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*
01-05-25	First draft	Arianna Rossi	First draft of the complete document v.0.1.	Giovanni Comandé
8-07-25	Draft review	Giovanni Comandé	Content review v.0.1.	Giovanni Comandé
27-08-25	Content update	Arianna Rossi Francesca Gennari	Reviewed draft v.0.2	Giovanni Comandé
07-09-25	Content update	Arianna Rossi Francesca Gennari	Final draft addressing comments of reviewers v.0.3	Giovanni Comandé
17/09/25	Draft for LaPoH AB	Arianna Rossi Francesca Gennari	Draft v.04 submitted to LaPoH AB	Giovanni Comandé
29/09/25	Final draft	Arianna Rossi Francesca Gennari	Final draft addressing comments of reviewers	Giovanni Comandé
29/09/25	Final version	Arianna Rossi Francesca Gennari	Submission	Giovanni Comandé

ACKNOWLEDGEMENT

This report is the outcome of collaborative work. The main authors are Arianna Rossi, Francesca Gennari and Prof. Giovanni Comandé. However, other researchers of the Law and Policy Hub, of its Advisory Board and of its network have contributed with policy recommendations and best practices, namely in alphabetical order: Floriana d'Ambrosio, Irene Aprile, Giovanna Elisa Calabrò, Vittoria Caponecchia, Irina Carnat, Federica Casarosa, Georgios Christou, Simona Crea, Tommaso Crepax, Michele Emdin, Ilaria Fagioli, Alessio Fasano, Chiara Gallese, Maria Gagliardi, Renza Barbon Galluppi, Marco Germanotta, Sabrina Grigolo, Maria Cristina Mauro, Alessandro Mazzarini, Andrea Parziale, Ciro Pappalardo, Ludovica Paseri, Claudio Passino, Robert Lee Poe, Milena Sirtori, Daniela Spajić, Stefano Tramacere, Leopoldo Trieste, Giuseppe Turchetti. We also thank them, as well as Domenico Camboni and Gianclaudio Malgieri, 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 analyses (published in deliverables D7.3, D7.4 and D7.5) have 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 Open Data Directive, the European Health Data Space Regulation, the Data Act, 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), 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 and on the previous version of this report published at the end of 2023, the report provides a 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, validate and integrate the present policy recommendations and best practices, as established in the sustainability plan.

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 3.1), Artificial intelligence law and governance (see 3.2), Regulation of medical devices and health law (see 3.3), Liability and insurance (see3.4), and Cybersecurity compliance and policy design (see 3.5). A mapping of the regulatory framework that highlights the complexity and interplay of the relevant legal provisions has been illustrated in the three iterative reports dedicated to the Cross-field regulatory analysis (D7.3, D7.4 and D7.5) 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 who 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 rise to legal uncertainty. An additional challenge is represented by the fact that many EU legislative proposals regarding technological aspects were still under negotiation within the EU Trilogue during the progress of the project and have evolved; while other approved regulations still need to be implemented into national laws or be adapted to the national legal system. Moreover, case law related to certain legislative acts that could provide legal certainty is still lacking.

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 even later on (e.g., in the view of commercialization) 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 contributions 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 lawmakers 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 (or were) under negotiation between the relevant European bodies. Some of these have been published as editorials by international journals, while others have been submitted to decision-makers as part of public consultations to increase their efficacy (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, for example, through awareness panels and webinars. Lastly, this deliverable also systematizes a broader set of interventions that encompass policy recommendations and best practices to enable the development of compliant-by-design outcomes of biorobotic research.

This deliverable must be understood as a living document, as it was published in two iterations: the first in December 2023 and the final one in September 2025, at the end of the project. Thus, the final version of the report contains two corresponding sets of recommendations. As mentioned earlier, in the period between the two publications, many regulations were approved. Moreover, national provisions and guidance were issued. For transparency, it was decided to maintain in the final version even those recommendations that referred to draft legislative proposals, as many of these contributions have been published in other venues and have therefore contributed to the lawmaking efforts and scholarly debate. Hence, the final version of the report contains an exhaustive mapping of the relevant topics, policy areas and best practices that are relevant to the various BRIEF's R&I activities. For clarity, the year when the contribution was provided has been added and all the parts that have been added since the publication of the first version of the report are written in green.

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 ties these inputs into the future work and the strategy foreseen in D7.8 "Sustainability Plan".

2. METHODOLOGY

BRIEF envisions the creation of a comprehensive, decentralized infrastructure with innovative laboratories and machinery for conducting cutting-edge research in the fields of robotics and biorobotics across a wide range of projects. We describe hereby a selection of the research projects that are 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. This information has been integrated with the content from the presentations sent by the BRIEF's labs[1] after the joint BRIEF workshop held on 5th March 2024 at Sant'Anna School of Advanced Studies titled "BioRobotics Research and Innovation Engineering Facilities: scenari di ricerca e d'innovazione".

[1] This overview is rich but incomplete. We hereby report only the content of the presentations we received.

2.1.Overview of the activities of the BRIEF's labs

The WP3 "BioRobotics Science to Engineering Translation" labs within the BRIEF project focus on developing and testing biomedical, robotic, and computational technologies that span from tissue engineering and microfluidics to soft robotics, photonic sensing, and neurophysiological monitoring. These labs produce or work with a wide range of devices and systems, including organotypic tissue constructs, microrobots, wearable and surgical robots, fluidic actuators, photonic circuits, and intelligent platforms for clinical decision support. They also support rapid prototyping, biohybrid systems, and high-performance computing for AI model training and simulation. The datasets generated across WP3 are equally diverse. Labs collect biosignals such as EEG, EMG, ECG, and fNIRS; imaging data from confocal and multiphoton microscopy; mechanical and material characterization data including force, pressure, and deformation profiles; and optical transmission spectra. Several labs maintain repositories of 3D design files and sliced models for fabrication, while others generate large-scale simulation data for machine learning. These datasets enable advanced analysis, modeling, and validation of biomedical and robotic systems, supporting translational research and personalized medicine.

The WP4 "BioRobotics Platforms" labs contribute to BRIEF by developing advanced platforms and devices for biomedical testing, diagnosis, rehabilitation, and surgical intervention. These include modular systems for cardiovascular device testing (C-LOOP), robotic tele-examination and scanning platforms for early diagnosis (SAPIO), embedded photonic sensors for haptic robotics (PHIS), flexible and modular robotic systems for surgery and diagnostics (FIRPADS), biomechanical evaluation setups for wearable robotics (BIOMECH), biohybrid and biomimetic prostheses and exoskeletons (B2R), and robot-assisted surgical navigation systems with telemedicine capabilities (ARTS). The labs collectively support the design, prototyping, and validation of high-TRL devices such as artificial hearts, smart prosthetics, capsule endoscopes, soft surgical robots, and AI-driven diagnostic tools. The datasets generated across WP4 are rich, including flow and pressure profiles, particle image velocimetry, and durability data for cardiovascular devices; volumetric, chromatic, and thermal scans of the human body; interaction forces, EMG, HR, and GSR signals during teleoperation; tactile sensing data from photonic sensors; physiological and biomechanical measurements during surgical simulations and daily living activities; and medical image datasets linked to robot and surgeon kinematics. These datasets enable validation, benchmarking, and algorithm development for robotics, AI, and biomedical applications, supporting both clinical and research advancements.

WP5 "BioRobotics and Health" establishes a science-based approach to biorobotics by integrating foundational knowledge from biology, neuroscience, and clinical sciences into the design and evaluation of robotic systems. The work package is structured into four interdisciplinary areas—CELL-Health, ANI-Health, CLI-Health, and TECH-Health—each hosting specialized laboratories that span molecular biology, animal models, clinical trials, and advanced sensing technologies. These labs support the development of biorobotic systems such as implantable neuroprosthetics, regenerative medicine platforms, robotic surgical tools, and telediagnostic infrastructures. The emphasis is on understanding biological mechanisms and translating that knowledge into more effective and personalized robotic solutions. Several labs imply the generation of complex biological, physiological, and clinical data. For instance, the CLI-Health area aims to construct Digital Twins of pathologies to simulate patient-specific treatment strategies, suggesting the integration of multimodal patient data. Similarly, labs focused on breath biomarkers, genomic analysis, and cardiovascular physiology produce biochemical, imaging, and signal-based data to support diagnostic and therapeutic innovation.

The eight labs of WP6 "BioRobotics and Sustainability" span a rich spectrum of expertise in sustainable materials, robotics, biofabrication, and environmental monitoring. B3MAT and SMAM4SoRo focus on eco-friendly materials and soft robotics, while STAR and FLARE advance biofabrication and human–robot interaction technologies. TERRA and UWR specialize in autonomous robotic systems for terrestrial and underwater environments, respectively, with strong emphasis on environmental safety and monitoring. COGITO provides digital tools for life cycle assessment and eco-design, and FLARE offers a high-precision testbed for aerial and mobile robotics. The WP6 labs develop and test a wide range of innovative technologies, including autonomous terrestrial, aerial, and underwater robots for environmental monitoring and infrastructure inspection (TERRA, UWR, FLARE); soft robots with embedded sensors and actuators fabricated using sustainable materials and additive manufacturing (SMAM4SoRo); biofabricated 3D materials for applications beyond biomedicine (STAR); eco-designed products and services supported by life cycle assessment tools and carbon calculators (COGITO); and sustainable biorobotic components such as biodegradable sensors and actuators (B3MAT). These labs also support the development of collaborative human–robot interaction systems, robotic swarms, and biohybrid energy harvesting solutions, contributing to high-TRL (Technology Readiness Level) prototypes and services.

Across all four work packages, there is a clear convergence in the use of multimodal, high-resolution datasets to support the development, validation, and personalization of biorobotic systems. While each WP has a distinct focus, they share several dataset themes. For example, physiological and biomechanical data are central across WP3 (e.g., EMG, joint angles, gait analysis), WP4 (e.g., surgeon kinematics, cardiovascular signals), WP5 (e.g., visceral reflexes, movement disorders), and WP6 (e.g., force sensors, user dynamics). Signal acquisition such as EEG, ECG, and respiration data appears in WP3 (N2LAB, INTOCADS), WP4 (BIOMECH, B2R), WP5 (PHYSIO & SLEEP, ISMI4PM), and WP6 (SAPIO-integrated platforms). Imaging datasets of various nature are used in WP3 (DONOR, REISSUE), WP4 (ARTS, SAPIO), WP5 (BBC, TELEMEDICINE), and WP6 (TERRA, COGITO). Human–robot interaction and control data are also common: WP3 and WP4 collect force feedback, tactile sensing, and teleoperation metrics; WP5 includes interaction data from robotic surgery and telediagnostic platforms; WP6 tracks pose and gesture data in collaborative robotic environments. Additionally, design and fabrication data such as CAD files and 3D printing models are used in WP3 (+Tech, SMAM4SoRo), WP4 (FIRPADS, B2R), and WP6 (STAR, B3MAT).

Despite their different domains, all WPs converge on the goal of data-driven validation, personalization, and high-TRL prototyping. The datasets cannot not only be used for device testing and optimization, but may also be used for training AI models, developing decision support systems, and enabling human–machine interaction. This shared emphasis on rich, structured, and interoperable datasets reflects a broader strategy within BRIEF to support reproducible, scalable, and interdisciplinary research.

2.2.BRIEF's relevant regulatory frameworks

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 the multiple fields described above (see Figure 1).

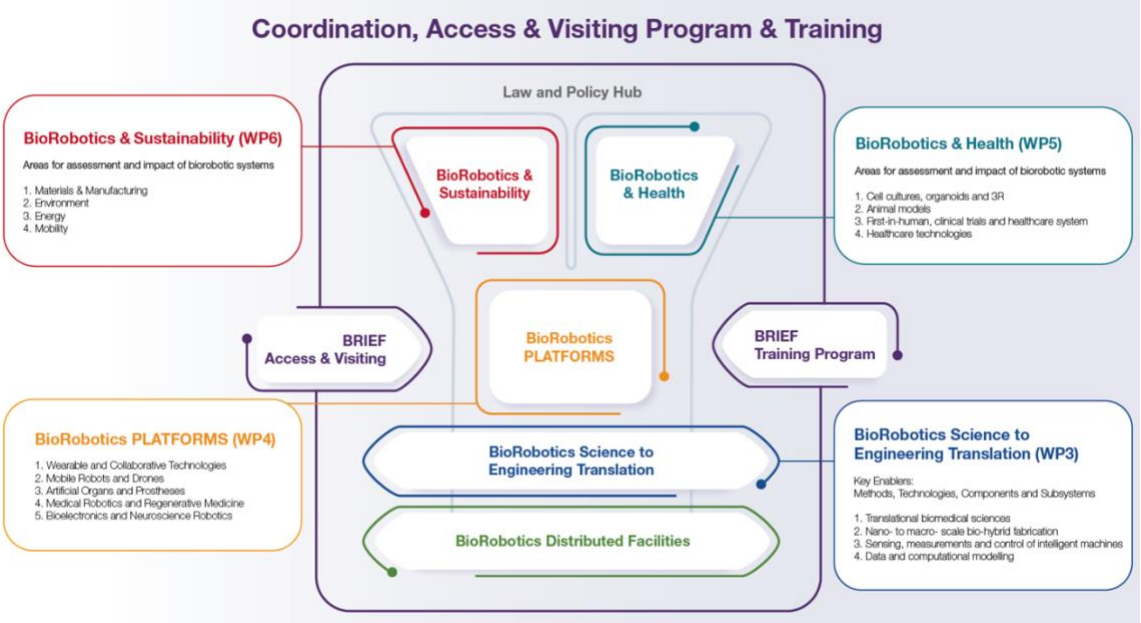

Figure 1. The schema illustrates the interplay between the various WPs and the framework set by the Law and Policy Hub. Source: "Annex B – Part 2: BRIEF – Biorobotics and Innovation Engineering Facilties" of the grant application (p. 21). Available at: https://www.santannapisa.it/it/pnrr-santanna/brief

The 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. The evolution of the legal framework over the progress of the project has been reflected in the ever-evolving mappings published in D7.4 and D7.5. These legislative endeavours are part of the recent EU Commission's initiatives concerning digitalisation, datafication and innovation, such as the EU Digital Strategy[2] and the EU Data Strategy.[3] An additional aspect that was highlighted concerns the complex ethical values that govern biorobotics research and that are transposed into general or sectoral 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 a need to establish a common framework that facilitates and regulates the performance of clinical trials and the development of safe-by-design medical devices. Research and innovation must uphold fundamental rights and protect research participants, including vulnerable populations, by ensuring the confidentiality of data, while striving to adhere to the principle of openness that emphasizes the need for replicable experiments and derives from open science policies.

Whereas we refer the reader to the in-depth analysis reported in D7.5, we summarize here the main outcomes in terms of applicable laws that need to be examined to understand enablers and challenges to be addressed.

[2] Shaping Europe's Digital Future, https://digital-strategy.ec.europa.eu/en

[3] 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

Data laws. The General Data Protection Regulation[4] 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[5] represents the counterpart of the GDPR and intends to encourage and govern the free movement of non-personal data across borders by abiding to cybersecurity requirements. The Data Governance Act[6] sets up novel mechanisms meant to enhance trust in data sharing and overcome 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 Regulation[7] lays down rules, standards, and practices for the primary use of data, as well as secondary use of data. The Data Act[8] 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. The Open Data Directive[9] requires public sector data to be made available in free and open formats, including data generated from publicly-funded research.

Health law framework. Regulation (EU) 2017/745 on Medical Devices[10] and Regulation (EU) 2017/746 on In Vitro Diagnostic Medical Devices[11] 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

[4] 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, pp. 1–88.

[5] 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 PE/53/2018/REV/1 *OJ L 303, 28.11.2018, pp. 59–68*.

[6] 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.

[7] Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847 (Text with EEA relevance) OJ L, 2025/327, 5.3.2025, ELI: http://data.europa.eu/eli/reg/2025/327/oj

[8] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) (Text with EEA relevance) PE/49/2023/REV/1 *OJ L, 2023/2854, 22.12.2023, ELI: http://data.europa.eu/eli/reg/2023/2854/oj*.

[9] 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 PE/53/2018/REV/1 OJ L 303, 28.11.2018, pp. 59–68.

[10] 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.

[11] 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.

and ethical review with the involvement of an Ethics Committee. The procedure to obtain 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[12] 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[13] would have instituted uniform requirements for non-contractual civil liability concerning damages caused with the involvement of AI systems, but it was retracted in 2025.

Cybersecurity. The legal framework encompasses the NIS Directive,[14] the NIS2 Directive,[15] and the Cyber Resilience Act proposal.[16]

Liability. The Product Liability Directive Update[17] concerns the liability of defective products and revises the existing Product Liability Directive 85/374/EEC.

Product safety. The Machinery Regulation[18] establishes health and safety requirements for the design and construction of machinery. The General Product Safety Regulation[19] modernises the EU general product safety framework and addresses the challenges posed to product safety by the digital economy.

[12] Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

[13] 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'.

