle vulnerabilità e il sistema delle responsabilità,* in Rivista Italiana di Medicina Legale, n.1/2023, pp.1-20.

[131] For a complete analysis of the importance of transparency requirement to have an effective principle of explainability of the internal functioning of algorithms, read B. Béviére-Boyer, *The French paradox of the Halftone Legislative Intervention* (n121) p. 280, where the Author notes *"the importance for health professionals to implement the transparency and explainability requirement for the benefit of the consolidation of the medical relationship, by distinguishing between informed and uniformed audiences (AI specialists, doctors, patients, etc.). The challenge was always to be able to explain to the interlocutor how the algorithmic system works, to justify the opportunity to use it, but also its potential limits which presupposes appropriate training for health professionals, as well as effective means of interaction making exchange and collaborations with the designers and providers of the devices possible"*.

[132] For a comprehensive discussion on civil liability in healthcare in Italy, read G. Comandé, *Medical Law in Italy (Second Edition),* Wolters Kluwer, 2020, pp. 155-173 where the Author write "*The basis of civil liability is (1) fault, (2), causation, and (3) damages. In particular, the trial judge must first identify separately the existence of a causal link between the unlawful conduct and the event of damage and then determine whether that conduct was negligent or willful. Only after finding a causative link must the existence of negligent and the consequent burden of proof be addressed. Note that the causal link between the failure to act on the part of the physician and the injury suffered by the patient should be configured through a necessarily probabilistic criterion […]. Moreover, in those cases where a discussion arose as to whether the harm could be sourced in the alleged medical malpractice, courts have requested that the patient (in line with the general principles on the burden of proof contained in Article 2697) shows the causal link between malpractice and the suffered harm."*

[133] Indeed, when AI is interposed between the act or omission of a person and the damage, the specific characteristics of certain AI systems, *e.g.,* opacity, may make it excessively difficult, if not impossible, for the injured person to meet this burden of proof. The opacity may make it difficult or prohibitively expensive for victims to identify the liable person and prove the requirements for a successful liability claim. It is precisely for these reasons that the proposed *AI Liability Directive (COM/2022/496final)* (n15) lays down common rules on (Article 3) disclosure of evidence concerning high-risk AI systems suspected of having caused damage and (Article 4) on the

who (subsequently) use AI systems to comply with legal rules on transparency, so that the provider or user (*e.g.,* the doctor) of an AI system is more aware of how the system works[134], thus reducing the risk of harming the end user (*e.g.,* the patient) and being held civilly liable.[135] For these reasons, the AI Act proposal[136] (which is under discussion between the European co-legislators at the moment of writing) intends to establish harmonized rules on AI, identifies among high-risk AI systems those that affect health (in its various aspects: diagnosis; treatment; therapy; medical assistance, including emergency; patient triage; etc.) and lays down rigorous legal requirements (Title III, Chapter 2 of the AI Act Proposal)[137].

Proposed best practices aimed at solving the problem:

The aim of these best practices is to regulate the use of AI systems in the performance of tasks in areas that have an impact on the fundamental rights and freedoms of the individual, such as the medical domain. Indeed, the main legal problem concerns the liability regime arising from the use of AI-based medical systems. In this regard, a possible solution may come from the use of the GDPR[138], which develops a risk-based approach to ensure an effective and accountable system. To this end, fundamental to the protection of personal data are the principles of *"privacy by design"* and *"by default"* (art. 25), which are effective expressions to summarize the grafting of rule onto technique and are themselves a concretization of accountability[139]. Such principles draw attention to the proactive attitude and the risk assessment approach aimed at starting personal data flows (by design) so that they can take place (by default) through those technical-organizational measures that guarantee compliance with the regulations in force[140].

This implies, for example, that if a processing presents a high risk to the rights and freedoms of natural persons, controllers must provide an impact assessment, so-called "DPIA" (art. 35), and keep records of the processing activities performed (art. 30). In addition, the GDPR guarantees

burden of proof (alleviated towards the injured person) in tort actions based on fault. In the latter respect, the presumption applies to damage produced by AI systems, provided that the injured party proves: (a) the defendant's negligent breach of duties of care established by European or national law aimed at preventing the damage from occurring; (b) the reasonable likelihood, inferred from the concrete circumstances, that such conduct affected the output of the system; (c) the origin of the damage from the output of the device. Hence, regarding the preparatory studies that led to the new AI Liability directive, read: European Commission, *Liability for Artificial Intelligence and other emerging technologies,* Report from the Expert Group on Liability and New Technologies, 2019.

[134] See R. Hamon, et al., *Bridging the Gap Between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making,* in IEEE Computational Intelligence Magazine, 2022, pp. 72-85.

[135] See A.G. Grasso, *Diagnosi algoritmica errata*, in Rivista di Diritto Civile, n.2/2023, pp. 335-360.

[136] AI Act proposal (n14).

[137] These best practices discuss the AI Act as proposed by the EU Commission. The proposal is currently being debated by the EU co-legislators (the EU Parliament and the EU Council) and therefore the content of the final legislation may differ from what is described here. References to the articles in the following parts have been included to indicate what the legal basis should be once the text is approved, so these references are not binding at this time.
Here, the common position (so called *General approach*) by EU Council, finalized on 28 November 2022: https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf. Moreover, the Parliament adopted its negotiating positions on 14 June 2023 with substantial amendments to the Commission's proposal (no 76).

[138] GDPR (n7*).*

[139] This reflection in D. Poletti, *Comprendere il Reg. UE 2016/679: un'introduzione*, in *Regolare la tecnologia: il Reg. UE 2016/679 e la protezione dei dati personali. Un dialogo tra Italia e Spagna* (a cura di A. Mantelero, D. Poletti), 2018, p.15.

[140] In this sense, read D. Amram et al., *La violazione della privacy in sanità tra diritto civile e penale*, in *Itinerari di medicina legale e delle responsabilità in campo sanitario* (a cura di M. Caputo, A. Oliva), 2021, p. 567.

several technical measures to ensure transparency in the processing of personal data (art. 5); appropriate measures for the processing of special categories of data, such as health data (art. 9); and specific rights for the data subject if there is automated decision-making system (art. 22).

Therefore, based on this Regulation and the interpretation offered by the Italian Data Protection Authority in a Decalogue of September 2023, transparency requirements are embodied in three key principles of the GDPR related to AI systems, which are also shared by the AI proposal[141]:

1. The principle of *knowability*[142], according to which the individual has the right to know about the existence of decision-making processes that concern them based on automated processing (i.e., the concept of "algorithmic legibility" in artt. 13, 14 and 15 GDPR)[143] and to receive meaningful information about the logic involved, so to have means/possibility to understand them (i.e., the principle of *comprehensibility*)[144] (art. 22, rec. 71 GDPR and art. 11 Annex IV (2)(b) AI Act proposal).

2. The principle of *non-exclusivity* of the algorithmic decision, according to which be the decision-making process should include a human intervention that is capable of controlling, validating, or refuting the automated decision, the so-called *human in the loop* (art. 22, rec. 71 GDPR and art. 13 and 14 AI Act proposal). This principle is necessary for *comprehensibility*, since to be able to control the decision-making process, it is necessary to understand the decision and the process that led to it.

3. The principle of *algorithmic non-discrimination*, according to which reliable AI systems should be used, namely systems that reduce opacities and errors caused by technological and/or human causes; their effectiveness should be periodically verified also in the light of the rapid evolution of technologies, by applying appropriate mathematical or statistical procedures for profiling, and by implementing appropriate technical and organizational measures to this end (rec. 71 GDPR, art. 15[145] AI Act proposal among other articles in the Chapter 2[146]).

In practical terms, there are several measures that must be implemented when setting up AI systems in healthcare to limit opacity. These include, as mentioned above, the obligations to

[141] For a detailed presentation, we refer to Autorità Garante per la protezione dei dati personali, *Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale*, settembre 2023.

[142] The principle of "knowability" - of the existence of automated decision-making processes and the logics used - is established in the judgments of the *Consiglio di Stato* (nos. 8472, 8473, 8474/2019; no. 881/2020; no. 1206/2021) and taken up in the Decalogue by the Italian Data Protection Authority in point 4 (n141).

[143] Regarding the important concept of "algorithmic legibility", read G. Malgieri, G. Comandé, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation,* in International Data Privacy Law, n.4/201, pp. 243-265.

[144] The principle of algorithm "comprehensibility" is established in the judgments of the *Consiglio di Stato* (nos. 8472, 8473, 8474/2019; no. 881/2020; no. 1206/2021) which state that any decision-making algorithm used by public administrations to make a decision must be able to provide a humanly comprehensible justification for the decision. These arguments are taken up in the Decalogue by the Italian Data Protection Authority cited. For more discussion on the subject read A. Simoncini, *Amministrazione digitale algoritmica. Il Quadro Costituzionale,* in Il Diritto dell'Amministrazione Pubblica Digitale (a cura di R. Cavallo Perin e D. Galetta), 2020, pp. 1-38.

[145] Article 15 is entitled *"Accuracy, robustness and cybersecurity"*.

[146] Article 9 *"Risk management systems"*; Article 10 *"Data and data governance"*; Article 11 *"Technical documentation"*; Article 12 *"Record-keeping"*; Article 13 *"Transparency and provision of information to users"*; Article 14 *"Human oversight"*.

inform users in compliance with art. 13, 14 and 15 of the GDPR in clear, concise, and comprehensible terms. In the context of AI applications, we propose to interpret the transparency obligations concerning the logics involved as follows:

I. whether the data processing is carried out in the learning phase of the algorithm (i.e., in the phases of experimentation and validation) or in the subsequent phase of its application, in the context of health services, the provider should represent the general logic and characteristics of data processing, especially with *black box* systems. Hence, the provider should indicate the metrics used to train the model and assess the quality of the adopted analysis model, the checks carried out to detect the presence of any biases, any corrective measures adopted, the measures suitable for verifying the performed operation, even *a posteriori*[147], etc.

II. the obligations and liability of the users of the medical AI system;

III. the advantages, in diagnostic and therapeutic terms, deriving from the use of these new technologies; and the risks deriving from such use.

Furthermore, in order to ensure the transparency and explainability of AI systems, it is fundamental to have high quality dataset, so that accurate predictions can be derived from the processing of such data, and to assign a central control role to humans[148], without delegating exclusively to AI systems the decision-making process (art. 14, rec. 48 AI Act proposal).

Constraints of the best practice:

The suggested best practices mainly considered the provisions of the GDPR concerning the processing of personal data. Therefore, they may evolve, change, and adapt to the new regulatory framework once the AI Act completes its legislative process and is finally adopted in EU.

1.7 Regulation of medical devices and health law

(PR5) Uncertainty and Slowdown in the MDR Regulatory Process and the lack of Notified Bodies

Main author: Georgios Christou

Addressee:

For local and national medical regulatory authorities, the European Commission and the European Council.

Medical Device Regulation in Context:

[147] Autorità Garante per la protezione dei dati personali, *Decalogo,* cited, points 7 and 8.

[148] Ivi, point 9.

The regulation of Medical Devices was initially regulated by three directives, the Medical Devices Directive (MDD) 93/42/EEC,[149] Active Implantable Medical Devices Directive (AIMDD) 90/385/EEC, and the In Vitro Diagnostic Medical Devices Directive 98/79/EC[150]. After a scandal in the 2000s involving Poly Implant Prothese (PIP) Breast Implants, which resulted in severe injuries and deaths due to the manufacturer using industrial grade silicone to make breast implants, it was becoming increasingly concerning that the MDD and its sister directive for In Vitro Diagnostics, were becoming outdated. While what happened constituted a violation of the regulations at the time, the medical device safety framework lacked sufficient checkpoints to prevent it from happening. For example, the UK's Medicines and Healthcare Regulatory Authority had completely failed to safeguard women who had received these implants despite first receiving a report of potential problems with PIP implants nearly a decade before the scandal had broken out, including a case of premature rupture of both implants in the same patient[151]. In light of this scandal, the EU introduced the IVDR[152] as well as the EU Medical Device Regulation (EU MDR)[153] to try and prevent such a tragedy from happening again[154]. That case as well as others, such as Johnson & Johnson recalling toxic on-metal hip system, were the cited reasons for the new regulations introduced in[155].