[14] 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.

[15] 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) .

[16] Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act) (Text with EEA relevance) PE/100/2023/REV/1 OJ L, 2024/2847, 20.11.2024

[17] Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on Liability for Defective Products and Repealing Council Directive 85/374/EEC (Text with EEA Relevance) OJ L, 2024/2853, 18.11.2024

[18] 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.

[19] 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.

Intellectual Property Rights. The Software Directive[20] intends to ensure the protection of software by copyright in all the EU Member States. The Database Directive[21] regulates the legal protection of databases by copyright or by sui generis rights, with respect to their defining characteristics. The InfoSoc Directive[22] encompasses a wide spectrum of copyright-related matters, including the technological protection measures (TPMs) and digital rights management (DRM) systems. The Copyright in the Digital Single Market Directive[23] adapts certain key E&Ls to copyright to the particularities of the digital and cross-border environment. The Term Directive[24] harmonizes the duration of the legal protection granted upon copyright-protected works. The Trade Secrets Directive[25] harmonizes the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure. The Design Directive[26] harmonizes design law across Member States, with the aim of enhancing the internal market's functioning and supporting innovation. Lastly, the EU Design Regulation[27] updates the legal framework for the protection of designs at the EU level in light of technological developments.

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 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

[20] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Codified version) (Text with EEA relevance) *OJ L 111, 5.5.2009, pp. 16–22*
[21] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases *OJ L 77, 27.3.1996, pp. 20–28*
[22] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society *OJ L 167, 22.6.2001, pp. 10–19*
[23] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance.) PE/51/2019/REV/1 *OJ L 130, 17.5.2019, pp. 92–125*
[24] Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights (codified version) *OJ L 372, 27.12.2006, pp. 12–18*
[25] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (Text with EEA relevance) OJ L 157, 15.6.2016, pp. 1–18
[26] Directive (EU) 2024/2823 of the European Parliament and of the Council of 23 October 2024 on the legal protection of designs (recast) (Text with EEA relevance) PE/97/2023/REV/1. OJ L, 2024/2823, 18.11.2024
[27] Regulation (EU) 2024/2822 of the European Parliament and of the Council of 23 October 2024 amending Council Regulation (EC) No 6/2002 on Community designs and repealing Commission Regulation (EC) No 2246/2002 (Text with EEA relevance) PE/96/2023/REV/1. OJ L, 2024/2822, 18.11.2024

and treatment. One of the policy recommendations (PR6) addresses the challenges of translating principles of justice into machine learning pipelines in a lawful manner. Another concern that we address in PR7 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 (PR5) 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 and ask their consent (BP1, BP2, BP3). Other guidelines illustrate how to make AI applications explainable (BP8-BP9), 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.

2.3.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. Many regulations have only recently been approved and some or all of the provisions are not yet applicable. In addition, case law and official guidance are scarce or nonexistent for certain domains. 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[28] 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;[29] similarly, it has been applied to enable employees to more easily follow the cybersecurity policies of their organization[30] 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[31] 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.[32]

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

---

[28] 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.

[29] 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.

[30] 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.

[31] Michie, van Stralen and West (n21) 4.

[32] ibid 4–5.

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 have those skills within their team, then appropriate *resources* should be dedicated to acquiring those skills (e.g., through the acquisition of 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 senior researchers teach such a norm 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 a 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:[33]

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.[34] The techniques that implement the

[33] ibid 7.
[34] ibid 8.

general interventions pertain to the broader family of policies that can be summarized as follows:[35]

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
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.

2.4.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 intervention techniques that cover various functions and address the goal of facilitating compliance with the normative framework briefly reported in Section 2.2. 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.

[35] ibid 7.

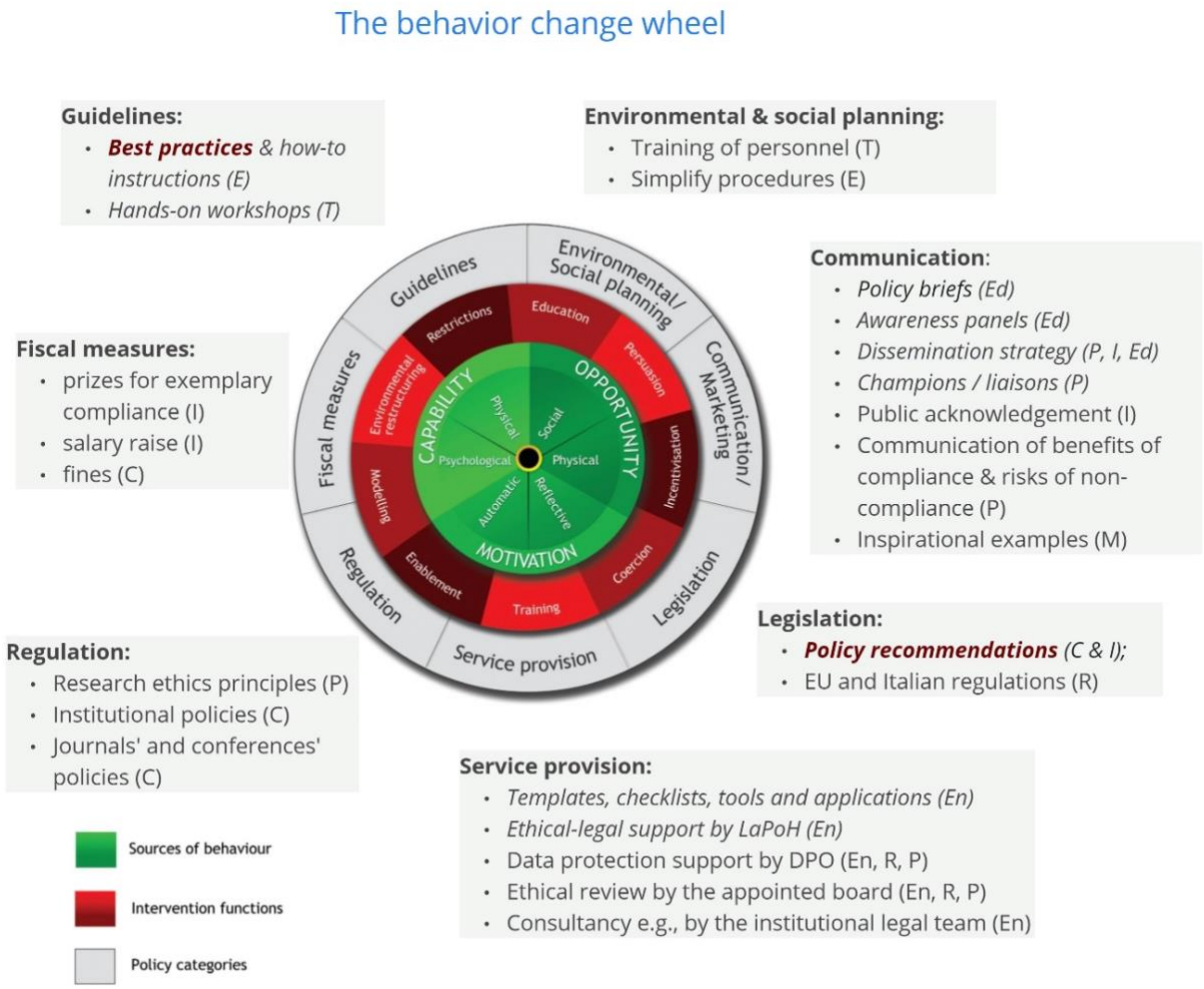

Figure 2. A non-exhaustive list of interventions that can facilitate compliance of researchers with legal and ethical norms governing biorobotics R&I, organized in the categories proposed by Michie, van Stralen and West[36] In italics, the intervention techniques that WP7 is putting in place to this 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[37]

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[38] 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.

[36] Ibid 7.
[37] Ibid 7.
[38] 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

2.4.1. 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 scientificaly grounded multi-layered solution that integrates personalized dynamic consent, user-centric interface design, and semantic interoperability that should inform the work of the Commission on the rulebook for data altruism consent (PR2);
- a redefinition of anonymization as a contextual, ethically grounded, and spectrum-based governance practice to guide harmonized and transparent implementation under the EHDS (PR3);
- a proposal for legislative and procedural alignment between Italy's FSE 2.0 and EDS systems and the EHDS, through opt-out provisions, institutional consolidation, and phased implementation (PR4);
- **a clarification of the roles and responsibilities of the actors that are involved in the accountability measures established for AI (PR5);**

- **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 (PR6);**
- **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 (PR7);**
- **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 (PR8);**
- a multifaceted strategy for integrating personalized medicine into healthcare, combining humanistic clinical practice with updated regulatory frameworks for equitable and secure implementation (PR9);
- **a proposal for extending the liability of manufacturers of defective components to importers and authorized representatives to ease the process of consumers' compensation (PR10);**
- a proposal for risk-based decision-making in software development for AI-powered products, supported by contractual safeguards and cybersecurity certification standards (PR11);
- **a revisitation of the concept of personal injury compensation within the robotic context (PR12);**
- a recommendation for harmonizing liability standards and clarifying legal concepts under the revised Product Liability Directive, while promoting insurer involvement and regulatory coordination (PR13);
- **a recommendation for a clearer involvement of the ENISA (European Union Agency for Cybersecurity) in the official definition of emerging cybersecurity issues in AI (PR14);**
- **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 (PR15).

These policy recommendations are timely and relevant, since they mostly address legislation that is currently (or was during the drafting of the first version) 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.

2.4.2. 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 report contains *proposed* best practices on a very broad set of topics, namely:

- a recommendation to adopt reusable, transparency-enhancing design patterns for privacy communication in research, supported by authoritative resources like CNIL's library (BP1);

- a redesign of privacy and consent policies using user-centered transparency, contextual integrity, and tailored communication to support ethical data sharing in digital health (BP2);
- a recommendation to design multimedia consent forms using strategic reading cues, layered content, and ethical co-design to support informed and context-sensitive decision-making (BP3);
- a recommendation for developers to adopt human-centered privacy engineering practices aligned with ISO standards (BP4).
- a best practice for implementing personalized dynamic consent platforms for data altruism to ensure ethical, adaptable, and legally compliant health data sharing (BP5).
- a recommendation to implement layered, GDPR-compliant consent mechanisms for data altruism in healthcare, supported by transparency, harmonized interpretation, and anonymization safeguards (BP6);
- a structured approach for public research actors to support responsible innovation in personalized medicine through FAIR data, accountability, proactive governance, and coordinated stewardship (BP7);
- a recommendation to apply GDPR principles to AI-based medical systems by ensuring transparency, human oversight, and non-discrimination through proactive, risk-based governance (BP8);
- a structured framework for AI developers to enhance MM-LLM reliability through expert validation, explainability techniques, and safeguards against automation bias (BP9);
- a recommendation for developers to design modular AI control systems for medical devices, integrating human-in-the-loop optimization, clinical validation, and early regulatory engagement (BP10);
- a recommendation for AI developers to standardize terminology and interpretation methods under the AI Act, enabling interoperable, accountable, and future-proof innovation (BP11);
- a recommendation to strengthen personalized medicine through inclusive research, stakeholder engagement, and the preservation of the therapeutic alliance in AI-supported care (BP12);
- a recommendation for dynamic evaluation models, risk-based assessment, real-world evidence integration, and participatory design to support safe and patient-centered adoption of digital health technologies (BP13);
- a proposal for multidimensional evaluation, stakeholder-specific indicators, simulation-based methods, and participatory frameworks to ensure sustainable and system-aligned HTA in personalized medicine (BP14);
- a strategy for structural digital health literacy, inclusive training, patient involvement, personalized support services, and outcome-based evaluation for participatory, equitable, and user-centered healthcare systems (BP15);
- an approach to economic, organizational, educational, regulatory, and ethical barriers through evidence-based models, interdisciplinary teams, participatory training, equitable access, and privacy-compliant integration of robotics in rehabilitation (BP16).
- a framework for reinforced safety expectations, post-market accountability, transparent liability rules, narrow interpretation of exemptions, and harmonized strict liability models for AI-powered medical devices (BP17);

- a guideline for transparent design, interdisciplinary collaboration, AI literacy, lifecycle accountability, and bias-aware data governance to ensure compliant and trustworthy deployment of AI systems in healthcare (BP18);
- a recommendation for proactive risk management, jurisdiction-sensitive deployment, compliance documentation, multimedia instructions, and litigation preparedness to ensure defensible and safe active prostheses under evolving liability regimes (BP19);
- a strategy for liability assessment through objective state-of-the-art evidence, mitigation documentation, and narrow interpretation of exemptions for software vulnerabilities in AI-powered medical devices (BP20)

Such best practices 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 by the punctual observations that arise from the close collaboration between technologists with legal-ethical expertise in WP7 and those with technical expertise in the other WPs. Various meetings have been held throughout the project to explore the specific technological development requirements within the research projects carried out by the various laboratories that are involved in BRIEF. The outcome of the first two meetings highlighted 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 has been set to enable the efficient addressing of the raised issues. As it is visible from the content of the report, the focus was then set on AI governance and data management, alongside complementary topics on regulation of medical devices, liability and cybersecurity. The workshops held with the bioengineers have been particularly fruitful and the results of such interdisciplinary discussions are included in BP10 and BP11.

2.4.3. Additional interventions

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 such actor that, through the elaboration of best practices stemming from the actual research needs of specific projects and the relevant domain knowledge, 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[39] 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 the website of the project[40] and 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[41].

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 can be adopted for disseminating the best practices and the policy recommendations (see the awareness panels and the webinars planned under D7.7). 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. Others were part of contributions to public consultations of the European Commission. Some of the policy recommendations and the best practices of the final version of the report were extracted from the chapters that compose the book on "Enabling and Safeguarding Personalized Smart Medicine", as outlined in "D7.7 Report on Research Dissemination and Awareness activities", with the support of generative AI. The authors of the chapters produced highly interdisciplinary contributions, ranging across various domains of law (data, AI, liability, health law), as well as medicine, bioengineering and economics. The drafting and the publication of the book, including the presentation and discussion of its contents at the dedicated BRIEF conference in October 2014, ensured a stricter engagement with the members of the LaPoH and of its Advisory Board, as well as novel ties with experts in other research institutions. This srtengthen the network of stakeholders, creating new synergies and the fruitful exchange of ideas across domains.

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

[39] https://altai.insight-centre.org/
[40] https://biorob-hub.eu/infrastructures/wp7/awareness-development-and-oa-dissemination/
[41] See the policy briefs already published on https://www.lider-lab.it/news/

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. In the previous version of this report, it was hypothesized to highlight to researchers the sectorial and institutional policies in place, as well as the specific actors who could support the implementation of their activities within the boundaries established by such policies. It was also suggested that, if needed, practical tools such as checklists could be developed to facilitate compliance with these procedures, particularly in collaboration with technologists and researchers from other work packages. However, as the project progressed, no specific requests or needs for such outputs emerged from the discussions with the involved partners. This absence of demand, combined with the diversity of institutional policies across participating institutions, confirmed that a centralized or standardized approach would have been difficult to implement and would have added limited value. As a result, these activities were not carried forward.

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.

2.5.Drafting and review process of the report

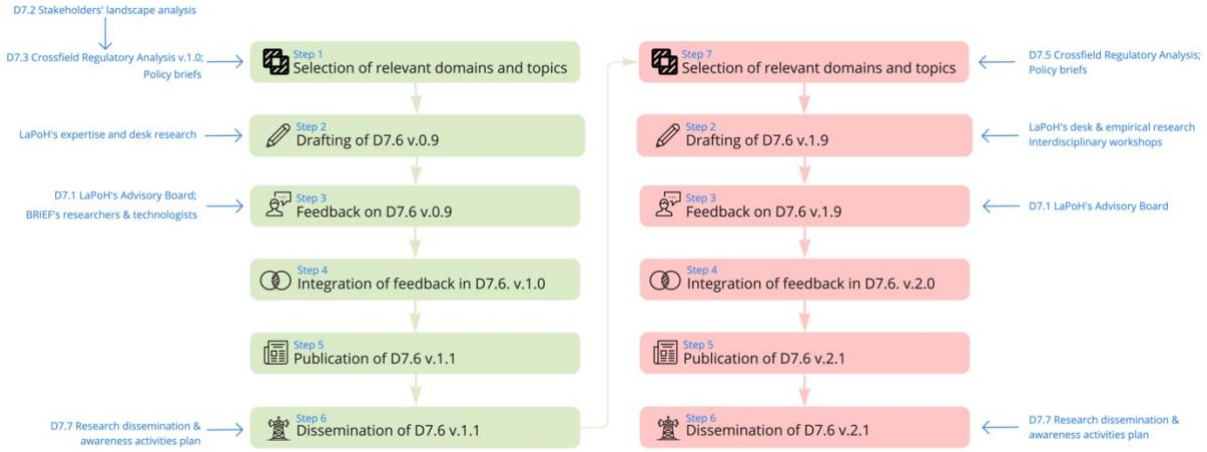

Figure 3. Diagram representing the steps of the methodology that has been followed for preparing this report in both its iterations, as well as the next envisioned steps. In blue on the lefthand side, the relevant input sources

This report has been created thanks to a collective effort and the participatory input of the relevant stakeholders, as Figure 3 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. After its publication, this report was further updated with novel policy recommendations and best practices that stem from the Cross-field regulatory analysis illustrated in D7.5 and the research activities of the LaPoH's members. The review and publication process followed the same steps and were concluded in September 2025.