Implementation Issues:

The new MDR is not without its growing pains. When the MDR was introduced, it foresaw that on 27 May 2024 all certificates issued under the former two directives would expire, requiring all devices on the market with such certificates to have an entirely new certification under the MDR. But as of July 2022, MedTech had reported that the vast majority of medical devices on the market had yet to obtain certification under the MDR, despite having less than two years remaining until the deadline of 26th of May 2024[156]. This included certificates that have not been issued yet for "more than 85% of the > 500,000 devices estimated to be covered by (AI)MDD certificates"[157]. Some scholars estimated that a full transition "will probably take even longer than this to complete, and devices certified under the former directives will continue to be used during this time and perhaps for decades if they are put into service or made available on the market on 26 May 2025 at the latest[158]", which is why there has been a reluctance in assessing the impact of the MDR currently. The lack of Notified Bodies (who are the qualified organisations that carry out the assessment procedures and issue certificates under the MDR) remains incredibly difficult, with the EU being very behind on schedule for their set

[149] Council Directive 93/42/EEC of 14 June 1993 concerning medical devices, OJ L 169, 12.7.1993, p. 1–43

[150] Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices OJ L 331, 7.12.1998, p. 1–37.

[151] Victoria Martindale, Andre Menache, 'The PIP scandal: an analysis of the process of quality control that failed to safeguard women from the health risks', May 2013, Journal of the Royal Society of Medicine

[152] Regulation on in vitro diagnostic medical devices (n13).

[153] Medical Devices Regulation (n12).

[154] Laura Maher, Niki Price, 'Ultimate Guide to IVDR for In Vitro Diagnostic Medical Device Companies', November 2022, Greenlight Guru.

[155] Zaide Frias, 'Update on EMA role in implementation of new legislation for medical devices (MDR) and in vitro diagnostics (IVDR)', 20 November 2019, Annual PCWP/HCPWP meeting with all eligible organisations

[156] MedTech, 'MedTech Europe Survey Report analysing the availability of Medical Devices in 2022 in connection to the Medical Device Regulation (MDR) implementation', 14 July 2022, at p6.

[157] Ibid.

[158] Kosta Shatrov, Cart Rudolf Blankart, 'After the four-year transition period: Is the European Union's Medical Device Regulation of 2017 likely to achieve its main goals?', December 2022, Elsevier Health Policy, Volume 126, Issue 12, Pages 1233-1240, p1235.

up, resulting in severe and unpredictable delays, which put the seamless availability of medical devices and the prioritization of innovation in the EU healthcare sector at risk[159].

Recommendations:

The issue was partially addressed already by the European Commission through the proposal 2023/0005 (COD) [160], amending the transitional provisions of the EU Medical Devices Regulation (MDR) and the sister regulation, In Vitro Diagnostic Medical Devices Regulation (IVDR). The Commission also acknowledges that "despite considerable progress over the past years, the overall capacity of conformity assessment ('Notified') Bodies remains insufficient to carry out the tasks required of them", and that "many manufacturers are not sufficiently prepared to meet the strengthened requirements of the MDR by the end of the transition period[161]". The proposal will seek to extend the deadline of the transitionary period "from 26 May 2024 until 31 December 2027 for higher risk devices (class III and class IIb implantable devices except certain devices for which the MDR provides exemptions, given that these devices are considered to be based on well-established technologies) and until 31 December 2028 for medium and lower risk devices (other class IIb devices and class IIa, class Im, Is and Ir devices)[162]". This extension is subject to certain conditions, such as the devices must continue to conform with the MDD and must not undergo substantial changes. While these extensions might delay a potential crisis, a long-term investment is required in order to support the regulatory procedure introduced with the MDR. But considering the length of the extension of a staggering four years, it could potentially be enough time for the Commission to resolve these issues and create more Notified Bodies to streamline and speed the conformity assessment procedure, as well as make the timeline for it more consistent.

Constraints and Considerations:

The industry has welcomed the EU proposal[163], but it is noted that this is only an extension and does not actually fix the fundamental underlying issues regarding Notified Body availability that was discussed above. Absence of a sufficient number of Notified Bodies to support the industry's demands to keep the process smooth and relatively fast, the negative impact is unlikely to change, and soon manufacturers will start deprioritizing the EU, much like the MedTech survey suggests[164]. The creation of more Notified Bodies is of course easier said than done, as the Commission has clearly struggled to meet this goal, but it is a necessary part if the MDR is to succeed, and perhaps an increase in budget is needed to hasten their creation. The Commission could also consider alternative ways of approaching the challenge such as following the Medical Device Coordination Group's suggestion of hybrid auditing[165], or following MedTech's suggestions such as speeding up the certification procedure.

[159] Ibid.

[160] Proposal REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, amending Regulations (EU) 2017/745 and (EU) 2017/746 as regards the transitional provisions for certain medical devices and in vitro diagnostic medical devices, Brussels, 6.1.2023, COM(2023) 10 final.

[161] Ibid. p1.

[162] Ibid, pp7-8.

[163] MedTech, 'MedTech Europe welcomes the adoption of amended transitional provisions of the Medical Devices Regulations and calls for continued work to address outstanding implementation challenges', 7 March 2023, MedTech Press Release.

[164] Ibid.

[165] MDCG Position Paper, "Transition to the MDR and IVDR: Notified body capacity and availability of medical devices and IVDs", August 2022, MDCG 2022-14.

1.8 Liability and Insurance

(PR6) Changing the draft Article 7 of the new Product Liability Directive Update. A few suggestions

Main author: Francesca Gennari

Addressee:

This recommendation can be suggested to consumer advocacies and implemented by the EU institutions and then, at a national level, by Member States (MS) parliaments.

Context/history of the problem:

In the application of the current Product Liability Directive (PLD)[166], Article 3 PLD specifies that the producer, intended as the manufacturer, is the person who is primarily liable for the product (Articles 1 and 3 PLD). Other subjects, such as the importer or supplier, can be considered liable only if the producer is not identified or identifiable. The main problem is that if consumers were not able to identify the producer, then they could be time-barred from bringing a product liability action based on the directive. In thirty-eight years of the PLD application, it has become clear that Article 3's rule- that the producer is the main person liable- could be difficult to apply in practice because of the increasingly complex international organizations of certain sets of products (such as vaccines[167]). The Court of Justice had to evaluate whether the complainant being time-barred was fair. It is maintained that this problem will resurface in different ways even with the updated Article 7 of the Product Liability Directive Update (PLDU)[168], which will substitute Article 3 PLD. This new Article 7 PLDU will be extremely relevant for new technologies as well. In fact, it is likely that the majority of IoT consumer objects (which might have very complex product and value chains) will be covered by the rules set in the novel PLDU.