3. POLICY RECOMMENDATIONS AND BEST PRACTICES

This section contains the policy recommendations (PR) and the best practices (BP) proposed by the members of the LaPoH. They have been organized according to their primary topic, however many of them pertain to more than one legal domain, reflecting the cross-domain nature of the WP7 research. Five principal areas have been identified, and the contributions are accordingly structured: personal and non-personal data management and data governance (Sec. 3.1.); Artificial Intelligence governance (Sec. 3.2.); regulation of medical devices and health law (Sec. 3.3); liability and product safety (Sec. 3.4.); and cybersecurity compliance and policy design (Sec. 3.5.).

3.1.(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 (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

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.[42] 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")[43], data portability lacked strength.[44] The challenges to its realization included:

(1) unclarities on textual interpretations of Article 20 GDPR,[45] 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

[42] EU Data Strategy (n6).
[43] GDPR (n7).
[44] 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).
[45] 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).

(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 portability in newer regulations, such as those found in Article 6 of the Free Flow of Non-Personal Data Regulation of 2018 [46] 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[47] (DMA) and the Data Act[48] have yet to produce their effects, others, like the European Health Data Space[49], 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[50] 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,[51] 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:

---

[46] 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; […]."

[47] 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.

[48] Data Act (n11).

[49] European Health Data Space (n10).

[50] 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).

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

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 ("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 [52] 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.[53] 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

[52] 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.

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

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 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.

Year of publication: 2023.

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

Main author: Arianna Rossi (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

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[54] (GDPR) that aim at "engendering trust in the processes which affect the citizens by enabling them to understand, and if necessary, challenge those practices".[55] Prior to the GDPR, the Directive 95/46/EC[56] 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.[57] However, the resulting disclosure has often resulted in lengthy, verbose, obscure privacy policies[58] 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"[59] related to the data processing practices and the individuals' rights about their data. Transparency is now understood as a "user-centric rather than legalistic"[60] 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,[61] as well as be subject to empirical tests to demonstrate their effectiveness.[62]

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

[54] GDPR (n7).
[55] 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>.
[56] 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
[57] 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.
[58] 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.
[59] Article 29 Data Protection Working Party (n 48) 5.
[60] ibid.
[61] 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.
[62] Article 29 Data Protection Working Party (n 48) 7.

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 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[63] by the Article 29 Working Party,[64] 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[65] 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,[66] 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[67] 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.[68] 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.

[63] Article 29 Data Protection Working Party (n 48).
[64] 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.
[65] Article 29 Data Protection Working Party (n 48) 12.
[66] Icon sets available at: https://www.garanteprivacy.it/temi/informativechiare#2
[67] 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).
[68] Available at: https://design.cnil.fr/en/design-patterns/ (English) and https://design.cnil.fr/fr/design-patterns/ (French).

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 in addressing design issues in privacy[69] 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):[70]

- **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.

[69] 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>

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

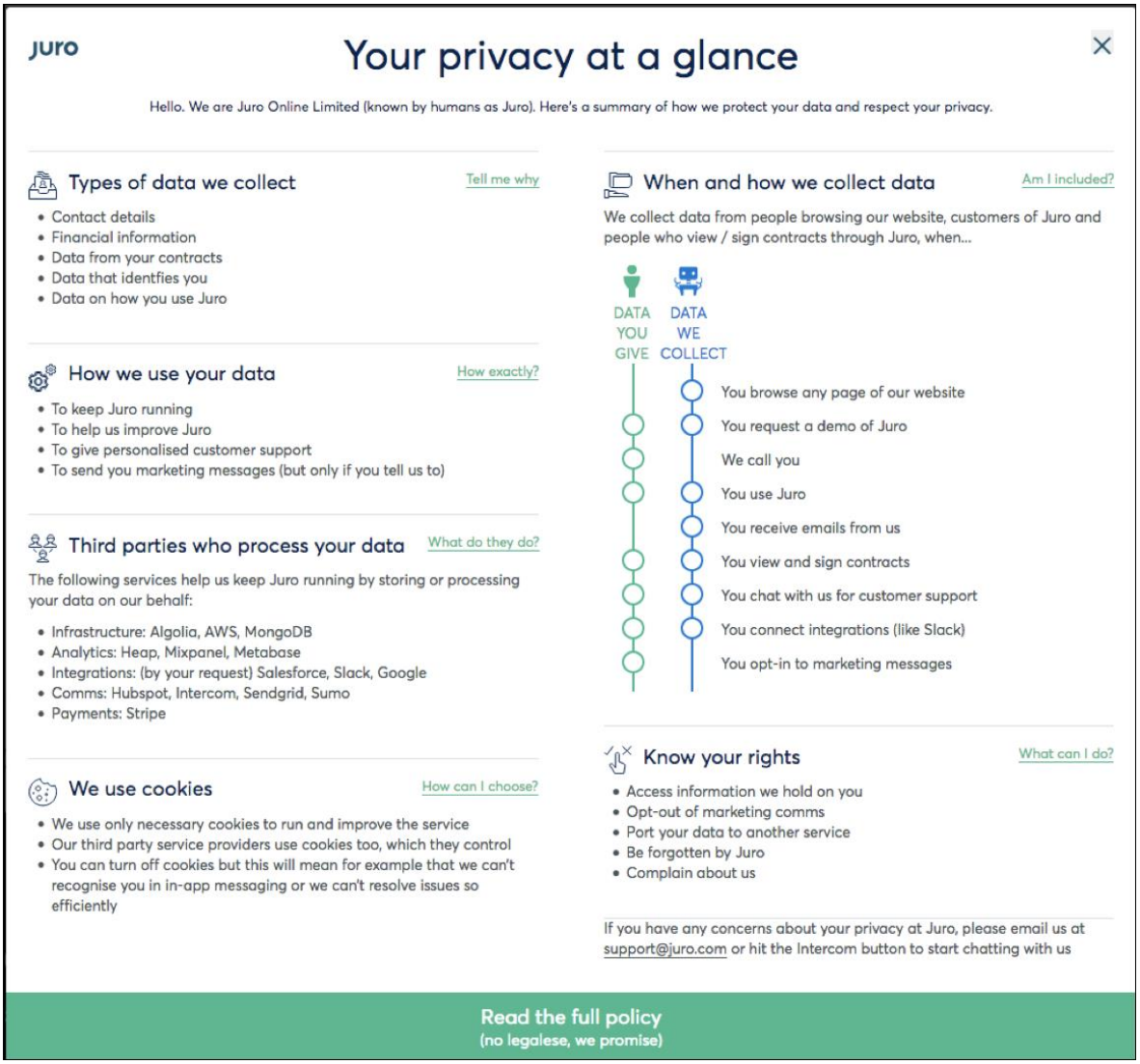

Figure 5. First layer of Juro's privacy policy, designed by Stefania Passera. Available at: https://stefaniapassera.com/portfolio/juro/

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[71] implements transparency in the process of managing

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

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/

Year of publication: 2023.

(BP2) Improving user-centered transparency in privacy policies about genetic data (re)use through contextual integrity

Author: Arianna Rossi (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

Addressees:
Privacy officers and legal teams within genetic companies and health data processors; policy makers and regulators overseeing GDPR compliance and digital health governance; UX designers and communication specialists working on consent interfaces; researchers in data ethics, privacy law, and biomedical innovation; consumer advocacy groups concerned with genetic and health data protection.

Context
Direct-to-consumer (DTC) genetic testing companies are expected to grow significantly in Europe, reaching over 2.7 billion USD by 2032. In the contex of the European Health Data Space, organizations collecting or processing genetic data will become major actors of data

sharing practices. These companies process highly sensitive personal data, which raises substantial privacy and ethical concerns.

Transparency is a cornerstone of the General Data Protection Regulation (GDPR), intended to enable individuals to understand how their data is processed and to exercise their rights meaningfully. However, privacy policies often fail to meet these goals, resulting in a misalignment between user expectations and actual data practices. The study we conducted examined the privacy and research consent policies of six leading DTC genetic companies operating in the EU, identifying 62 distinct data-sharing flows.

The analysis revealed that 81% of these flows were vague and 37% were contextually distinct and confusing, suggesting that GDPR transparency requirements may not be adequately fulfilled. Moreover, the information provided was not user-relevant and failed to address collective risks associated with genetic data sharing. To assess these issues, the study applied the theory of contextual integrity (CI), which defines privacy as the appropriateness of data transmission within specific social contexts. CI offers a structured framework to audit data flows and evaluate whether transparency obligations are met, particularly in terms of quality, accessibility, and comprehensibility of information. This approach aligns with Article 12 and Recital 39 of the GDPR, which emphasize the need for clear, plain language and user-centered communication tailored to the informational needs and abilities of the intended audience.

Definition of the challenge

The main challenge lies in the lack of clarity, completeness, and contextual relevance in the privacy and consent policies of DTC genetic companies. Information flows are often vague, bloated, or confusing, making it difficult for users to understand how their data is processed and shared. This informational opaqueness contributes to customer misinterpretation and misaligned privacy expectations, especially when risk information is insufficient or obscured. Users may consent to data sharing without fully grasping the implications, such as the difficulty of anonymizing genetic data or the potential impact on family members. This disconnect becomes evident in the aftermath of data breaches, which have led to class action lawsuits and public backlash. Furthermore, the study found that transparency requirements under GDPR were not met in many cases, with missing information about data controllers, recipients, and processing purposes. The use of vague terms and parameter bloating, such as listing multiple transmission principles in a single flow, undermines the clarity and specificity required by Article 13. While companies may use vague language to reduce the frequency of policy updates, this compromises the effectiveness of transparency.

In addition, the quality of information regarding user-relevant attributes was found to be lacking. Vague or missing details about the type of data collector and the reasons for data use can lead to confusion and further misalignment between customer expectations and reality. This is particularly problematic in consent policies, where individuals may be unaware of the nature of third-party collaborations. For example, while a consent policy mentions partnerships with academic and nonprofit institutions, it does not clearly distinguish these from for-profit collaborations. Users may assume altruistic motives behind research participation, without realizing that their data may be used for commercial drug development. This is particularly relevant in the context of data altruism. Moreover, the lack of information about whether data will be publicly available or used internally, and the absence of financial disclosures, further obscure the ethical dimensions of data sharing. Ethical review mechanisms, if transparently

communicated, could help alleviate the burden on individuals and support informed decision-making.

The challenge is compounded by the absence of communication around collective risks and harms. Genetic data inherently carries implications for relatives and descendants, yet privacy policies rarely mention these shared consequences. The risk of reidentification through genetic relatives is increasing as datasets grow. Although such practices are not yet reported in the EU, they may still affect European citizens. The lack of guidance on collective deliberation and consent is particularly concerning, given that privacy is contextual and networked. Individuals may inadvertently expose their relatives to risks, or fail to share potential benefits. While collective consent models have been explored in indigenous biomedical research, their digital implementation remains underdeveloped. The absence of collective safeguards in DTC genetic testing policies reflects a broader gap in regulatory and ethical frameworks.

Although the study focused on six companies and only analyzed publicly available privacy and consent policies, its findings are indicative of broader challenges in the governance of sensitive health data. The methodology was exploratory and expert-driven, and future work should include user studies and broader corpora to validate and extend the results. Nevertheless, the recommendations derived from this analysis are relevant not only for DTC genetic companies but for any organization processing genetic or highly sensitive health data. The issues of informational opaqueness, lack of user-relevant transparency, and absence of collective safeguards are systemic and must be addressed across sectors to ensure compliance, trust, and ethical data governance.

Proposed best practice
To address these shortcomings, privacy and consent policies should be redesigned to prioritize user-centered transparency. This involves clearly mapping data flows and ensuring that each flow is described in a way that is both specific and comprehensible to non-expert users. In addition, policies should be evaluated and structured using the framework of contextual integrity, which assesses whether data practices align with social norms and user expectations in specific contexts. Furthermore, companies should include explicit and accessible information about the risks and benefits of data sharing, including collective risks that may affect groups or communities.

Communications should be tailored to the informational needs and abilities of the intended audience, incorporating elements of information design to enhance readability and usability. This includes moving away from lengthy, jargon-heavy documents and toward empirically tested formats that support informed decision-making. Policies should also be audited regularly to ensure alignment with GDPR requirements and evolving user expectations. Algorithmic tools may assist in scanning and identifying gaps, but expert analysis remains essential to assess compliance and usability. By adopting these practices, organizations handling genetic or sensitive health data can foster greater trust, improve compliance, and support ethical innovation in digital health and biomedical research.

Constraints
Implementing user-centered transparency requires interdisciplinary collaboration between legal experts, data protection officers, UX designers, and technical teams. Moreover, organizations must balance the need for clarity with the complexity of their data ecosystems, which may involve third-party sharing, research partnerships, and evolving technologies. Regulatory

compliance must be maintained without oversimplifying the information or omitting critical details. Additionally, the contextual integrity framework must be adapted to the specific operational and cultural settings of each organization. The risks associated with genetic data sharing, such as reidentification, discrimination, and familial impact, must be communicated effectively without causing undue alarm or confusion. The lack of established models for collective notice and consent further complicates the implementation of safeguards for shared genetic privacy.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Doan X, Doğan FS and Rossi A, 'Analysis of Transparency and User-Relevancy of DTC Company Policies' in Jaap-Henk Hoepman and others (eds), *Privacy Symposium 2024* (Springer Nature Switzerland 2025). https://link.springer.com/chapter/10.1007/978-3-031-76265-9_9

Year of publication: 2025.

(BP3) Designing effective consent through multimodal communication: insights from user attitudes toward consent mediums

Author: Arianna Rossi (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

Addressees
Researchers working with health consent; legal and compliance teams in health data organizations and research institutes; policy makers developing consent standards; researchers in digital health, privacy, and human-computer interaction; UX designers and communication specialists working on consent interfaces; ethics committees and data protection authorities.

Context
As digital health data sharing becomes more prevalent, especially in the framework set by the EHDS and the DGA, the design of consent forms must evolve beyond traditional plain-text formats. The European data strategy aims to create a single market for data, built on the GDPR's foundational principle of giving individuals more control over their personal data. Informed consent (IC) under the GDPR must be freely given, specific, informed, and unambiguous, and presented in an intelligible and accessible form using clear and plain language. It must also be transparent in terms of completeness, comprehensibility, and accessibility, and compliant with data protection by design and by default. These requirements imply a user-centric approach to consent design, which includes not only plain language but also visual and structural elements that support understanding.

Despite these legal obligations, conventional privacy communications are often characterized by lengthy, jargon-heavy documents that fail to engage users or communicate key information effectively. Recent attention has turned to legal document design and multimedia formats, such as comics, videos, and infographics, as potential tools to improve transparency and user engagement. However, the effectiveness of these formats in the context of health data consent remains underexplored.

We conducted an empirical study that intended to address that gap by investigating how different consent formats are perceived by users in a health data sharing scenario. A plain-text consent form was created to simulate a request for transferring personal data from an intermediation service to a hospital. Based on this, four additional versions were designed in infographic, comic, newsletter, and video formats. All versions contained the same core content, specifically the section "What happens if you agree?", but were adapted to their respective mediums using best practices in information transparency and visual communication. The infographic used a step-by-step format with icons; the comic employed literal illustrations and story elements; the newsletter mimicked familiar email formats with more visuals than text; and the video combined color, animation, and voiceover to convey the same message. These variations were designed collaboratively and iteratively.

The study draws on human-computer interaction research to develop archetypes (i.e., general user profiles based on goals and motivations) to better understand how individuals engage with consent materials. It also builds on multimedia communication research, which suggests that dual-channel approaches and repetition across formats can enhance comprehension and retention. In a digital attention economy, where users are constantly bombarded with information, understanding what captures and retains attention in consent forms is crucial for effective communication.

Definition of the challenge

Designing consent forms that are both legally compliant and user-friendly is a complex task. Traditional text-based formats often fail to engage users or communicate key information effectively. While alternative formats such as infographics or videos offer potential, their appropriateness and impact on user understanding remain underexplored. This study highlights that user preferences for consent mediums are highly contextual and shaped by individual goals and expectations. Through qualitative interviews with 24 participants, the study identified archetypes such as the "Fully Informed," the "Record Keeper," and the "Trust Seeker," each representing different motivations and expectations in the consent process. All participants expressed a desire for high levels of understanding before making a consent decision, with some emphasizing the importance of retaining copies of their decisions and others focusing on the trustworthiness of institutions.

Participants strongly preferred short, concise, and direct consent forms, ideally no longer than one page, and often engaged in strategic reading rather than attentive reading. This underscores the need for surface-level cues such as headings, bullet points, and highlights that allow users to skim effectively and identify key information at a glance. Step-by-step organization, readability, and structure were consistently ranked as the most engaging elements, and these can be integrated across different mediums. While the infographic format incorporated several of these elements and was ranked highest overall, other formats such as text could be improved by adopting similar design principles. The tone and perceived audience fit of each medium also influenced user preferences: newsletters were associated with marketing and comics with childishness, while text was seen as routine and acceptable, though uninspiring. Rather than prioritizing one format over another, the study suggests focusing on embedding the most effective engaging elements, such as structure and step-by-step clarity, into whichever medium is used.

Proposed best practice

Consent forms should be designed to support strategic reading by incorporating surface-level cues that allow users to quickly identify and prioritize key information. This includes using clear headings, bullet points, highlights, and step-by-step organization. Regardless of the medium, text, infographic, video, comic, or newsletter, designers should ensure that the structure and readability of the content are optimized for comprehension and engagement. Archetypes can be used to tailor content to general user profiles, such as the Fully Informed or Trust Seeker, without requiring full personalization. Layering techniques should be employed to present essential information upfront while allowing access to more detailed content on request. For example, a short video or summary can accompany a full written notice, and privacy icons can be used to visually reinforce key points. Co-designing with the intended audience is essential to ensure that the chosen medium fits the context and avoids negative associations. Finally, ethical design principles should be followed to avoid manipulative elements such as overly persuasive icons or emotionally charged visuals, and to ensure that multimedia consent supports informed decision-making rather than influencing it unduly.

Constraints
The best practices proposed are based on a qualitative study involving 24 German adults and mock consent forms focused on a single section of a health data sharing scenario. While the findings offer valuable insights, they may not be generalizable across all populations, data types, or regulatory contexts. The study relied on self-reported preferences, which may differ from actual behavior, and the materials were designed by a non-professional designer, potentially influencing participant responses. Archetypes, while useful for general tailoring, require further validation and refinement through broader user studies. Implementing layered multimedia consent requires interdisciplinary collaboration and may involve increased costs and complexity. Moreover, cultural perceptions of certain mediums, such as comics, must be carefully considered to avoid undermining the seriousness of the consent process. Laslty, ethical concerns around nudging and manipulation must also be addressed, particularly when using visual or audio elements that may influence user decisions. Before deploying such consent formats, expert input should be sought to ensure compliance, usability, and contextual appropriateness.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Doan X and others, ‘Comparing Attitudes Toward Different Consent Mediums: Semistructured Qualitative Study’ (2024) 11 JMIR Human Factors e53113. Availablet at: https://humanfactors.jmir.org/2024/1/e53113

Year of publication: 2025.

(BP4) Using ISO standards to engineer privacy: Lessons from interface-level violations and design risks

Author: Arianna Rossi (LIDERLab, DIRPOLIS, Sant’Anna School of Advanced Studies, Pisa, Italy)

Addressees

Privacy engineers and software designers; compliance officers and data protection teams; UX researchers and interface auditors; ISO certification bodies and regulators; academic researchers in privacy, ethics, and human-computer interaction; organizations conducting privacy impact assessments and risk evaluations.

Context
Privacy-by-design is increasingly recognized as the standard approach in ICT system engineering. However, developers often struggle to translate high-level privacy requirements into concrete, actionable design solutions. This challenge is particularly evident when trying to align system functionalities with legal obligations such as those set out in the General Data Protection Regulation (GDPR), and international standards like ISO/IEC 29100:2011. While several privacy implementation frameworks exist, they tend to offer broad guidance without sufficient practical examples, leaving developers uncertain about whether their design choices meet compliance and standardization criteria.

The ISO/IEC 29100 Privacy Framework provides a set of eleven privacy principles—including consent and choice, openness and transparency, data minimization, and purpose legitimacy—that are intended to guide the ethical and compliant design of ICT systems. Unlike legal frameworks, ISO standards are globally applicable and align closely with industry perspectives, making them particularly useful for organizations operating across jurisdictions. However, the lack of detailed implementation guidance within ISO 29100 has limited its practical uptake. This study addresses that gap by analyzing how privacy-invasive design practices, commonly referred to as dark patterns, violate ISO privacy principles. The analysis offers a structured way to identify privacy risks and supports the development of actionable guidelines for privacy engineering.

Definition of the challenge
Most existing taxonomies of interface design patterns that are contrary to transparency principles (i.e., dark patterns) do not focus exclusively on privacy violations, nor are they structured to support systematic privacy-focused evaluations of ICT systems. Moreover, many are tailored to GDPR compliance, limiting their relevance to organizations operating under global standards. This study investigates which ISO/IEC 29100 privacy principles are most frequently violated by dark patterns, thereby offering a more universally applicable framework for identifying privacy risks.

The findings reveal that the principles most frequently violated include consent and choice, openness and transparency, and individual participation and access. These violations often stem from interface-level manipulations that obscure user choices, overload users with information, or hinder access to personal data. Other principles, such as data minimization, use limitation, and purpose legitimacy, are also at risk, though they are harder to detect without access to backend data practices. The study highlights the need for privacy engineering to move beyond abstract principles and incorporate concrete design guidance that addresses both front-end and back-end risks.

Proposed best practice
To fulfill ISO 29100, engineers should adopt a human-centered design approach that prioritizes user autonomy, clarity, and control. This involves avoiding manipulative interface designs and instead implementing privacy-enhancing patterns that support informed decision-making. The

ISO/IEC 31700 standard reinforces this approach by emphasizing usability and user experience in privacy engineering.

Research teams should integrate privacy threat modeling methods such as LINDDUN and PriS, using the mapped violations as examples of threats like data disclosure, unawareness, and non-compliance. These mappings can also support ISO certification processes and help organizations demonstrate conformity with privacy standards. Moreover, the findings can inform risk assessments under Article 24 of the GDPR, where deceptive design patterns may lead to autonomy harms, reputational damage, psychological distress, or discriminatory outcomes.
Layered transparency techniques—such as combining summaries, icons, and videos with full-text disclosures—should be used to meet accessibility and clarity requirements. Organizations should also work toward ethical design practices that extend beyond privacy engineering. Finally, the study encourages a reconceptualization of privacy-related dark patterns in terms of their violations of specific privacy requirements. This can help resolve terminological ambiguity and support the development of automated tools for detecting and mitigating deceptive designs.

Constraints
The study's findings are based on expert analysis involving a small group of participants and a taxonomy corpus finalized in mid-2022. Future research should involve a broader, more diverse group of experts and include newer taxonomies and standards, such as ISO/IEC 31700. While one-third of the analyzed patterns do not pose immediate privacy risks, they may still hinder informed decision-making and should be considered in ethical design evaluations.

The mapping focuses primarily on interface-level violations, which are more visible and easier to analyze than backend practices. However, deceptive data uses that occur behind the scenes—such as repurposing data without user awareness—also pose significant privacy risks and require more sophisticated auditing methods. Additionally, while the study provides practical guidance for privacy engineering, it does not address systemic incentives that promote manipulative design. These broader issues must be tackled through interdisciplinary collaboration and policy reform.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Valoggia P and others, 'Learning from the Dark Side About How (Not) to Engineer Privacy: Analysis of Dark Patterns Taxonomies from an ISO 29100 Perspective' (2025) <https://www.scitepress.org/Link.aspx?doi=10.5220/0012393100003648> accessed 8 September 2025.

Year of publication: 2025.

Main author: Arianna Rossi (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

Addressees:
This recommendation is directed to policymakers and technical teams within the European Commission responsible for drafting the rulebook on data altruism under the Data Governance Act (DGA). It also concerns data altruism organizations (DAOs) and researchers working with health-related personal data for scientific and public interest purposes.

Context / History of the Problem

The European Union has introduced the concept of data altruism to counteract the negative effects of data monopolies and promote the voluntary sharing of personal data for purposes of general interest. This initiative is particularly relevant in healthcare and scientific research, where access to diverse datasets can significantly improve public services and policy-making. The COVID-19 pandemic illustrated the potential of data altruism through projects like the Corona Data Donation, which enabled real-time insights into virus spread and vaccine effectiveness. However, despite the promise of such initiatives, their implementation is hindered by legal ambiguities, technical fragmentation, and trust-related concerns. The DGA seeks to address these issues by establishing DAOs and proposing standardized consent mechanisms.

Definition of the Problem

The implementation of data altruism faces several interrelated challenges. First, the DGA does not clearly define key concepts such as "scientific research" and "public interest," leading to legal uncertainty and inconsistent interpretations across organizations. This ambiguity places the burden on DAOs to assess whether data use purposes align with public interest. Second, the validity of consent is compromised by the tension between the broad nature of research and the specificity required for informed consent. Without consent obtained at the point of data collection, DAOs struggle to access diverse datasets. Third, there is no standardized digital infrastructure for managing consent, resulting in fragmented practices and potential disengagement. Existing tools vary in functionality and fail to simultaneously address identification, consent management, and data portability. These issues undermine user autonomy, transparency, and trust, and without robust solutions, the potential of data altruism in healthcare remains difficult to realize.

Proposed Policy Recommendation Aimed at Solving the Problem

To address these challenges, this recommendation proposes a multi-layered solution that integrates personalized dynamic consent, user-centric interface design, and semantic interoperability. Dynamic consent replaces static, one-time consent with ongoing, two-way communication, allowing individuals to manage their preferences over time and engage with research outcomes. This model supports autonomy, transparency, and legal compliance, while reducing administrative burdens for researchers and DAOs.

To mitigate consent fatigue and information overload, the recommendation includes tiered layered consent, which presents information in structured categories and allows individuals to choose their preferred level of detail. Personalized privacy assistants, intelligent agents that learn user preferences and automate decision-making, can further support users in navigating complex consent options without being overwhelmed.

The design of consent interfaces plays a critical role in influencing individuals' willingness to share data. Therefore, the European data altruism consent form should be developed using user-centric design principles that avoid manipulative patterns and promote informed decision-making. Tailored communication formats, such as graphics, videos, or simplified text, should accommodate diverse user profiles and literacy levels.

To ensure scalability and legal robustness, the recommendation also includes the adoption of machine-readable semantic technologies for expressing and managing consent. Semantic interoperability is essential for enabling automated, cross-sectoral data sharing while maintaining transparency and accountability. Open standards such as the W3C's Open Digital Rights Language (ODRL) and the Data Privacy Vocabulary (DPV) allow DAOs and researchers to define permissions, obligations, and restrictions. International standards like ISO/IEC 29184 and TS 27560 further support this approach by specifying how consent notices and records should be structured, documented, and communicated.

Together, these measures aim to create a unified, user-friendly, and legally compliant infrastructure for data altruism, fostering trust and enabling meaningful data sharing for public benefit.

Constraints of the Policy Recommendation

While the proposed solutions offer promising avenues for improving consent in data altruism, their implementation is subject to several constraints. First, the diversity of data sharing contexts means that a single consent model may not be universally applicable. The effectiveness of consent tools depends on factors such as the trustworthiness of the DAO, the degree of user participation in determining data use, and the presence of robust safeguards. Second, the technological maturity of proposed solutions, such as privacy assistants and semantic standards, is still evolving. Their adoption requires interdisciplinary evidence, iterative testing, and careful evaluation of usability and legal compliance.

Third, the integration of these tools must account for regulatory overlaps with the EHDS Regulation, the Open Data Directive, the Medical Devices Regulation, and the Clinical Trials Regulation. Without harmonization, there is a risk of fragmented implementation and legal uncertainty. Additionally, the success of semantic interoperability depends on consensus around vocabulary definitions and stakeholder engagement in standardization efforts. Finally, unresolved issues such as the lack of clear definitions for "scientific research" and "public interest" remain outside the scope of this recommendation and require further legal clarification. These constraints highlight the need for a cautious, evidence-based approach to designing consent experiences that empower individuals without exploiting cognitive biases.

Proposed Best Practice Aimed at Solving the Problem

For researchers, developers, and medical personnel involved in health-related data collection and processing, a practical best practice is the implementation of personalized dynamic consent platforms integrated with semantic interoperability standards. These platforms should allow participants to review and modify their consent choices over time through personalized digital interfaces. Researchers should ensure that consent options are presented in a tiered and layered format, enabling individuals to engage with information at their preferred level of detail and avoid information fatigue.

Consent interfaces should be designed using user-centric principles, offering multiple formats for information delivery, such as simplified text, graphics, videos, or voice-based communication, depending on the digital and health literacy of the target population.

Researchers should avoid manipulative design patterns and instead use ethically framed messaging that transparently communicates both the benefits and risks of data sharing.

To ensure legal compliance and interoperability, researchers should adopt machine-readable standards for documenting and managing consent. This includes using the W3C's Open Digital Rights Language (ODRL) and the Data Privacy Vocabulary (DPV). International standards such as ISO/IEC 29184 for privacy notices and ISO/IEC TS 27560 for consent record structures should be implemented to ensure traceability and accountability. These standards also support personalization and can be adapted to national legal frameworks.

By following this best practice, researchers can reduce administrative burdens, improve legal compliance, and foster trust among participants. Moreover, they contribute to a more transparent and accountable data sharing ecosystem, enabling meaningful reuse of health data for scientific and public interest purposes.

Constraints of the Best Practice

The implementation of this best practice is subject to several practical constraints. Researchers must have access to adequate technical infrastructure and expertise to deploy dynamic consent platforms and integrate semantic standards. This may require financial support, new skill acquisition, and collaboration with IT specialists or legal experts. The effectiveness of the consent experience also depends on the trustworthiness of the research institution and the clarity of communication with participants. Personalization must be carefully balanced to avoid overwhelming users or introducing bias through design.

Furthermore, researchers must ensure that their practices align with evolving regulatory frameworks, including the EHDS Regulation and other sectoral laws. The adoption of semantic technologies depends on the availability of standardized vocabularies and the willingness of institutions to harmonize their data governance practices. Despite these constraints, the best practice offers a concrete and scalable approach to improving consent management in health data altruism.

References:

For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Arianna Rossi, 'One Form to Rule Them All: Towards a Personalized, But Standardized, European Data Altruism Consent Form' (2025) in *Enabling and Safeguarding Personalized Medicine*, DSMIL vol 7, Springer, pp 49–76. Available at: https://link.springer.com/chapter/10.1007/978-3-031-99709-9_4

Year of publication: 2025

(BP6) Towards Solving Legal Ambiguities of Data Altruism Consent in the European Health Data Space

Main author: Daniela Spajic (KU Leuven, Centre for IT & IP Law, Leuven, Belgium)

Re-elaborated by: Arianna Rossi

Addressee
European Commission (DG SANTE, DG CNECT, DG JUST); National Health Ministries and Data Protection Authorities

Context / History of the Problem or Challenge
The Data Governance Act (DGA), applicable since September 2023, introduced a horizontal framework for data sharing across sectors, including healthcare. One of its key mechanisms is data altruism, which allows individuals to voluntarily share their data for objectives of general interest, such as scientific research and public health. This is facilitated through a standardized data altruism consent form and the involvement of registered data altruism organizations. The DGA lays the groundwork for sector-specific data spaces, including the European Health Data Space (EHDS), which aims to enable the re-use of electronic health data for both primary and secondary purposes. However, the DGA's reliance on consent as a legal basis intersects with the General Data Protection Regulation (GDPR), creating a complex legislative triangle. The GDPR remains the primary legal framework for data protection, and the EHDS builds upon it while introducing its own provisions for secondary data use. This layered structure raises questions about the compatibility and clarity of consent mechanisms, particularly in the sensitive context of health data.

Definition of the Problem or Challenge
The concept of data altruism consent, as introduced by the DGA, is legally ambiguous when applied to the healthcare sector. Although the DGA refers to the GDPR for its definition of consent, it does not establish a separate legal basis, instead relying on the GDPR's framework. This creates tension with the GDPR's strict requirements for consent, which must be freely given, specific, informed, and unambiguous. The DGA's approach, which allows consent for broad objectives of general interest without specifying the exact purpose at the time of data collection, challenges the GDPR's purpose limitation and data minimization principles. These ambiguities are carried into the EHDS, which also relies on data altruism consent for secondary use of health data. The EHDS regulation permits opt-out mechanisms for secondary use but allows Member States to introduce exceptions, potentially leading to fragmentation across jurisdictions. Moreover, the broad definition of general interest in the DGA, left to national interpretation, risks inconsistent application and undermines cross-border data sharing. The absence of a clear definition for general interest and the removal of dedicated provisions for data altruism in the final EHDS text further complicate the legal landscape. Additionally, the involvement of private actors in accessing sensitive health data under the guise of altruism raises ethical concerns about fairness, reciprocity, and medical confidentiality.