Definition of the problem:

The application of the new Article 7 PLDU as it is in the proposal is likely to become a sub-efficient set of norms that will not address the complexity of the product and value chains of connected objects such as IoT devices. In the PLDU explanatory memorandum[169], Article 7 PLDU is considered a mean to help the consumer because it establishes that there is always a person that is responsible for compensation in the EU. Nevertheless, the text of Article 7 is straightforward and, if a manufacturer (a producer in the PLD text) exists, the consumer must contact them, independently from where they are located, with the help of their Member States (MS). Nevertheless, the Article does not give any further indication about how MS should help consumers find the manufacturer. Only after having found out that it is not possible to reach the manufacturer, because it is either i) unknown/reachable ii) located outside the EU consumers can reach out to other subjects in the list. The list of economic operators to ask for compensation is quite rigid and is structured as follows. Beyond the manufacturer, the other

[166] Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products [1985] OJ L/29.

[167] See the case C-127/04. *Declan O'Byrne v Sanofi Pasteur MSD Ltd and Sanofi Pasteur SA* ECLI:EU:C:2006:9.

[168] Product Liability Directive Proposal (n19).

[169] PLDU Explanatory Memorandum (within the proposal see footnote 2) p. 2.

economic operators mentioned are the importer, the authorized market representative, the fulfillment service provider, the refurbished product trader/seller, and the distributor (former supplier). The criteria to scroll down the list of these economic operators is the same as for the manufacturer: the economic operator contacted by the consumer must be unknown, unreachable or located outside of the EU. In particular, the distributor could be considered liable if they do not help the consumer who endured the damage to contact the manufacturer. In fact, Article 7(5) PLDU states that the distributor will be considered liable if " *(a) the claimant requests that the distributor identify the economic operator or the person who supplied the distributor with the product, and (b) the distributor fails to identify the economic operator or the person who supplied the distributor with the product within 1 month of receiving the request*"[170]. Finally, the last category of economic operators that could be liable are online platforms that allow consumers to conclude distant contracts with traders. They are the only economic operators which could be liable at the same conditions as the distributors, as they do have a specific duty to provide the consumer with the identity of the manufacturer within one month of the consumer's request[171].

This solution is suboptimal as it makes it extremely complicated for the consumer to get compensated and they risk being time-barred as they only have 3 years to ask compensation for damages since the damage occurs[172]. Besides, this would be the opposite outcome of the application of the explicit rationale of Article 7 PLDU and that could be found in the explanatory memorandum which is to always provide a subject that is liable in the EU.

The fixed order of this new list of potentially liable economic operators constitutes a problem, especially for connected objects such as low-risk IoT devices and robotics applications as their product and value chains are much more complex than the ones of traditional consumer objects, even electronic ones. The further level of complexity is given by the fact that it is difficult for the average consumer and the average lawyer to understand whether the damage was caused by the object, by its software, or by the interaction between the product's software and applications downloaded from a third party. Moreover, among scholars, some have rightfully highlighted that the PLDU does not consider the transnational dimensions of future product liability claims[173]. This will become a major problem, especially for connected objects such as IoT devices, since the major IoT device manufacturers are located outside of the EU and it is not a given that they have an authorized representative or a distributor in the EU. If they do have it, they will re-direct the consumer to a foreign jurisdiction which might not offer the same level of protection as an EU MS. In practice, this more than probable scenario clashes with the Explanatory memorandum that specifies that the long and rigid list of economic operators is justified by a pro-consumer concern, namely, to always identify a subject that is liable in the EU.

Article 7 PLDU is a problem not only for consumers who need to scroll through the list from the further subject to the closer one to get compensation, but also for all the economic operators. Some of them, such as distributors, might not be informed about the exact details of the manufacturer's whereabouts and might need to take more time than what Article 7 PLDU allows to contact the manufacturer and redirect the consumer to them. Despite this hierarchy of

[170] Article 7(5) PLDU

[171] Article 7(6) PLDU.

[172] Article 14(1) PLDU.

[173] Jean- Sébastien Borghetti, "Taking EU Product Liability Law Seriously: How Can the Product Liability Directive Effectively Contribute to Consumer Protection?".(2023)(1) French Journal of Public Policy, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4502351>.

subjects that need to be sued, it is likely that consumers will start suing platforms and distributors first, because they are the subjects they have dealt directly with, rather than obscure and far-away manufacturers.

Proposed policy recommendation aimed at solving the problem:

This policy recommendation consists of two alternative drafts to the latest version of Article 7 suggested in the proposal. Both drafts propose a modification of the legal basis of the PLDU. At the moment, the PLDU's legal basis is Article 114 TFEU which is the clause concerning market harmonization. This means that its rationale should be to create a balance among the different stakeholders (consumers and manufacturers alike).

It is hereby recommended to change the legal basis of the proposal and adopt Article 169 TFEU to better protect consumers. It would be the only way to provide a solid and coherent legal basis to the explicit references to consumer protection that are contained in the explanatory memorandum[174]. As a consequence, redrafting Article 7 PLDU would entail identifying the distributors and the online platforms (which most of the time are more solvable than manufacturers) as the subjects to which the complainant should ask for compensation first. Besides, their importance as the subjects that are closer to consumers is implicitly highlighted by the same text of Article 7(5)(6) PLDU which gives a specific span of time to redirect the complainant to the manufacturer. If they do not comply within the given time, they are considered liable. Moreover, MS, with the help of the EU, should lay the basis for a pan-European recovery action. This would mean that the distributor or the platform could ask for repayment from the manufacturer after they have compensated the complainants that have demonstrated the correctness of their claim. This solution would not be a new one: France and Denmark fought for years with the European Union Court of Justice (EUCJ) on this matter, as this rule was the basis of their product liability laws, although they showed differences in the implementation[175]. This solution would grant the consumer a faster and more effective remedy and the distributor or online platform would have sufficient economic leverage to compel the manufacturer to pay, especially if it is located outside of the EU.

Despite this, it is not likely that the aforementioned solution will be adopted since the current PLD's legal basis is Article 114 TFEU, and the PLDU is just an update of that document, rather than an entirely new one. An alternative that could be more easily adopted would be to abstain from switching legal bases while amending Article 7 PLDU. The new Article 7 would include the following modification: that the manufacturer is not the primary responsible subject that is liable if it is not based in the EU. This would give the importer and the authorized representative the role of subjects that can compensate the victim of product liability damage. Then, the other subjects that are mentioned would follow in the cascade of responsibilities (i.e., the fulfillment service provider, the distributor, and the online platform). However, the new text should also include a mention that MS must guarantee a recovery action for importers and authorized representatives (as well as for the other economic operators) towards the manufacturer. This result could be achieved (but maybe not as easily) by also using the current regulations on

[174] Francesca Gennari, "A tale of two cities? Fennia v Philips and Article 7 of the Product Liability Directive Update", Forthcoming EuCML December issue 2003

[175] Case 52/00 *Commission of the European Communities v French Republic* ECLI:EU:C:2002:252 ; Case C-402/03 *Skov Æg v Bilka Lavprisvarehus A/S and Bilka Lavprisvarehus A/S v Jette Mikkelsen and Michael Due Nielsen* ECLI:EU:C:2006:6.

private international law. Specifically, there should be an explicit reference to Article 5 of the Rome II regulation[176], which sets rules about product liability cases even beyond the EU.