Proposed Best Practice
To mitigate the legal and ethical challenges posed by data altruism consent in the healthcare context, public authorities and data altruism organizations should adopt a layered consent mechanism that aligns with GDPR standards while accommodating the operational needs of the EHDS. This mechanism should ensure that consent is specific enough to meet GDPR requirements, even when broad objectives are pursued. Where possible, dynamic consent models should be explored to allow individuals to update or refine their consent over time. Data altruism organizations should implement robust transparency measures, including clear

communication about data use, safeguards, and the identity of data users. Additionally, national authorities should coordinate to develop harmonized interpretations of general interest in healthcare to support cross-border data sharing. The rulebook foreseen in Article 22 of the DGA should be leveraged to establish sector-specific criteria for data altruism in healthcare, ensuring alignment with the sensitive nature of health data and the ethical standards of medical research. Finally, data processing should prioritize anonymization or pseudonymization techniques to uphold data minimization and accuracy principles, especially when the specific use of data is not known at the time of collection.

Constraints of the Best Practice

The implementation of this best practice faces several constraints. First, the broad and undefined notion of general interest in the DGA, left to national interpretation, may lead to fragmented approaches across Member States, complicating cross-border data sharing. Second, the absence of a dedicated legal basis for data altruism consent under the DGA means that all processing must still comply with GDPR requirements, which may limit flexibility in designing consent mechanisms. Third, the sensitive nature of health data and the potential involvement of commercial actors raise ethical concerns about fairness and reciprocity, especially if data donated altruistically is used for profit. Fourth, the removal of Article 40 from the final EHDS text leaves uncertainty about the formal role of data altruism organizations in the health data space. Fifth, the compatibility clause under GDPR for further processing of personal data for scientific research lacks clarity in the context of data altruism, making it difficult to determine whether secondary use under the EHDS is legally permissible without renewed consent. Lastly, national opt-out mechanisms under the EHDS may introduce further fragmentation and require careful balancing between individual rights and public interest objectives.

References:

For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
D Spajić, 'Data Altruism Consent: A Move Forward Towards the Creation of a European Health Data Space?' in F Casarosa, F Gennari and A Rossi (eds), *Enabling and Safeguarding Personalized Medicine* (Data Science, Machine Intelligence, and Law, vol 7, Springer, Cham 2025) https://doi.org/10.1007/978-3-031-99709-9_3

Year of publication: 2025

(PR3) Rethinking Anonymization in the European Health Data Space: Legal, Ethical, and Technical Imperatives for Inclusive Data Governance

Main author: Chiara Gallese, University of Turin

Re-elaborated by: Arianna Rossi

Addresses: Policy officers and legal experts within the European Commission, particularly those working in DG CNECT, DG SANTE, and DG JUST; members of the European Parliament involved in digital and health policy; national health authorities and Health Data Access Bodies across EU Member States**;** stakeholders in EU-funded research projects dealing

with health data, AI, and digital infrastructure; civil society organizations and patient advocacy groups concerned with data protection and digital rights.

Context

The European Health Data Space (EHDS) Regulation represents a major milestone in the EU's digital strategy, aiming to facilitate both the primary and secondary use of health data across Member States. It builds upon existing legislative instruments such as the General Data Protection Regulation (GDPR), the Data Governance Act, the Data Act, and the NIS2 Directive. The EHDS seeks to enable cross-border healthcare delivery through interoperable electronic health records, create a unified digital market for health data systems, and promote the reuse of health data for research, innovation, and policymaking.

Health data under the EHDS is categorized for primary use, which includes treatment, prescriptions, diagnostics, and administrative services, and for secondary use, which encompasses research, public health, education, and AI development. The regulation introduces a decentralized infrastructure known as HealthData@EU and mandates that data access be granted through Health Data Access Bodies, which are responsible for ensuring privacy, security, and compliance. These bodies are also tasked with implementing anonymization and pseudonymization protocols, although the regulation does not provide clear guidance on how these should be carried out.

The EHDS operates within a fragmented legal landscape, relying heavily on pre-existing instruments like the GDPR without offering clear definitions for key concepts such as "anonymization." This reliance has led to interpretative inconsistencies and implementation challenges across Member States. Moreover, the regulation introduces opt-out mechanisms for patients, but these are poorly publicized and ethically problematic, especially when overridden by vaguely defined notions of public interest. Vulnerable populations often lack the means to understand or exercise their rights under the EHDS framework, and public awareness of the regulation remains limited.

The concept of anonymization is central to the EHDS, yet it is treated inconsistently across legal and technical domains. The only official EU guidance on anonymization remains the 2014 Opinion of the Article 29 Working Party, which predates the GDPR and has not been updated by the European Data Protection Board. This document outlines techniques such as randomization and generalization, including methods like k-anonymity, l-diversity, and differential privacy. However, these techniques are not uniformly applied, and their effectiveness varies depending on the context and the data involved. The EHDS defers to the GDPR for definitions but fails to provide operational clarity, leaving implementation to national discretion.

In practice, anonymization is not a binary process but a spectrum of techniques that must balance privacy protection with data utility. This is particularly relevant in health research, where overly aggressive anonymization can render datasets useless for analysis or AI training. The regulation's emphasis on anonymization as a default safeguard is undermined by the lack of a coherent framework and by the increasing feasibility of re-identification through data linkage and advanced analytics. The risks are especially pronounced in cases involving rare diseases, where the granularity of data makes full anonymization nearly impossible without losing critical information.

Definition of the Challenge

The central challenge posed by the EHDS lies in the definition, implementation, and governance of anonymization. The regulation's reliance on pre-existing legal instruments without offering a coherent conceptual framework has created a patchwork of overlapping regimes, each shaped by its own historical contingencies and normative compromises. This has resulted in a legal landscape that is difficult to interpret and navigate, leaving fundamental questions unresolved and practitioners uncertain about compliance.

Health Data Access Bodies are granted significant discretion in interpreting and applying anonymization standards, yet there is no binding guidance on the qualifications required for those performing these tasks. This opens the door to inconsistent practices and uneven levels of protection, particularly between countries with differing resources and institutional capacities. The variability in anonymization protocols can lead to significant disparities between wealthier and less wealthy Member States, potentially fragmenting the protection offered by the EHDS and undermining its overall purpose.

The tension between data utility and privacy protection remains unresolved. While high standards of anonymization may reduce the risk of re-identification, they can also compromise the analytical value of datasets, especially in fields like AI development and rare disease research. This trade-off is particularly problematic when anonymization techniques remove attributes that are essential for understanding health disparities or conducting intersectional analysis. In such cases, the anonymization process may inadvertently reinforce existing inequalities or obscure critical insights.

Anonymization decisions are inherently political and ethical. They determine who is protected and whose data is used, often reflecting power imbalances and institutional priorities. These decisions include selecting the methods used to anonymize data, choosing which datasets or individuals' data will be subject to anonymization, and deciding the extent to which the data will be modified. The ethics of anonymization extend beyond technical procedures and raise broader questions about the responsibilities of those who anonymize data. It is not sufficient to meet legal requirements; there must be a commitment to protecting vulnerable groups and ensuring that anonymization practices do not inadvertently cause harm.

The lack of stakeholder involvement further exacerbates these challenges. Patients and marginalized groups are rarely consulted in anonymization processes, despite being the most affected. The removal of certain data attributes, such as race or gender, in an attempt to protect privacy can result in the essentialization of individuals and the loss of relevant information for secondary analysis. This can lead to discrimination and harm, particularly when such attributes are correlated with specific health outcomes.

Proposed Policy Recommendations

Addressing the challenges posed by the EHDS requires a redefinition of anonymization that moves beyond a binary understanding and adopts a spectrum-based approach. Anonymization must be recognized as a political and ethical decision, not merely a technical safeguard. Legislators and policymakers must provide clear and detailed guidelines on anonymization standards and practices, including the qualifications and expertise required for those performing these tasks.

There is an urgent need to harmonize anonymization protocols across the EU to ensure consistent protection and data utility. This harmonization should be achieved through EU-wide regulations or guidelines that offer actionable instructions for Health Data Access Bodies. These guidelines must take into account the power dynamics involved in data governance and ensure that anonymization practices do not exacerbate existing inequalities.

Transparency and accountability must be central to the anonymization process. Health Data Access Bodies should be required to publish anonymization protocols and decisions, and data users must report research outcomes derived from shared data. Public reporting and oversight mechanisms can help build trust and ensure that anonymization practices are aligned with ethical standards.

Inclusive governance is essential. Patients and vulnerable communities must be engaged in the design and evaluation of anonymization practices. Intersectional frameworks should be applied to avoid essentializing individuals and erasing relevant attributes. The involvement of ethicists, legal scholars, IT experts, and social scientists can support fairness assessments and co-creation processes for secondary data use.

Cybersecurity must be strengthened to protect health data from unauthorized access and exploitation. Legislators should consider extending GDPR protections to anonymized datasets to mitigate residual risks and ensure robust data protection. This would require data users to secure datasets in accordance with GDPR standards, reducing the likelihood of breaches and misuse.

Ultimately, the protection of citizens' rights in the context of health data sharing requires a reevaluation of the definition and implementation of anonymization. Clear guidelines, harmonized practices, and inclusive governance are essential to achieving the EHDS's goals and safeguarding individuals from the risks associated with data breaches and exploitation.

Constraints

The implementation of best practices within the EHDS framework faces several structural and operational constraints. The absence of a unified definition of anonymization across EU Member States creates legal uncertainty and technical fragmentation. Health Data Access Bodies are granted significant discretion in interpreting and applying anonymization standards, yet there is no binding guidance on the qualifications required for those performing these tasks. This opens the door to inconsistent practices and uneven levels of protection, particularly between countries with differing resources and institutional capacities.

The tension between data utility and privacy protection remains unresolved. While high standards of anonymization may reduce the risk of re-identification, they can also compromise the analytical value of datasets, especially in fields like AI development and rare disease research. This trade-off is particularly problematic when anonymization techniques remove attributes that are essential for understanding health disparities or conducting intersectional analysis. In such cases, the anonymization process may inadvertently reinforce existing inequalities or obscure critical insights.

The regulatory framework itself is layered and complex. The EHDS relies heavily on pre-existing instruments such as the GDPR, without offering a coherent conceptual framework of

its own. This reliance exacerbates interpretative challenges and forces practitioners to navigate a dense legal landscape shaped by case law, doctrinal debates, and administrative guidance. The absence of clear procedural standards for anonymization and pseudonymization further complicates compliance and enforcement.

Cybersecurity remains a pressing concern. The EHDS infrastructure is likely to become a target for cyberattacks, and while the regulation references the need for secure environments, it does not mandate specific technical safeguards or breach response protocols. Smaller institutions and SMEs may lack the capacity to implement advanced security measures, leaving them, and the data subjects they serve, more vulnerable to exploitation.

Finally, there are constraints related to public awareness and democratic participation. The opt-out mechanisms provided by the EHDS are not sufficiently publicized, and vulnerable populations often lack the means to understand or exercise their rights. Without inclusive engagement and transparent governance, the regulation risks excluding those most affected by its provisions and undermining its legitimacy.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
C Gallese, 'Redefining Anonymization: Legal Challenges and Emerging Threats in the Era of the European Health Data Space' in F Casarosa, F Gennari and A Rossi (eds), *Enabling and Safeguarding Personalized Medicine* (Springer, Cham 2025) vol 7 <https://doi.org/10.1007/978-3-031-99709-9_5>

Year of publication: 2025

(PR4) Harmonising National Health Data Systems with EHDS Requirements

Author: Andrea Parziale (LIDER-Lab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy; Health Science Interdisciplinary Center, Sant'Anna School of Advanced Studies, Pisa, Italy)

Addressee
This recommendation is addressed to the Italian Ministry of Health, AGENAS (National Agency for Regional Health Services), and regional health authorities responsible for implementing and managing the Electronic Health Record (FSE) 2.0 and Health Data Ecosystem (EDS)[72].

Context / History of the Problem
The European Health Data Space (EHDS), established by Regulation 327/2025, will become progressively applicable from March 2029. This regulatory framework aims to facilitate cross-border access to health data for both primary use (healthcare delivery) and secondary use

[72] Corso, S. (2024). Il Fascicolo Sanitario Elettronico 2.0. Spunti per una lettura critica. *Nuove Leggi Civili Commentate*, *2*, 334-362.

(research, policymaking, statistics)[73]. Italy has developed parallel national systems through the FSE 2.0 (established September 2023) and EDS (established December 2024, operational March 2026). While these systems share similar objectives with EHDS, they operate under different consent mechanisms and procedural frameworks[74]. The COVID-19 pandemic highlighted the critical need for timely access to quality health data, making the harmonisation of these systems increasingly urgent as the EHDS implementation deadline approaches.

Definition of the Problem

The primary problem is a fundamental misalignment between consent requirements and data access procedures in Italy's national health data systems (FSE 2.0 and EDS) and the incoming EHDS framework.

On the one hand, the EHDS establishes an opt-out system for secondary use of health data, allowing individuals to exclude their data from research and policy applications unless overridden by public interest considerations. On the other hand, Italian law generally requires explicit consent for processing health data for research purposes, with limited exceptions.

This misalignment poses several critical risks. First, legal uncertainty can emerge when researchers and institutions must navigate conflicting consent requirements, potentially hampering research investment and cross-border collaboration. Second, administrative inefficiency can result from duplicated procedures. Technical integration challenges arise when different institutional mechanisms must coexist, creating complexity in data governance systems.

This problem affects multiple stakeholders across the health ecosystem. Researchers face legal uncertainty and administrative burden when accessing health data for scientific purposes. Healthcare institutions must implement and maintain multiple systems, increasing operational costs. Patients might experience confusion about their data rights and may need to make multiple decisions about the same data. Policymakers struggle with fragmented data access that undermines evidence-based decision-making. Industry partners encounter barriers to health data innovation and cross-border research collaboration.

The urgency of addressing this problem intensifies as the EHDS implementation timeline approaches. The EHDS does not provide clear guidance to Member States on how to ensure standardised and interoperable systems[75].

Proposed Policy Recommendation

First, to avoid misalignment in consent requirements, Italy should enact comprehensive legislative amendments that aligns FSE 2.0 and EDS mechanisms with the EHDS, particularly

[73] Spajic, D. (2025). Transforming the secondary use of patient data in the European Health Data Space: A challenge for the patient's right to medical confidentiality? In S. Slokenberga, K. Ó Cathaoir, & M. Shabani (Eds.), *The European Health Data Space: Examining a new era in data protection* (pp. 87-109). Routledge. Paul Quinn, Emma Ellyne and Chao Yao, 'Will the GDPR restrain health data access bodies under the European Health Data Space (EHDS)?' (2024) 54 Computer Law & Security Review 105993

[74] Cacini, F., & Arcuri, M. A. (2024). Uso secondario dei dati personali relativi alla salute: panoramica della normativa europea e nazionale. *Diritto dell'informazione e dell'informatica*, *1*, 837.

[75] Marelli, L., Stevens, M., Sharon, T., Van Hoyweghen, I., Boeckhout, M., Colussi, I., Degelsegger-Márquez, A., El-Sayed, S., Hoeyer, K., van Kessel, R., Zając, D. K., Matei, M., Roda, S., Prainsack, B., Schlünder, I., Shabani, M., & Southerington, T. (2023). The European health data space: Too big to succeed?. *Health policy (Amsterdam, Netherlands)*, *135*, 104861. https://doi.org/10.1016/j.healthpol.2023.104861. Rak, R. (2024). Anonymisation, pseudonymisation and secure processing environments relating to the secondary use of electronic health data in the European Health Data Space (EHDS). *European Journal of Risk Regulation*, 15(4), 928-938. https://doi.org/10.1017/err.2024.67

with its opt-out provisions. This requires amending existing decrees to establish an opt-out system for secondary use of health data across all national systems.

Secondly, to avoid institutional and procedural duplications, AGENAS, which already plays an institutional role in the EDS, could be designated as a health data access body under the EHDS. This consolidation would eliminate procedural duplication by creating a single point of contact for all secondary use data requests, whether from national or international researchers. AGENAS should develop standardised procedures that satisfy both national requirements and EHDS authorisation criteria.

Importantly, a phased implementation plan should be devised that gradually transitions from current national procedures to the aligned system as the EHDS becomes fully applicable. To this end, clear timelines should be established for legislative changes and technical implementations. Also, adequate transition periods should be provided to institutions to adapt their systems and processes while ensuring continuity of data access for ongoing research projects. Finally, training programs should be implemented for data stewards, researchers, and healthcare institutions on the aligned procedures.

Constraints of the Policy Recommendation
This recommendation is constrained by several practical limitations that must be acknowledged. The scope is limited to aligning consent mechanisms and procedural frameworks, and cannot address broader issues of data standardisation, technical infrastructure, or resource allocation for health data systems. The recommendation assumes that existing technical architectures of FSE 2.0 and EDS can be adapted to the EHDS rather than completely rebuilt.
Essential enablers include the political will of national and regional authorities to prioritise harmonisation over local system preferences. Adequate financial support is also crucial for technical system modifications, staff training, and transition management. The recommendation depends on rapid implementation before EHDS becomes applicable. Coordination between the Ministry of Health, AGENAS, and regions or autonomous provinces is essential to ensure consistent implementation across Italy's regionalised health system.

Year of publication: 2025.

(BP7) Clarifying the Definition of Scientific Research in EU Data Governance for Personalized Smart Medicine

Main Author: Ludovica Paseri, Law Department, University of Turin, Torino, Italy

Re-elaborated by: Arianna Rossi

Addressee:

Public research institutions and consortia active in personalized smart medicine in the EU; European Commission

Context / History of the Problem

Personalized smart medicine is transforming healthcare by enabling treatments tailored to individual genetic, environmental, and lifestyle factors. This approach depends on the processing of large volumes of both personal and non-personal data, often in ways that blur the boundaries between the two. Consequently, the legal frameworks governing data use, particularly in the context of scientific research, are central to enabling or constraining innovation. Within the EU, the concept of "scientific research" is defined inconsistently across multiple legal instruments, including the GDPR, EHDS, ODD, DGA, and AI Act. This fragmentation creates uncertainty for researchers, especially in public-private collaborations, and may hinder the translation of research into clinical practice. The challenge is particularly pressing in personalized smart medicine, where ethical, legal, and technological considerations converge.

Definition of the Problem

EU legislation reflects three distinct approaches to defining scientific research. First, narrow definitions found in instruments such as the Open Data Directive and Directive 2019/790 restrict the concept to publicly funded entities or those pursuing public interest missions. These frameworks offer benefits such as open access policies and exemptions for text and data mining, but exclude private actors or hybrid collaborations. This creates legal uncertainty for public-private partnerships, which are common in personalized smart medicine. Second, broad definitions in instruments like the GDPR and EHDS include public, private, and not-for-profit research. While this flexibility supports innovation, it raises ethical concerns about data use, public trust, and technological dependence. Third, some regulations such as the Data Governance Act and the Artificial Intelligence Act refer to scientific research without providing any definition or pointing to another legal source. This vagueness is problematic when exemptions or benefits are granted to research activities, as it becomes unclear which actors or projects qualify. The fragmentation of definitions, narrow, broad, and undefined, creates a complex and contradictory legal environment for researchers in personalized smart medicine. It undermines legal certainty, complicates compliance, and may hinder ethical oversight, cross-border collaboration, and the development of robust governance mechanisms.

Proposed Best Practice

To address the fragmented legal definitions of scientific research and support responsible innovation in personalized smart medicine, public research actors should adopt a structured approach based on three operational criteria and three governance mechanisms. The first criterion is FAIRness, which involves applying the FAIR data principles, findability, accessibility, interoperability, and reusability, to ensure high-quality data stewardship and facilitate compliance with frameworks like the Open Data Directive. This supports flexible, case-by-case management of research data, especially in public-private collaborations. The second criterion is accountability, drawn from Article 5(2) of the GDPR, which should guide the management of both personal and non-personal data. This principle promotes transparency, ethical conduct, and public trust in biomedical research. The third criterion is proactivity, which encourages public actors to take initiative in contexts where legal definitions are absent, such as the DGA and AI Act. By positioning themselves as data altruism users or organizations, institutions can access certified data and reduce reliance on private infrastructures, while also preparing for future commercialization and regulatory compliance.

In addition to these criteria, three governance mechanisms are proposed. First, public actors should establish clear rules of participation in research projects to promote transparency, ethical data use, and equity. This includes prioritizing approaches that consider socio-economic

heterogeneity and avoid exclusion of disadvantaged groups. Second, investment in responsible data stewardship is essential. This involves training and infrastructure to support ethical and legal data management, requiring interdisciplinary expertise in research, data curation, ethics, and law. Third, horizontal coordination among national research entities should be promoted through consortia and shared infrastructures. This helps mitigate regulatory fragmentation and supports harmonized participation in EU-level initiatives. Existing models such as the European Partnership for Personalised Medicine and the International Consortium for Personalised Medicine provide examples of how coordinated efforts can align diverse national frameworks and foster collaborative research.

Constraints of the Best Practice

The implementation of this best practice is subject to several constraints. Public actors operate within a fragmented and evolving set of EU and national regulations, which complicates consistent application of definitions and governance mechanisms. They may also face dependency on private technological infrastructures, particularly in data processing and AI development, which can limit autonomy and increase operational costs. Effective data stewardship requires significant investment in training, infrastructure, and interdisciplinary expertise, which may be challenging for smaller institutions or underfunded research centers. Finally, without deliberate inclusion strategies, personalized medicine risks reinforcing socio-economic disparities in access to treatment and participation in research, particularly affecting individuals from non-privileged backgrounds.

References:

For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
L Paseri, 'Defining Scientific Research Within the EU's Politics of Data: The Impact on Personalized Smart Medicine' in F Casarosa, F Gennari and A Rossi (eds), *Enabling and Safeguarding Personalized Medicine* (Springer, Cham 2025) vol 7, Data Science, Machine Intelligence, and Law <https://doi.org/10.1007/978-3-031-99709-9_2> accessed 6 September 2025.

Year of publication: 2025.

3.2.Artificial intelligence governance

(PR5) The principle of accountability for responsible innovation

Main author: Irina Carnat (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

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 (AI), robotics, autonomous vehicles, etc.[76]. 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[77], 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[78], 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')[79], the revised Product Liability Directive and an *ad hoc* AI Liability Directive[80]. 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*

[76] 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.
[77] 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.
[78] https://ec.europa.eu/research-and-innovation/en/horizon-magazine/opening-black-box-artificial-intelligence
[79] AI Act Proposal (n14).
[80] AI Liability Directive Proposal (n15).

forum can pose questions and pass judgement, and the actor may face consequences"[81]. Thus, 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[82]. 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[83]. 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[84].

- Algorithmic impact assessments[85]: 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[86]: 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.

[81] 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.

[82] 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.

[83] 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.

[84] 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.

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

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

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 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.[87] 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.[88]

To learn more about the topic:
Carnat, I. (2023). Ethics Lost in Translation: Trustworthy AI from Governance to Regulation. *Opinio Juris in Comparatione*, *4*. Available at: link

Year of publication: 2023.

(PR6) Redefining Algorithmic Fairness for High-Impact Automated Decision-Making

Main author: Robert Lee Poe (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

Addressee:

[87] 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.
[88] Irina Carnat, 'Ethics Lost in Translation: Trustworthy AI from governance to regulation' (pre-print 2023) 4 Opinio Juris in Comparatione 30-31.

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 questions about what should be distributed, to whom it should be distributed, and in what manner.[89] 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.[90] 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.[91]

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.

[89] Sven Ove Hansson, *Equity, Equality, And Egalitarianism*, 87 ARSP: ARCHIV FÜR RECHTS- UND SOZIALPHILOSOPHIE / ARCHIVES FOR PHILOSOPHY OF LAW AND SOCIAL PHILOSOPHY 529 (2001).
[90] JOHN RAWLS, A THEORY OF JUSTICE: ORIGINAL EDITION (1971), https://www.jstor.org/stable/j.ctvjf9z6v (last visited Nov 2, 2023).
[91] ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA (1974).

The conception of justice embodied in fair machine learning metrics and techniques is based on theories of distributive justice, characterized as egalitarian and equitable. [92]
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 unlawfulness in AI systems inherently undermines their ethical standing.[93] 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).[94] 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.[95]

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

---

[92] 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).

[93] 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.

[94] *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.

[95] 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.

positive action[96] 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.

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.[97] 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.[98] 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.[99] By

[96] 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.
[97] Poe and Mestari (n81).
[98] Alycia N. Carey & Xintao Wu, *The Statistical Fairness Field Guide: Perspectives from Social and Formal Sciences*, 3 AI ETHICS 1 (2023).
[99] Case 158/97 Badeck v Hessischer Ministerpresident [2000] ECR I-1875, § 19.

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".[100] 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, 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.[101]

Year of publication: 2023.

(PR7) Subliminal, manipulative and deceptive techniques in the context of the AI Act: new definitions proposal

Main author: Vittoria Caponecchia (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

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.[102] 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[103], the aim of which is to create a single European data space (single market for data) while leaving a central position for humans[104].

The AI Act establishes harmonised rules for artificial intelligence, with the aim, among others, of meeting the requirements of a well-functioning internal market[105], ensuring a high level of data protection, digital rights and ethical standards[106], 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

[100] Id. at §60.

[101] 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

[102] "*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.

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

[104] EU Data Strategy (n6) 4.

[105] AI Act proposal (n14).

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

rights and to facilitate the enforcement of legal rules[107].

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 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*"[108].

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[109]). 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

[107] 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.

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

[109] 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.

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[110]; Digital Markets Act - DMA[111]; Data Act[112]). In addition to these, the most important consumer protection legislation was studied (Unfair Commercial Practices Directive - UCPD[113]; and, at Italian level, Legislative Decree No. 145/2007[114] and Legislative Decree No. 146/2007[115]), 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*"[116]. 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.

[110]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).
[111]DMA (n40).
[112]Data Act (n11).
[113]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).
[114]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;
[115]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.
[116]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.

The same decree, in article 1, states that "*advertising must be clear, truthful and correct*"[117], 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"*[118] (since, if an advertisement must be transparent in order not to be considered subliminal, then it must also be clear)?;
- Assuming that "*transparent*" is equivalent to "*clear*"[119], 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[120] (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[121]) and language models with strategic reasoning (e.g. CICERO by Meta[122], strategy game based on negotiation and persuasion of opponents[123]) 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*

[117]Personal translation of art. 1, Legislative Decree 145/2007, which states: *"La pubblicità deve essere palese, veritiera e corretta".*
[118]*Ibid*.
[119]*Ibid*.
[120]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.
[121]*Ivi*, p. 1.
[122]https://ai.meta.com/research/cicero/ ; https://ai.meta.com/research/cicero/diplomacy/
[123]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.

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.

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[124]). 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*".

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

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.

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[125] and non-commercial terms (e.g. they could aim at obtaining consent and personal data, just think of online phishing).
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[126].

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[127]. 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

[125]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.
[126]Harry Brignull, *Deceptive patterns. Exposing the tricks tech companies use to control you*, Testimonium Ldt, 2023, p. 241.
[127]Il nuovo Zingarelli minore, vocabolario della lingua italiana, Zanichelli, Milano, 2008, p. 1220.

be clarified and specified by the legislator] *to one or more persons*"[128].

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 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.

Year of publication: 2023.

(BP8) Guidelines for researchers to ensure the transparency of AI systems used in bio-robotics context

[128]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.

Main author: Stefano Tramacere (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

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[129]. One of the main problems is the opacity of most state-of-the- art AI systems, *i.e., black box models*[130]. 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[131]. 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[132]. From a technical point of view, some solutions to provide greater transparency are eXplainability techniques (*XAI*)[133]. 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[134], such as feature importance scores or counterfactuals[135].

---

[129] 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.

[130] 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.

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

[132] 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.

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

[134] 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."*

[135] 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).

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[136]. Due to the opacity of the models used, these phenomena can lead to two opposing physicians' reactions: either *overreliance* or *distrust* in AI systems[137]. 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 hide incorrect inferences[138] and algorithmic discriminations[139] that could endanger the health and safety of patients[140], in violation of their fundamental rights.[141] 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[142]. From a legal perspective, opacity could interfere with the attribution of civil liability in case the

---

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

[137] 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.

[138] 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.

[139] 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.

[140] 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.

[141] 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.

[142] 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"*.

AI system's output cause harm to the patient because it is more difficult to prove the causal link[143]. Thus, the use of *black box* medical AI systems could undermine the liability of healthcare professionals by leaving injured patients unprotected[144]. In this respect, it is necessary for both those who have trained and those 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[145], thus reducing the risk of harming the end user (*e.g.*, the patient) and being held civilly liable.[146] For these reasons, the AI Act proposal[147] (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)[148].

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

---

[143] 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.*"

[144] 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 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.

[145] 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.

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

[147] AI Act proposal (n14).

[148] 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).

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[149], 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[150]. 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[151].

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 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[152]:

1. **The principle of *knowability*[153], 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)[154] and to receive meaningful information about the logic involved, so to have means/possibility to understand them (i.e., the principle of *comprehensibility*)[155] (art. 22, rec. 71 GDPR and art. 11 Annex IV (2)(b) AI Act proposal).**

[149] GDPR (n7*).*

[150] 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.

[151] 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.

[152] 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.

[153] 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).

[154] 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.

[155] 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.

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[156] AI Act proposal among other articles in the Chapter 2[157]).**

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 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*[158], 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[159], 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

[156] Article 15 is entitled *"Accuracy, robustness and cybersecurity"*.
[157] 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"*.
[158] Autorità Garante per la protezione dei dati personali, *Decalogo,* cited, points 7 and 8.
[159] Ivi, point 9.

regulatory framework once the AI Act completes its legislative process and is finally adopted in EU.

Year of publication: 2023.

(BP9) Toward best practices for using large language models in research: transparency, validation, and compliance

Author: Arianna Rossi (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

Addressees
Academic researchers; developers of AI-based research tools, with a focus on LLMs; ethics committees; institutions funding or overseeing AI research; regulatory bodies and policymakers interested in trustworthy and auditable AI systems; educators and public interest organizations working on digital literacy and consumer protection.

Context
Multimodal Large Language Models (MM-LLMs), such as GPT-4o, are increasingly integrated into research workflows to automate complex tasks involving both textual and visual data. Their ability to detect patterns, generate structured reasoning, and operate across languages makes them promising tools for decision-support systems. In domains such as deceptive pattern (DP) detection, MM-LLMs offer a scalable alternative to traditional rule-based or machine learning systems, which often lack multimodal capabilities and generalizability. The DeceptiLens study proposes an exploratory framework for using MM-LLMs to detect DPs in user interfaces, combining prompt engineering, Retrieval Augmented Generation (RAG), and expert validation.

Definition of the challenge

Traditional machine learning tools for DP detection are limited by their reliance on large, well-labeled datasets and their inability to process multimodal inputs. MM-LLMs overcome these limitations but introduce new risks, such as hallucinations, sensitivity to visual noise, and overreliance due to perceived authority. Expert disagreement on DP classification highlights the subjectivity and contextual dependence of current definitions. Moreover, static UI screenshots may be insufficient for reliable assessment, pointing to the need for richer datasets and more nuanced evaluation frameworks. Legal and ethical implications also vary depending on the intended use — whether for research, enforcement, or public education.

Proposed approach

The DeceptiLens study introduces a structured, human-in-the-loop framework for evaluating MM-LLM outputs in research. It uses GPT-4o, selected for its multilingual and multimodal capabilities, and applies Chain-of-Thought prompting and RAG to improve factual accuracy and explainability. The model is guided to produce structured reasoning that includes measurable features, step-by-step analysis, and references to

source documents. Expert evaluation was conducted in three stages: classification accuracy, explanation assessment (using clarity, correctness, completeness, and verifiability), and qualitative interviews. Results showed high recall and strong agreement when experts were unanimous, but also revealed limitations in completeness and verifiability.

Experts appreciated the transparency of the system but warned against automation bias, especially when bibliographic citations are used as authority cues. Future iterations may include cognitive forcing functions to promote user reflection, context-sensitive explanation formats, and tailored outputs for different stakeholders — from researchers to regulators and the general public. Instruction fine-tuning based on expert feedback, layered explanation structures, and enriched datasets (e.g., user journeys, HTML code) are also recommended.

Constraints
This approach inherits the limitations of MM-LLMs, including the risk of generating incorrect or biased content. The current implementation relies exclusively on UI screenshots, which may be insufficient for detecting certain DPs. The dataset is restricted in size and scope, excluding DP categories with limited reported examples. Expert input was limited to academic researchers, and the evaluation task was simplified by focusing on specific DPs rather than open-ended detection. The effectiveness of RAG depends on the relevance and representativeness of retrieved sources. Legal compliance must be context-specific and documented. Use by enforcement agencies would require safeguards, AI literacy, and clarity on the tool's role in the decision-making process. Public-facing versions would need to meet transparency obligations under the AI Act.

Reference:

For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Kocyigit E and others, 'DeceptiLens: An Approach Supporting Transparency in Deceptive Pattern Detection Based on a Multimodal Large Language Model', *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2025) <https://dl.acm.org/doi/10.1145/3715275.3732129> accessed 8 September 2025.

Year of publication: 2025.

(BP10) Best practices for the personalized and legally compliant control of robotic lower limb prostheses using AI and machine learning[160]

Authors: Ilaria Fagioli, Alessandro Mazzarini & Simona Crea (The BioRobotics Institute, Sant'Anna School of Advanced Studies; Department of Excellence in Robotics and AI,

[160] Fagioli I, Mazzarini A, Gennari F and Crea S, 'The Role of Artificial Intelligence and Machine Learning in Personalizing the Control of Robotic Lower Limb Prostheses' in Casarosa F, Gennari F and Rossi A (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) <https://doi.org/10.1007/978-3-031-99709-9_12>

Sant'Anna School of Advanced Studies, Pisa, Italy); Francesca Gennari (LIDER-Lab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

Re-elaborated by: Arianna Rossi

Addressees
Researchers and developers of robotic prostheses; regulatory experts in medical devices and AI systems; EU policymakers involved in digital health and AI regulation; clinical rehabilitation professionals; Notified Bodies and members of the Medical Devices Coordination Group (MDCG); representatives of the AI Office.