Constraints of the policy recommendation:

The recommendation does not take into consideration the specifics of an EU recovery action for damage as far as the first alternative (with Article 169 as a legal basis) is suggested. This goes partly outside the scope of the present policy recommendation which focuses mainly on product liability and not on judicial remedies. It is true that also for the second alternative (with Article 114 TFEU as a legal basis), there is a means to effectively ensure a recovery action through private international law. It is thus recommended that policymakers try to make these two elements of product liability (substantial law) and recovery actions (procedural law) communicate with each other, for instance by referring to articles of substantial law concerning product liability in procedural laws and vice-versa. In addition to that, there will also be the need to consider that all the relevant rules to the issues should be updated for the challenges caused by objects with digital content such as the IoT. For instance, as the Rome II regulation's Article 5 on product liability has not been modified since 2007 yet, it does not explicitly consider data or IoT objects, whereas the PLDU does. Because of the procedural law aspect that is inherent to these policy recommendations, it is suggested that more financial support is provided to train judges and lawyers to these new kinds of disputes.

(PR7) "Robotics and biorobotics in the law of personal injuries compensation and rehabilitation"

Main author: Maria Gagliardi

Title:
Robotics and biorobotics in the law of personal injuries compensation and rehabilitation

Addressee:
judges in personal injury cases, bioengineers, (mainly forensic) doctors, researchers in law

Context / history of the problem:
Similar to the tendencies occurring in other legal systems, in the last 50 years, there has been an evolution in the way in which the Italian legal system gives a right to compensation for personal injury damages to injured persons. The evolution dealt with: (i) the conditions under which the right to health is actionable against a tortfeasor; (ii) the heads of damages to which the victims are entitled; and (iii) the definition of personal injury itself and the distinction between pecuniary and non-pecuniary losses. These aspects were defined when personal injuries and their medical treatment were "traditional" and could not include in almost any way the availability of technological solutions such as robotics, prosthesis, bio-materials and so on. Now, the development of such technological solutions that are useful also in the medical treatment of injuries and in rehabilitation rises many questions about their relevance for the legal concepts, doctrines and rules that have evolved in the last 50 years.

Definition of the problem:

[176] Regulation (EC) No 864/2007 of the European Parliament and of the Council of 11 July 2007 on the law applicable to non-contractual obligations (Rome II) [2007] OJ L 199/40.

There is uncertainty about the application of existing rules[177] to the cases where new technological solutions and supports (both biotechnological and robotic ones) enable injured people to recover the ability to perform at least some of the activities they were able to perform before the injury.
The notion of personal injury under Italian law aims to compensate the consequences of the psycho-physical impairment with "equivalent in money". In order to obtain evidence of the impairment and to measure it in an objective manner, judges, lawyers and also the legislator have established the necessity for an interaction with physicians (i.e., forensic or medico-legal doctors), based on a sort of sharing of the assessment: it is up to the doctors to assess the degree of impairment as a percentage referred to the functionality of the person as a whole. Drawing on this evaluation, the court applies the legal rules which give an economic value to such a medical percentage.

However, no legal rule explicitly includes or explains if and how a technological support (such as a prosthesis, among others biotechnological or bioengineering tools) can be considered as a substitutive means of at least a part of the percentage. At the same time, as far as personal injury litigation is concerned, in the decisions and in the motivations, judges and medical experts do not disclose if and when they take into consideration the availability of biotechnological solutions in the assessment of damages, even when it comes to that part of non-pecuniary loss which quantifies the difficulty of performing activities in a different way compared to not performing them at all.

The fact that the existence of biotechnological tools is not clearly included, neither in the rules governing the assessment of damages for personal injury, nor in the courts' reasoning, could produce differences among injured persons, above all as unequal results in the assessment for compensation, depending for instance on the consideration of the availability of biotechnological tools, on the varying sensibilities of judges, or on the cost of a tool. It is not possible for a victim (or for her practitioner, or for her insurer) at the moment of a claim to foresee if the assessment made by a court will take into consideration the availability of specific robotic or biorobotic tools, for instance by reducing the amount of any head of damages. This creates uncertainty for practitioners, for victims, for public and private insurers, and ultimately for the system. Furthermore, under uncertainty it is not possible to introduce, for instance, specific services, insurance coverages, tailored premiums and so on.

Proposed policy recommendation aimed at solving the problem:
We aim at inserting the new technological opportunities into the conceptual legal framework of personal injury, in order to better understand the impact that the developments in the biorobotic research field can have on rules and doctrines.

Policy recommendation 1, for lawyers, engineers and doctors: We suggest the co-operation of researchers and practitioners from the legal, medical and engineering domains with the goal of clarifying if it is possible (and whit which methodology) to measure, and thus to assess, the amount of functionality (as the percentage of the integrity of the person) that can be recovered with the adoption or use of specific biotechnological and biorobotic tools. The results could become both an amendment or specification of existing law, and an update of existing policy during the procedures of personal injury compensation.

[177] Articles 138 and 139 of the Italian Code of Insurance (D. Lgs. 7 settembre 2005, n. 209)

Policy recommendation 2, for judges and experts: in the lack of formal introduction of new rules or policies, we suggest that in all the personal injury cases, judges, experts and other actors (such as mediators or facilitators) should explain: if they take into consideration the availability of different types of technological tools, under which head of damages, and how they eventually quantify their contribution to the overall assessment.

To learn more about the topic and the problem:
Gagliardi Maria, 'Brevi note sulle tecnologie e la "riduzione" del danno alla persona. Prospettive di ricerca interdisciplinare in tema di cd reversibilità del danno alla persona in connessione con l'ausilio di biotecnologie (domande per I giuristi e domande per I medici legali)' (2022), XLIV Rivista Italiana di Medicina Legale 245;

Amram Denise, 'Post fata resurgo. Innovazione tecnologica e medicina rigenerativa: l'impatto sul danno alla persona' (2021), XLIII Rivista Italiana di Medicina Legale 1.