Context
Recent advances in human-in-the-loop optimization have enabled the personalization of control strategies for robotic lower limb prostheses. These approaches rely on cost functions, such as energy expenditure, gait symmetry, or cadence, to evaluate and adapt control parameters in real time. However, the optimization process is computationally intensive and time-consuming, especially when using energy-based cost functions that require prolonged walking trials. To address this, alternative cost functions with shorter adaptation periods have been explored, allowing for faster tuning and improved user experience. The integration of machine learning algorithms such as Bayesian optimization and CMA-ES has shown promise in navigating the complexity of non-convex, user-specific physiological responses, thereby enhancing the personalization of prosthetic control.

At the same time, robotic prostheses are subject to multiple legal frameworks, including the MDR, GDPR, and the AI Act. These frameworks govern both the hardware and software components of the device, with the AI Act classifying prostheses as high-risk AI systems due to their adaptive and autonomous behavior. The convergence of technical innovation and regulatory complexity makes this a critical area for coordinated action.

Definition of the challenge
The challenge lies in efficiently identifying optimal control parameters for robotic prostheses in a way that accounts for individual variability and minimizes the metabolic cost of walking. Traditional optimization methods like grid search are limited by high dimensionality and computational expense. Gradient descent, while faster, may fail in non-convex scenarios typical of human-in-the-loop systems. Machine learning-based approaches offer more robust solutions but require careful selection of cost functions and surrogate models. Moreover, the variability in user responses and the need for real-time adaptation complicate the implementation of universally effective strategies.

From a technical standpoint, one of the most pressing challenges is designing controllers that make interaction with the prosthesis natural and intuitive, while maintaining safety and reliability. This is essential for translating research into clinical applications. Human-in-the-loop optimization and machine learning methods have shown promise in tailoring control strategies to individual users, but their integration into clinical practice remains limited due to complexity and lack of standardization.

In parallel, the legal and regulatory landscape is rapidly evolving and presents significant compliance challenges. Robotic prostheses must simultaneously adhere to the Medical Device Regulation (MDR), the General Data Protection Regulation (GDPR), and the AI Act. The AI

Act classifies these systems as high-risk AI, triggering obligations such as technical documentation, human oversight, and cybersecurity. If the control algorithms are deemed to have a medical purpose, they may be classified as Software as a Medical Device (SaMD), requiring separate MDR compliance and risk classification.

The principle of complementarity in the AI Act attempts to streamline overlapping obligations, but aligning MDR and AI Act requirements, especially around quality management systems and certification, is legally complex. The lack of clear guidance from the Medical Devices Coordination Group (MDCG) and the AI Office further exacerbates uncertainty. Moreover, the classification of control algorithms depends on the manufacturer's intended purpose, which can be strategically framed to either accelerate market entry or secure long-term regulatory protection. This flexibility introduces ambiguity and risk, making legal conformity a major hurdle for innovation.

Proposed best practices

To address both the technical and legal challenges, developers should adopt modular and transparent control architectures that allow for clear separation between safety-critical and medically functional components. This facilitates regulatory classification and simplifies compliance with both MDR and AI Act requirements. Human-in-the-loop optimization strategies should be implemented using machine learning algorithms such as Bayesian optimization or CMA-ES, which enable real-time personalization of control parameters based on user-specific physiological data. These approaches must be validated through clinical trials to support MDR conformity assessments.

AI literacy and human oversight should be ensured throughout the development and deployment process. Developers, clinicians, and users must be trained to understand the system's intended purpose, risks, and operational boundaries, in line with Article 4 of the AI Act. Early engagement with regulatory bodies, including Notified Bodies and the AI Office, is essential to clarify the classification of control algorithms and to align quality management systems across MDR and AI Act frameworks. This includes preparing robust technical documentation and record-keeping protocols.

Stakeholders should actively support the development of harmonized guidance by contributing to consultations and expert groups such as the MDCG. This will help shape future frameworks for complementary compliance and reduce legal uncertainty for AI-based SaMDs. Interdisciplinary collaboration between engineers, clinicians, legal experts, and regulators is crucial to ensure that prosthetic technologies are not only technically advanced but also legally robust and clinically applicable.

Constraints

The MDR and AI Act are still undergoing phased implementation across EU Member States, which may result in inconsistent interpretation and enforcement. Further, clinical validation of machine learning-based control strategies is time-consuming and resource-intensive. In addition, manufacturers must balance innovation with regulatory risk, especially when deciding whether to classify control algorithms as SaMD. Moreover, there is currently no harmonized EU guidance on complementary compliance between MDR and AI Act obligations. Lastly, the classification of AI systems depends on the manufacturer's declared intended purpose, which may be influenced by strategic market positioning.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Fagioli I, Mazzarini A, Gennari F and Crea S, 'The Role of Artificial Intelligence and Machine Learning in Personalizing the Control of Robotic Lower Limb Prostheses' in Casarosa F, Gennari F and Rossi A (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) <https://doi.org/10.1007/978-3-031-99709-9_12>

Year of publication: 2025.

(BP11) Clarifying the definition of AI systems under the AI Act: Towards a shared interdisciplinary vocabulary

Authors: Arianna Rossi, Francesca Gennari, Denise Amram, Andrea Parziale (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy); Ilaria Fagioli, Alessandro Mazzarini, Fabrizio Moncelli, Simona Crea (WRLab, The BioRobotics Institute, Sant'Anna School of Advanced Studies, Viale Rinaldo Piaggio 34, 56025 Pontedera, Italy; Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna, Pisa, Italy)

Re-elaborated by: Arianna Rossi

Addressees
Policy makers drafting or revising AI legislation; legal experts interpreting the AI Act; technical experts developing AI systems; standardization bodies; compliance officers in regulated sectors; interdisciplinary research teams working on AI governance.

Context
The AI Act introduces a definition of "AI system" in Article 3(1), further elaborated in Recital 12 and the accompanying Guidelines. Although the Guidelines exclude certain systems, such as rule-based systems and optimization techniques, interdisciplinary analysis reveals persistent terminological and interpretative ambiguities. These issues risk undermining legal certainty and compliance, particularly in domains like healthcare and robotics, where AI systems are increasingly deployed and where regulatory clarity is essential. As soft law instruments, the Guidelines issued by the AI Office are intended to support responsible and accountable innovation by offering reliable criteria for compliance and oversight. However, the lack of a coherent interdisciplinary method for interpreting the AI Act and its Guidelines complicates this task. This work contributes to bridging that gap by proposing a comparative lexicon and interpretative framework that can be used by developers, legal scholars, regulators, and policy-makers. It is particularly relevant for high-risk applications such as medical devices, where classification under the AI Act has significant regulatory consequences.

Definition of the challenge
Despite the clarifications provided in the Guidelines, several semantic and conceptual issues remain unresolved. The notion of autonomy is vague and fails to account for varying degrees of independence across system components and stakeholder perspectives. Moreover, self-learning systems are incorrectly equated with adaptive systems, overlooking their distinct phases of operation. The distinction between machine learning and logic- or knowledge-based

approaches is often blurred in practice, and terms such as “inference” are used inconsistently, where “learning” or “training” would be more appropriate. In addition, pattern recognition is mischaracterized as rule-based, despite its data-driven nature. The term “performance” lacks a clear definition, even though it is central to determining exemptions. Furthermore, vague descriptors such as “simple”, “basic”, and “traditional” lack standard meaning in technical communities, making it difficult to apply the Guidelines consistently.

These inconsistencies hinder the operationalization of the AI Act and complicate the assessment of whether a system qualifies as an AI system or falls under an exception. From an operational point of view, it remains unclear whether it is more effective to first determine if a system meets the definition of an AI system and then assess whether it falls under an exception, or to begin by evaluating potential exceptions. The team leans toward the first approach, as the Guidelines define what is not considered an AI system, which is conceptually different from affirming that a machine-based system is not an AI system. This distinction is particularly important in the case of robotic prostheses, which may be classified as high-risk AI systems under Article 6(1) and Annex I(11) if they are considered medical devices.

Proposed best practices

To address these challenges, it is essential to develop a shared interdisciplinary vocabulary that clearly defines key terms used in the AI Act and its Guidelines, drawing from both legal and technical domains. In addition, the definition of autonomy should be clarified to distinguish between system components and stakeholder perspectives, such as those of developers and users. The classification of adaptive systems should be refined to ensure that self-learning capabilities are correctly contextualized within system lifecycle phases. Moreover, terminology for machine learning processes should be standardized, replacing ambiguous terms like “inference” with more precise alternatives. The examples used in the Guidelines, such as pattern recognition, should be reassessed to ensure they reflect current technical realities. Further, the concept of “performance” should be defined in a way that encompasses both system behavior and outcomes, enabling clearer exemption criteria. Non-standard descriptors should be avoided in favor of terminology recognized across engineering and computer science disciplines.

Additionally, the method of interpretation should combine a common vocabulary with a careful analysis of the grammar and logical structure of the legal text. This interoperable interpretation framework should be tested and refined with more complex use cases to support generalization and pre-standardization efforts. These efforts aim not only to address ethical and legal compliance within a given R&D lifecycle but also to contribute to the development of interoperable tools of interpretation for a fragmented and evolving legal framework. From a compliance perspective, this work is a necessary precondition for an accountable, future-proof approach to technological innovation.

Constraints

The interdisciplinary nature of AI regulation requires consensus across legal, technical, and policy communities. Moreover, definitions must remain flexible enough to accommodate evolving technologies while ensuring legal clarity. The vocabulary must also be applicable across diverse use cases and deployment contexts. Furthermore, coordination between the AI Act and other regulations, such as the Medical Devices Regulation (EU Regulation 2017/745), must avoid overburdening developers and ensure a streamlined compliance process. The consequences of misclassification are significant, as systems deemed high-risk must comply with both regulatory frameworks, even if developed within research settings.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Arianna Rossi et al, 'The AI system definition under the AI Act, a new nomen rosae?' (in press) in Davide dall'Anna, Gizem Gezici and Giulio Rossetti (eds), *Proceedings of HHAI-WS 2025: Workshops at the Fourth International Conference on Hybrid Human-Artificial Intelligence (HHAI), Pisa, Italy, 9–13 June 2025.*

This policy recommendation has also been submitted to the European Commission's call for evidence on "A European Strategy for AI in science – paving the way for a European AI research council" in 2025 (more details are provided in D7.7).[161]

Year of publication: 2025.

3.3.Regulation of medical devices and health law

(PR8) Uncertainty and Slowdown in the MDR Regulatory Process and the lack of Notified Bodies

Main author: Georgios Christou (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

Addressee:

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

Medical Device Regulation in Context:

The regulation of Medical Devices was initially regulated by three directives, the Medical Devices Directive (MDD) 93/42/EEC,[162] Active Implantable Medical Devices Directive (AIMDD) 90/385/EEC, and the In Vitro Diagnostic Medical Devices Directive 98/79/EC[163]. 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

[161] Available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14547-A-European-Strategy-for-AI-in-science-paving-the-way-for-a-European-AI-research-council/F3564147_en
[162] Council Directive 93/42/EEC of 14 June 1993 concerning medical devices, OJ L 169, 12.7.1993, p. 1–43
[163] 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.

scandal had broken out, including a case of premature rupture of both implants in the same patient[164]. In light of this scandal, the EU introduced the IVDR[165] as well as the EU Medical Device Regulation (EU MDR)[166] to try and prevent such a tragedy from happening again[167]. 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[168].

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[169]. 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"[170]. 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[171]", 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 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[172].

Recommendations:

The issue was partially addressed already by the European Commission through the proposal 2023/0005 (COD) [173], 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

[164] 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
[165] Regulation on in vitro diagnostic medical devices (n13).
[166] Medical Devices Regulation (n12).
[167] Laura Maher, Niki Price, 'Ultimate Guide to IVDR for In Vitro Diagnostic Medical Device Companies', November 2022, Greenlight Guru.
[168] 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
[169] 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.
[170] Ibid.
[171] 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.
[172] Ibid.
[173] 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.

MDR by the end of the transition period[174]". 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)[175]". 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[176], 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[177]. 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[178], or following MedTech's suggestions such as speeding up the certification procedure.

Year of publication: 2023.

(BP12-PR9) Personalized Medicine in the Age of Complexity: Reintegrating Empathy, Evidence, and Innovation in Clinical Practice

Authors: Michele Emdin, Claudio Passino (Health Science Interdisciplinary Center, Sant'Anna School of Advanced Studies, Pisa, Italy; Fondazione Toscana Gabriele Monasterio, Pisa, Italy)

Re-elaborated by: Arianna Rossi

Addresses: This policy recommendation is addressed to, European Commission officials working in DG SANTE, DG CNECT, and DG RTD; National health ministries and public

[174] Ibid. p1.
[175] Ibid, pp7-8.
[176] 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.
[177] Ibid.
[178] MDCG Position Paper, "Transition to the MDR and IVDR: Notified body capacity and availability of medical devices and IVDs", August 2022, MDCG 2022-14.

health agencies; Medical education and training institutions; Healthcare providers and hospital administrators; Research funding bodies and ethics committees; Patient advocacy groups and civil society organizations.

Context

Personalized medicine has emerged as a transformative paradigm in healthcare, promising to tailor prevention, diagnosis, and treatment to the unique biological and experiential profile of each patient. Rooted in the historical ethos of Hippocratic medicine and enriched by contemporary advances in genomics, systems biology, and information technology, personalized medicine seeks to reconcile scientific precision with humanistic care. The evolution from empirical therapies to evidence-based guidelines has been further refined by translational research and biomarker-driven interventions, enabling more targeted and effective treatments.

The concept of personalized medicine, also referred to as precision or P4 medicine (predictive, preventive, personalized, and participatory), is grounded in the integration of molecular profiling, environmental factors, and patient engagement. It has demonstrated clinical success in oncology, infectious diseases, and chronic conditions, with examples such as HER2-targeted therapies in breast cancer and pharmacogenetic screening for HIV treatments. The increasing use of genome-wide association studies (GWAS) and systems pharmacology has expanded the scope of personalized interventions, while digital infrastructures such as electronic health records and biomedical informatics grids have facilitated data sharing and clinical decision-making.

Despite its promise, personalized medicine remains unevenly implemented across healthcare systems. While developed countries have begun integrating it into policy frameworks, resource-limited settings face significant barriers. Moreover, the complexity of the individual phenotype, encompassing genetic, physiological, emotional, and social dimensions, requires a holistic approach that combines advanced diagnostics with empathy and narrative understanding. The challenge is not only technological but also relational: the therapeutic alliance between patient and provider must be preserved and strengthened in the face of increasing digitalization.

Definition of the challenge

The implementation of personalized medicine raises several interrelated challenges. First, the sustainability of personalized prevention and care is uncertain. While precision interventions may reduce long-term healthcare costs, they also risk generating inefficiencies through overtesting, overdiagnosis, overtreatment, and overcharging. Without clear guidelines and cost-effectiveness assessments, these practices may strain healthcare systems and divert resources from essential services.

Second, the scientific foundations of personalized medicine are still evolving. Pharmacogenetic data remain biased toward specific populations, and the genotype-phenotype relationship is far from fully understood. Environmental and lifestyle factors further complicate treatment responses, requiring multidisciplinary research and inclusive data collection. The lack of comprehensive support from governments and healthcare organizations, particularly in developing countries, exacerbates these limitations.

Third, ethical, legal, and social concerns must be addressed. Stratifying patients by genetic or ethnic markers risks reinforcing social segregation and misunderstanding among the public. The denial of treatment based on genetic classification, if not carefully communicated and

justified, may undermine trust in healthcare systems. Regulatory disparities in drug approval and distribution also contribute to unequal access and delayed implementation.
Finally, the increasing reliance on artificial intelligence and digital tools in clinical decision-making must not come at the expense of human judgment and empathy. The complexity of patient care cannot be fully captured by algorithms alone. Physicians must continue to exercise clinical reasoning, consider emotional and cognitive fragilities, and maintain a compassionate presence throughout the care continuum.

Proposed best practice and policy recommendation
To ensure the responsible and effective integration of personalized medicine into healthcare systems, a multifaceted strategy is required. First, policymakers should promote a balanced approach that combines scientific rigor with humanistic care. Personalized medicine must not be reduced to a purely technical exercise; it should be grounded in the relational and ethical dimensions of clinical practice. Medical education should reinforce the importance of empathy, narrative competence, and interdisciplinary collaboration.

Second, regulatory frameworks must be updated to support equitable access and sustainability. This includes developing evidence-based standards for testing, diagnosis, and treatment, as well as mechanisms for evaluating cost-effectiveness and minimizing unnecessary interventions. Health information technologies should be leveraged to facilitate data sharing and protect patient privacy, with robust cybersecurity measures and transparent governance.