1.9 Cybersecurity compliance and policy design

(PR8) Enhancing the participation of ENISA in the definition of cybersecurity requirements

Main author: Federica Casarosa

Addressee:

European bodies involved in the trilogues

Context:

In April 2021, the European Commission published a draft proposal for a Regulation on Artificial Intelligence systems[178] (AI Act or AIA) aimed at striking a balance between the market need for a competitive and dynamic ecosystem and the need to minimise risks to the safety and fundamental rights of users and citizens. Among the numerous obligations that apply to high-risk AI technologies, the AI Act includes a provision addressing cybersecurity of AI systems. However, the wording provided by the Commission proposal fell short of addressing the wide variety of cybersecurity threats that AI can face throughout the design, development, and deployment phases. Moreover, the certification mechanism set up by the AI Act, though, does not provide for sufficient guarantees such as stakeholder engagement, expert evaluation, subsequent updates, etc. Although the amendments proposed by the European Parliament[179] improved the proposed text, there are some further considerations that need to be considered by policymakers.

Definition of the problem:

The AI Act proposal sets up a detailed organisational structure requiring Member States to establish a certification network that includes notifying authorities and notify conformity assessment bodies. Both are part of the process leading to the issuing of CE labels to high-risk AI systems that have passed the conformity assessment which is based on the general

[178] AI Act Proposal (n14).

[179] Amendments to the AI Act (n76).

requirements defined in Articles 8-15, that are relevant to any AI system developer and manufacturer. This certification process, however, does not include sufficient details and stakeholder involvement and improvements are needed to uphold the goal of certification mechanisms as trust-enhancing and transparency-enhancing instruments for manufacturers and consumers (i.e. users and deployers in the language of AIA). These improvements are not only relevant for the cybersecurity perspective, but more generally for the overall effectiveness of the certification mechanism. In the Commission's version, Article 15 refers only to resilience to attacks that may affect the integrity of the AI system, such as data poisoning and adversarial examples. This approach did not account for the wide number of potential threats that have been already mapped by the ENISA study on AI cybersecurity risks.[180] The amended version of art. 15 AIA has widened the type of envisaged risks, including for instance also model poisoning and model evasion. Although these are important updates, the most relevant amendment is to be found in the added para 1b, where the Parliament proposed to establish a dialogue between the ENISA and the newly created European AI Board to address any emerging issues across the internal market about cybersecurity. This provision is crucial, as it will allow the AI board to establish a liaison with the European agency that is devoted to study and analyse cybersecurity issues and challenges on a wider scale.

The role of the AI Board is clearly set in art. 56 b AIA (as amended by the EP), that gives the Board the task of examining, on its own initiative or upon the request of its management board or the Commission and issuing opinions on technical specifications or existing standards as well as on the Commission's guidelines. No specific guideline is provided as regards the role, the forms of communication and collaboration of ENISA.

Policy recommendation:

Clarify when and how the ENISA can be involved alongside the AI board to contribute to the definition of emerging cybersecurity issues.

Possible operational applications

Modify art. 41 (2) in the following way:

"2. The Commission shall, throughout the whole process of drafting the common specifications referred to in paragraphs 1a and 1b, regularly consult the AI Office and the Advisory Forum, the European standardisation organisations and bodies**, and ENISA** or expert groups established under relevant sectorial Union law as well as other relevant stakeholders. The Commission shall fulfil the objectives referred to in Article 40 (1c) and duly justify why it decided to resort to common specifications."

Modify art. 56 b in the following way:

"k) organise meetings **and publish common positions** with Union agencies and governance bodies **(e.g. ENISA)** whose tasks are related to artificial intelligence and the implementation of this Regulation;

To learn more about the topic:

[180] ENISA (2020) AI Cybersecurity challenges – Threat Landscape for Artificial Intelligence. https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges

Casarosa Federica (2022) 'Cybersecurity Certification of Artificial Intelligence: A Missed Opportunity to Coordinate between the Artificial Intelligence Act and the Cybersecurity Act' (2022) 3 International Cybersecurity Law Review 115.

(PR9) Reducing the risks of outdated cybersecurity requirements in European standardisation

Main author: Federica Casarosa

Addressee:

European bodies involved in the trilogues

Context:

In April 2021, the European Commission published a draft proposal for a Regulation on Artificial Intelligence systems[181] (AI Act or AIA) aimed at striking a balance between the market need for a competitive and dynamic ecosystem and the need to minimise risks to the safety and fundamental rights of users and citizens. Among the numerous obligations that apply to high-risk AI technologies, the AI Act includes a provision addressing cybersecurity of AI systems. However, the wording provided by the Commission proposal fell short of addressing the wide variety of cybersecurity threats that AI can face throughout the design, development, and deployment phases. Moreover, the certification mechanism set up by the AI Act, though, does not provide for sufficient guarantees such as stakeholder engagement, expert evaluation, subsequent updates, etc. Although the amendments proposed by the European Parliament[182] improved the proposed text, there are some further considerations that need to be considered by policymakers.

Definition of the problem:

The general requirements set up in arts 8-15 should be operationalised for (and adapted to) the specific type or class of AI systems. In order to do so, the AIA relies on harmonised standards that should be adopted according to the procedure for technical standardisation (art. 40 AIA). In the absence of such harmonised standards, the Commission may adopt common (technical) specification (art. 41 AIA). In this case the procedure is only sketched in the article: the responsibility for defining the common specification is allocated to the Commission through the creation of an internal committee. This should "gather the views of relevant bodies or expert groups established under relevant sectorial Union law." An advisory role is also allocated to the newly created European Artificial Intelligence Board, which shall issue opinions, recommendations, or written contributions on the use of harmonised standards and common specifications.

Before any AI system is put on the market, the AI system providers should follow a conformity assessment procedure, which can either be a self-assessment or performed with the involvement of a notified body. Except for the case of AI systems based on facial recognition (listed in point 1 in Annex 3 AIA), all the high-risk AI system providers may use the self-assessment procedure as conformity assessment. Thus, when harmonised standards are lacking and common specifications have not been adopted, the high-risk AI system providers will not only be able

[181] AI Act Proposal (n14).

[182] Amendments to the AI Act (n76).

to set up their own self-defined standards, but also be able to self-assess their own compliance to the standards.

The amendments proposed by the European Parliament have improved the original text by setting up a system that is more accountable and transparent.

First, the amended Art 40 AIA acknowledges the need to start the standardisation process at the European level, without relying on other international initiatives that may not completely overlap with the standards set up by the European legislation. The standardisation role is allocated to the CEN/CELEC. Yet, it is important to mention that the process is not left only to the standardisation entity, but the provision requires also to '*ensure a balanced representation of interests and effective participation of all relevant stakeholders*'. Second, the procedure for adopting the Common specifications by the Commission, according to the amended provision of the EP of Art. 41 AIA, is also more detailed, transparent and participatory in the EP's amendments when compared to the Commission's proposal: it requires a preliminary consultation of the Commission with the newly created AI office and AI Advisory Forum, a regular coordination with the latter as well as with the European standardisation organisations and bodies or expert groups established under relevant sectorial Union law, as well as with other relevant stakeholders. Then, the Commission is also required to provide reasoned explanations when diverging from the opinion of the AI Office.

The amendments are helpful to include the views and comments by the relevant stakeholders in the drafting phase of the harmonised standards as well as the common specifications. However, considering the rapid developments that characterise this type of technologies, and when considering the emerging cybersecurity threats, the AI Act is missing a timeline for the revision of the adopted standards.

Policy recommendation:

Introduce a deadline for the reconsideration of the adopted standards and common specifications to account for technical developments and emerging cybersecurity threats.

Modify art. 40 AIA adding the following para:

1d. At least every five years, the adopted standards shall be re-evaluated, considering the feedback received from the AI office, the Union agencies, and the governance bodies **(e.g. ENISA)** whose tasks are related to artificial intelligence, as well as interested parties. If necessary, the Commission may request standardization bodies to revise the existing standard.

Modify art. 41 AIA adding the following para:

5. At least every five years, the adopted standards shall be re-evaluated, considering the feedback received from AI office, with Union agencies and governance bodies **(e.g. ENISA)** whose tasks are related to artificial intelligence and the implementation of this Regulation and interested parties. If necessary, the Commission may revise the existing standard.

To learn more about the topic:

Casarosa Federica (2022) 'Cybersecurity Certification of Artificial Intelligence: A Missed Opportunity to Coordinate between the Artificial Intelligence Act and the Cybersecurity Act' (2022) 3 International Cybersecurity Law Review 115.

4. FUTURE WORK

The publication of the future iteration of this deliverable is planned for March 2025, i.e., the final month of the project. Following a similar method as the one exemplified in Section 2, D7.6 v.2.0 will leverage the acquired knowledge of the LaPoH concerning the research activities undertaken during the development of BRIEF and the needs that the technologists and scientists of WP3-WP6 express, for example about the missing topics they would need to receive guidance on and the gaps that need to be bridged. Moreover, additional relevant knowledge may be derived from ancillary independent endeavors within the other interdisciplinary projects where LaPoH's members are involved. Note that the priorities in terms of topics for the policy recommendations and the best practices will be adapted over the course of the project, so this is a non-exhaustive list that may undergo modifications.

Concerning **data management**, there are a number of challenges that need to be overcome when it comes to research undertaken in the biomedical domain. The secondary use of health data carries great promises of scientific progress but is also surrounded by risks and uncertainties, for example for what concerns the legitimate legal basis that foregrounds the sharing of data generated in a certain context for certain purposes (e.g., MRI scans for medical diagnosis) and then repurposed (e.g., testing the accuracy of organ simulators), and the rules applicable to such BRIEF's cases. Guidance may also be provided for what concerns how to run a Data Protection Impact Assessment (needed for instance when data of vulnerable populations is processed), while useful resources about Privacy-Enhancing Technologies (PETs) and more in general about good practices for data management and data sharing may also be produced. The obligations originating from the ePrivacy regulation should also be included in the regulatory analysis. The interplay of data protection requirements with exceptions for scientific research should also be analysed to provide for enablers that can be leveraged by BRIEF's scientists.

It will also be necessary to analyse the types and purposes of the specific **AI models and robotic machineries** involved in the various projects, as they may pertain to different legal regimes with varying obligations, for example in reason of the level of risk that they carry. For instance, the AI Act differentiates between unacceptable, high, limited and minimal risk systems. It is very likely that some of the AI applications can be categorized as high-risk systems if they pose a threat to human safety and fundamental rights and therefore be subject to stringent requirements. If our analysis will identify BRIEF's researchers as providers (as opposed to deployers) and the systems as safety components deployed in medical settings, they will nevertheless have obligations concerning, for example, the implementation of risk management procedures (Article 9), the use of high-quality training, validation and testing data (Article 10), the insurance of an appropriate level of transparency (Article 13) and of human oversight measures (Article 14), as well as robustness, accuracy and cybersecurity (Article15), among the others. A thorough examination of the actual use of AI models in context will also determine the liability regime that can be applicable to the developers of such robotic systems in case harms occur (e.g., whether or not the AI Liability Directive and the Machinery Regulation may apply) and will influence the strategy that should be devised to enhance safety and lower the chances of researchers' liability.

Furthermore, the framework of **public health** could also be integrated with additional insights. In the healthcare system, patient-centricity[183] is increasingly recognized as an approach that puts people at the heart of healthcare and social services, including care, support, and enablement, by respecting their needs and preferences, as well as by encouraging them to take on an active role. For example, patients could also be better involved in the research activities, since studies demonstrate that there are mismatches between the priorities of researchers and those of patients and clinicians.[184] In addition, the World Health Organization proposes the concept of "one health"[185] as an approach to policy and research design where there is the recognition that human, animal, plants and environmental health are interconnected and there should be a sustainable balance for the health of all. Given the pivotal role that animal testing has in many experimental phases of biorobotic devices and BRIEF's specific focus on sustainability, expanding the notion of health beyond that of individuals, communities and societies appear promising. Related to this notion is the Do Not Significant Harm approach that avoid supporting economic activities that do significant harm to any environmental objective within the meaning of Article 17 of Regulation (EU) 2020/852,[186] such as climate change mitigation, the circular economy and the sustainable use and protection of water and marine resources.

In the next deliverable, an important place will be dedicated to the complex matters related to **intellectual property**. Not only the need for guidance emerged from the survey reported in D7.2; but the expertise of LaPoH pointed out various aspects and rights that need to be addressed, encompassing copyright, patents, design rights and their interplay with the openness requirements of scientific research. Another focus will be on the **procedures** that researchers need to follow **for testing the technologies** that they develop, such as exoskeletons, for instance when it comes to the internal ethical policies of the participating academic institutions regarding research studies, as well as those mandated by the CTR and its national implementation concerning the authorization required by Minister of Health for clinical trials. If the necessity arises from the discussion with the researchers, we may want to draft checklists and procedures to enable BRIEF researchers to navigate the documentation requirements in a seamless manner. Another important aspect is the **FAIR management of research data** that is closely interrelated to issues pertaining to confidentiality, both from an IP and a data protection point of view.

Additional insights could be added to better determine a common vision concerning **research and innovation**. For example, among the various existing models of innovation, the Quintuple helix model by Carayannis et al.[187] frames the creation and circulation of knowledge and the related innovation as determined by five factors: the political system, the education system, the economic system, the media-based and culture-based public and the natural environment. The

[183] See e.g., the Eight Picker Principles of Person Centred Care. Available at: https://picker.org/who-we-are/the-picker-principles-of-person-centred-care/

[184] Sally Crowe and others, 'Patients', Clinicians' and the Research Communities' Priorities for Treatment Research: There Is an Important Mismatch' (2015) 1 Research Involvement and Engagement 2 <https://doi.org/10.1186/s40900-015-0003-x> accessed 21 December 2023.

[185] https://www.who.int/health-topics/one-health

[186] Regulation (EU) 2020/852 of the European Parliament and of the Council of 18 June 2020 on the establishment of a framework to facilitate sustainable investment, and amending Regulation (EU) 2019/2088 (Text with EEA relevance)

[187] Elias G Carayannis, Thorsten D Barth and David FJ Campbell, 'The Quintuple Helix Innovation Model: Global Warming as a Challenge and Driver for Innovation' (2012) 1 Journal of Innovation and Entrepreneurship 2 <https://doi.org/10.1186/2192-5372-1-2> accessed 21 December 2023.

identification of the role that such factors can play within the specific context of BRIEF will also better inform the interventions that can be applied.

The observations of this last section do not aspire to offer an exhaustive understanding of the regulatory and practical challenges of BRIEF. We are aware that many of the challenges will be elicited through the close collaboration with the researchers and technologists of the other WPs and through the ongoing cross-field regulatory analyses that will be illustrated in D7.4 and D7.5.

CONCLUSIONS

This deliverable focused on the policy recommendations and the best practices stemming from the cross-field regulatory analysis carried out during the first year of BRIEF project and published in D7.3, as well as from the stakeholders' needs illustrated in D7.2 and gathered through collaborative meetings with the technologists of the other WPs. This must be understood as a living document that will be integrated with the elements identified in the last section, as well as other potential topics of interest stemming from the close collaboration with the real-world R&I challenges faced in the other WPs.

APPENDIX I: TEMPLATE FOR POLICY RECOMMENDATIONS

Expected length: Ca 1000-1200 words

Structure

- **Title** (10-15 words max that clearly indicates the envisaged best practice)
- **Addressee** (50 words max)
- To whom is it addressed? Who should apply the best practice?
  Specify role, responsibilities and domain e.g., bioengineering researcher working in a public research institution and collecting data from sensors; medical personnel of the hospital in charge of collecting consent from patients; etc
- **Context / history of the problem** (150 words ca.)
  - How and where did the problem arise? Why is it important to solve it now?
    E.g., in a specific geographical area / time / domain of law; it is a new problem / well-known problem
- **Definition of the problem** (300 words ca.)
  - What kind of problem is it?
    E.g., a legislative gap, conflicting interplay of norms, ineffective government strategy, etc.
  - Why is it a problem? What are the risks arising from the problem if it is not solved?
    E.g., legal uncertainty that can hamper economic investment in a certain area, ignoring the needs of specific populations, etc.
  - For whom is it a problem?
    E.g., manufacturers, researchers, citizens, patients, policymakers, etc.
- **Proposed policy recommendation aimed at solving the problem** (400-600 words ca.)
  - What kind of policy recommendation it is?
    E.g., changes to existing laws, introduction to new legislation, new strategy for government, update of existing policy/service, etc.
  - How does the recommendation solve the problem?
    Depends on how you formulated the problem
- **Constraints of the policy recommendation** (150 words ca.)
  - Which margins were taken into consideration to limit the scope of the recommendation?
    A good solution is concrete and specific: it cannot solve overly big or broad issues
  - What additional enablers does it need to work?
    E.g., adequate financial support, adequate political support, rapid implementation before a certain regulation is adopted, etc.
- **References**
  All cited bibliographic sources (regulations, articles, webpages, etc) + any useful resource for the reader to learn more about the subject. Use OSCOLA style

Useful resources:

- https://www.wordlayouts.com/free/policy-brief-overview-with-templates-examples/
- https://www.icpolicyadvocacy.org/sites/icpa/files/downloads/icpa_policy_briefs_essential_guide.pdf

APPENDIX II: TEMPLATE FOR BEST PRACTICES

Expected length: Max 1000-1200 words

Structure:

- **Title** (10-15 words max that clearly indicates the envisaged best practice)
- **Addressee** (50 words max)
    - To whom is it addressed? Who should apply the best practice?
      Specify role, responsibilities and domain e.g., bioengineering researcher working in a public research institution and collecting data from sensors; medical personnel of the hospital in charge of collecting consent from patients; etc.
- **Context/history of the problem/challenge** (150 words ca.)
    - How and where did the problem/challenge arise? Why is it important to solve it now?
      E.g., in a specific geographical area / time / domain of science or practice; it is a new problem / well-documented problem; etc.
- **Definition of the problem/challenge** (300 words ca.)
    - What kind of problem/challenge is it?
      E.g., an overly complex process? The concrete application of (abstract) legal requirements?
    - Why is it a problem/challenge? What are the risks arising from the problem/challenge if it is not solved?
      E.g., impossibility to test, distribute or sell a developed product, impossibility to publish research results, liability risks, risks to the safety of users, etc.
    - For whom is it a problem?
      E.g., manufacturers, research subjects, researchers, citizens, patients, policymakers, etc.
- **Proposed best practice aimed at solving the problem** (400-600 words ca.)
    - What kind of best practice is it?
      E.g., practical instructions, helpful applications and tools, international standards, procedures, etc.
    - How does the best practice solve the problem/challenge?
      Depends on how you formulated the problem/challenge
- **Constraints of the best practice** (150 words ca.)
    - Which margins were taken into consideration to limit the scope of the best practice?
      A good solution is concrete and specific: it cannot solve overly big or broad issues
    - What additional enablers does it need?
      E.g., adequate financial support, the responsible person's authorization, new skill acquisition, new machineries, novel work organization; etc.
- **References**
  All cited bibliographic sources (regulations, articles, webpages, etc) + any useful resource for the reader to learn more about the subject matter. Use OSCOLA style