Third, research efforts must be expanded and diversified. Funding should prioritize inclusive studies that reflect the genetic and environmental diversity of global populations. Multimodal data sources, including -omics, imaging, and patient-reported outcomes, should be integrated to refine disease models and therapeutic strategies. Ethical oversight must ensure that stratification does not lead to discrimination or exclusion.

Fourth, stakeholder engagement is essential. Patients, caregivers, and advocacy groups should be involved in shaping personalized medicine policies and practices. Their insights can inform the design of interventions that are not only clinically effective but also socially acceptable and culturally sensitive. Public awareness campaigns should clarify the benefits and limitations of personalized approaches, fostering informed consent and shared decision-making.

Finally, the therapeutic alliance must remain central. Personalized medicine should enhance, not replace, the relationship between patients and providers. Early diagnosis and appropriate treatment must be pursued through a synthesis of advanced scientific tools and the Hippocratic tradition of care. Artificial intelligence can support clinical reasoning, but it must be guided by the wisdom and empathy of healthcare professionals working in teams.

Constraints
The successful implementation of personalized medicine is constrained by several factors. Scientific limitations, including incomplete pharmacogenetic data and complex genotype-phenotype interactions, hinder the development of universally applicable interventions. Economic disparities between countries and within healthcare systems affect the availability and affordability of personalized treatments. Regulatory fragmentation and inconsistent approval processes delay access and create inequities.
Ethical concerns around genetic stratification and treatment denial require careful navigation. Public misunderstanding and mistrust may arise if personalized medicine is perceived as

exclusionary or opaque. The integration of digital tools, while beneficial, introduces risks related to data security, algorithmic bias, and the erosion of human judgment.
Institutional inertia and limited interdisciplinary collaboration further impede progress. Without coordinated efforts among policymakers, researchers, clinicians, and patients, personalized medicine may remain a niche innovation rather than a systemic transformation. The challenge is to align technological advancement with ethical responsibility and social inclusivity, ensuring that personalized care is not only scientifically sound but also humanely delivered.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Emdin M and Passino C, 'Personalized Medicine: A Medical Perspective' in Federica Casarosa, Francesca Gennari and Arianna Rossi (eds), *Enabling and Safeguarding Personalized Medicine* (Springer Nature Switzerland 2025) <https://doi.org/10.1007/978-3-031-99709-9_7> accessed 8 September 2025

Year of publication: 2025.

(BP13) Adapting Health Technology Assessment Frameworks for Digital Health: Towards Value-Based Personalized Care

Authors: Ciro Pappalardo & Floriana d'Ambrosio (Section of Hygiene, University Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Rome, Italy); Giovanna Elisa Calabrò (Department of Human Sciences, Society and Health, University of Cassino and Southern Lazio, Cassino, Italy and Value in Health Technology and Academy for Leadership & Innovation, Spin-Off of Università Cattolica del Sacro Cuore, Rome, Italy

Re-elaborated by: Arianna Rossi

Addressees:
This brief is addressed to European Commission policymakers; national and regional HTA agencies; regulatory bodies overseeing digital health and AI; healthcare providers and hospital administrators; developers and manufacturers of digital health technologies; patient advocacy groups and civil society organizations; academic and research institutions involved in HTA methodology development.

Context
Healthcare systems are increasingly shifting towards value-based models that prioritize personalized care and efficient resource allocation. This transformation is guided by paradigms such as Porter's definition of value and the European Commission's four-pillar framework, personal, allocative, technical, and societal value. Digital health technologies (DHTs), including telemedicine and wearable devices, are emerging as key enablers of this shift, offering solutions to challenges like aging populations, rising costs, and access disparities. However, their responsible and cost-effective implementation requires adherence to principles such as transparency, privacy, and scalability. Health technology assessment (HTA), as a multidisciplinary and lifecycle-based evaluation tool, is positioned to support this integration. Frameworks like EUnetHTA and INAHTA provide structured methodologies to assess clinical,

economic, ethical, and social dimensions of health technologies, ensuring that DHTs contribute meaningfully to sustainable, value-driven healthcare systems.

Despite their foundational role, current HTA frameworks struggle to evaluate DHTs effectively. These technologies evolve rapidly, often requiring iterative and dynamic assessments that traditional models cannot accommodate. Ethical, privacy, and cybersecurity concerns are frequently underrepresented, and the lack of universally accepted methodologies further complicates evaluations. Moreover, the digital divide, cultural resistance, and infrastructural limitations pose significant barriers to equitable access and adoption. Organizational challenges such as workflow disruption, data quality, and sustainability also impact the successful integration of DHTs. Regulatory frameworks like the GDPR, AI Act, EHDS, and NIS2 introduce new compliance requirements that HTAs must consider. Finally, the absence of clear liability guidelines for digital technologies raises concerns about accountability and trust. These multifaceted challenges underscore the urgent need to adapt HTA frameworks to the realities of digital health.

Definition of the challenge
Traditional HTA frameworks are not designed to accommodate the dynamic nature of digital health technologies. The rapid evolution of software, the need for continuous updates, and the integration of real-world data challenge static evaluation models. Ethical dimensions such as privacy, transparency, algorithmic fairness, and patient autonomy are often inadequately addressed. The digital divide and infrastructural disparities risk exacerbating health inequalities, while cultural resistance and low digital literacy hinder adoption. Organizationally, DHTs disrupt workflows, require new competencies, and demand robust data governance. Environmental sustainability is another emerging concern, with DHTs contributing to energy consumption and resource use. Regulatory compliance with GDPR, AI Act, EHDS, and NIS2 adds complexity, and the lack of clear liability frameworks for digital errors undermines trust. These challenges collectively reveal the limitations of current HTA models and the necessity for more flexible, inclusive, and multidisciplinary approaches.

Proposed best practice
HTA bodies should adopt dynamic evaluation models that allow for continuous reassessment of DHTs throughout their lifecycle. These models must integrate real-world data and evidence to capture the actual performance, safety, and value of technologies in diverse contexts. A risk-based approach should be implemented, aligning the depth of evaluation with the potential impact of the technology, thereby optimizing resource allocation and accelerating safe adoption. Outcome measures should be expanded to include patient-reported data, such as PROMIS, to reflect holistic health and wellbeing. HTA processes must involve multidisciplinary expertise, including cybersecurity, health informatics, user-centered design, and environmental science, to ensure comprehensive and context-sensitive evaluations. Design methodologies should be iterative and participatory, incorporating human factors and ergonomics to enhance usability and safety. These practices will enable HTAs to better support the integration of DHTs into healthcare systems and promote value-based, patient-centered care.

Best practices
To effectively evaluate digital health technologies (DHTs), health technology assessment (HTA) bodies should adopt dynamic models that allow for cyclical reassessments and frequent updates throughout the technology's lifecycle. These models must be capable of integrating

real-world data (RWD) and real-world evidence (RWE), enabling evaluators to monitor clinical effectiveness, safety, and patient adherence in real-world contexts. A risk-based evaluation approach should also be implemented, inspired by frameworks such as the EU AI Act and the NHS Evidence Standards Framework, to ensure that the depth and complexity of the assessment are proportionate to the potential clinical and societal impact of the DHT. Furthermore, HTA methodologies should incorporate patient-centered outcome measures, such as those provided by PROMIS, to capture physical, mental, and social wellbeing and enhance the relevance of evaluations for both patients and clinicians. The assessment process should be enriched through multidisciplinary collaboration, involving experts in cybersecurity, interoperability, user-centered design, ergonomics, and environmental sustainability. These professionals can contribute to more nuanced evaluations by addressing technical vulnerabilities, data integration challenges, usability, and ecological impact. Iterative design methods and co-creation with stakeholders should be embedded into HTA practices to ensure that digital solutions are aligned with real-world healthcare needs and user expectations.

Policy recommendations

Policymakers should support the reform of HTA frameworks to make them adaptive, modular, and responsive to the evolving nature of digital health technologies. This includes promoting the development of validated European models capable of continuous and context-sensitive evaluation. Real-world evidence should be systematically incorporated into regulatory and decision-making processes, including pre-market assessments, to facilitate responsible innovation and reduce barriers for small and medium-sized enterprises. Harmonization of HTA practices across EU member states is essential, and should be pursued through the establishment of shared terminology, open repositories of methodologies, and coordinated guidelines to enhance interoperability and transferability. HTA protocols must also integrate compliance with key regulatory instruments such as the GDPR, AI Act, EHDS, and NIS2 Directive, ensuring robust data protection, transparency, and cybersecurity. Equity and inclusivity should be central to HTA policy, with explicit consideration of the impact of DHTs on vulnerable populations, infrastructural sustainability, and data representativeness. Finally, a multi-stakeholder governance model should be institutionalized, involving patients, clinicians, developers, regulators, and civil society actors in the design and implementation of HTA frameworks. This collaborative approach will foster trust, accountability, and innovation, enabling digital health technologies to be safely and effectively integrated into value-based, personalized healthcare systems.

Constraints

The implementation of dynamic HTA frameworks and the integration of real-world evidence face several limitations. Data quality, representativeness, and methodological standardization remain unresolved challenges, particularly when relying on heterogeneous or unsupervised data sources. Privacy and security concerns are heightened in digital environments, where anonymization techniques may be vulnerable to re-identification, and consent mechanisms often lack transparency and user control. The absence of universally accepted methods for incorporating environmental sustainability into HTA evaluations further complicates efforts to assess the ecological impact of DHTs. Organizational constraints, including limited digital infrastructure, interoperability issues, and the need for continuous training, may hinder the adoption of new assessment models. Additionally, the lack of clear legal frameworks for assigning liability in cases of digital errors or adverse events undermines trust and slows uptake. Finally, cultural resistance, low digital literacy, and insufficient inclusion of marginalized

populations in evaluation processes risk perpetuating existing health disparities if not adequately addressed..

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
C Pappalardo, F d'Ambrosio and GE Calabrò, 'Navigating the Increasing Complexity of Health Technology Assessments in the Digital Era: How to Support Value-Based Personalized Healthcare?' in F Casarosa, F Gennari and A Rossi (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) <https://doi.org/10.1007/978-3-031-99709-9_8>

Year of publication: 2025.

(BP14) Health Technology Assessment for Personalized Medicine: Addressing Economic and Organizational Complexity

Authors: Leopoldo Trieste & Giuseppe Turchetti (Institute of Management, Sant'Anna School of Advanced Studies, Pisa, Italy)

Re-elaborated by: Arianna Rossi

Addressees
This best practice is addressed to health technology assessment experts; hospital managers and clinical directors; policymakers and regulatory authorities; pharmaceutical and medical device companies; digital health platform developers; healthcare professionals involved in personalized care delivery; patient advocacy groups; and payers and reimbursement bodies. These stakeholders must collaborate to ensure that personalized medicine is evaluated and implemented as a system-level transformation, not merely as a set of discrete technological tools. They must also align around shared goals to overcome economic, organizational, and cultural barriers to sustainable adoption.

Context
Personalized medicine is reshaping healthcare by tailoring treatments and interventions to individual genetic, environmental, and lifestyle factors. This transformation affects a wide range of stakeholders, including healthcare providers, technology developers, payers, and policymakers. The shift from standardized approaches to personalized ones introduces new technological, economic, and organizational dynamics. As healthcare systems begin to adopt personalized medicine, it becomes essential to understand its implications not only for clinical outcomes but also for system-wide sustainability, stakeholder alignment, and the transformation of healthcare delivery models.

Definition of the challenge
Evaluating and implementing personalized medical strategies presents multifaceted challenges across methodological, organizational, technological, and industrial domains. Traditional frameworks such as pharmaco-economics and health technology assessment (HTA) struggle to accommodate the individualized nature of personalized medicine. These tools are often ill-equipped to assess interventions tailored to specific genetic profiles or disease subtypes,

limiting their applicability and relevance. From an industry perspective, personalized therapies are reshaping drug development and service delivery. While targeted treatments and narrowed clinical trials may reduce costs and accelerate development, they also demand new business models and regulatory pathways. The rise of modular services, delivered via apps, robotics, and gamified platforms, expands the scope of health technologies but complicates their integration into existing evaluation and reimbursement systems. Organizationally, personalized care pathways increase service complexity.

The shift from standardized to individualized care risks placing patients at the center of fragmented and inefficient systems. Designing sustainable, patient-centered pathways requires integrating clinical, organizational, economic, and technological dimensions within HTA frameworks. Technologically, the integration of big data, AI, and digital platforms offers opportunities to enhance access and efficiency. However, there is a risk that technological solutions prioritize efficiency over clinical effectiveness and interpersonal care. Over-reliance on ICT platforms may distance patients from providers and lead to solutions that address technological possibilities rather than actual healthcare needs. Ultimately, the challenge lies in aligning innovation with real-world needs, ensuring that personalized medicine enhances care quality and system sustainability without increasing fragmentation or complexity.

Proposed best practices

To effectively evaluate and implement personalized medicine, health technology assessments (HTAs) must evolve to reflect the complexity, stakeholder diversity, and dynamic nature of personalized healthcare services. Assessments should go beyond clinical efficacy to include organizational, economic, usability, acceptability, and ethical dimensions. Evaluation should incorporate tailored key performance indicators for each stakeholder group (patients, healthcare professionals, hospitals, diagnostic companies, and technology providers) and these indicators must be updated when medical devices are involved. Static HTA approaches are insufficient for technologies that exhibit increasing returns or depend on economies of scale; simulation-based and real-time monitoring methods should be used to assess sustainability and cost-effectiveness as patient volumes and usage patterns evolve.

A checklist-based approach should be used to identify technologies, services, personalization elements, and stakeholder roles. For each stakeholder, their objectives, perceived value, willingness to pay, and the factors that drive or limit their actions should be defined. Managers should promote open communication, shared goals, and transparent decision-making. Neutral mediation and structured frameworks such as weighted matrices can help resolve conflicts and align stakeholders around common objectives. HTA processes must integrate advanced statistical and computational methods to handle high-dimensional data, longitudinal studies, and heterogeneous datasets. Technologies should be designed around recognized healthcare needs rather than technological possibilities to ensure relevance, usability, and alignment with clinical and organizational goals.

Constraints

The evaluation and implementation of personalized medicine within a health technology assessment (HTA) framework face several constraints. Standard economic models used in HTA rely on assumptions such as resource scarcity, fixed technologies, diminishing returns, and patient independence. These assumptions do not hold in the context of personalized medicine, especially when digital platforms and AI-driven solutions exhibit increasing returns and cross-side network effects. Personalized medicine often involves modular services delivered through

digital platforms, which complicate organizational analysis and require tailored evaluation frameworks. These services also introduce new stakeholder dynamics that must be accounted for. High-dimensional data, small sample sizes, multiple testing, population stratification, and longitudinal data pose significant challenges for traditional statistical methods. These issues reduce the reliability and generalizability of findings and necessitate advanced analytical techniques and high-performance computing.

The sustainability of personalized technologies often depends on economies of scale. Static HTA models are insufficient for evaluating these solutions in real time. Simulation-based and continuous monitoring approaches are needed to assess cost-effectiveness as patient volumes change. Key performance indicators vary widely across stakeholders and must be adapted when medical devices are involved. This fragmentation complicates the synthesis of economic and organizational impacts. Economic impacts include direct health and non-health costs, as well as indirect costs like productivity losses. These vary significantly depending on the stakeholder and the type of personalized solution, making comprehensive cost assessment challenging. The use of sensitive genetic and health data in personalized medicine raises concerns about data protection. Statistical methods must incorporate anonymization strategies without compromising analytical validity. The difficulty in pricing and reimbursing personalized treatments due to the need to demonstrate both clinical and economic efficiency is a constraint that affects market access. Successful implementation requires investments in infrastructure, training, and IT systems, which may not be readily available. Finally, acceptability, usability, ethical, and legal barriers can reduce the effectiveness of personalized medicine even when technologies are clinically sound.

References

For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:

Trieste L and Turchetti G, 'The Evaluative, Economic, and Organizational Dimensions of Personalized Medicine' in Francesca Casarosa, Federica Gennari and Arianna Rossi (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) <https://doi.org/10.1007/978-3-031-99709-9_9>

Year of publication: 2025.

(BP15) Empowering patients and caregivers in the digital transformation of healthcare: Building competencies for inclusive and sustainable personalized medicine in Europe[179]

Authors: Sabrina Grigolo & Milena Sirtori (Patient Expert EUPATI, Rome, Italy); Renza Barbon Galluppi (Associazione Rete Malattie Rare, Venice, Italy)

Re-elaborated by: Arianna Rossi

[179] Grigolo S, Galluppi RB and Sirtori M, 'New Perspectives in Research and Development for Patients and Caregivers: The Challenges of Digital Health Competencies in Europe for Personalized Medicine' in Francesca Casarosa, Federica Gennari and Arianna Rossi (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) <https://doi.org/10.1007/978-3-031-99709-9_10>

Addressees
This best practice is addressed to national and European policymakers; ministries of health and education; digital health platform developers; telemedicine service providers; hospital managers and clinical directors; patient advocacy groups; training institutions; and healthcare professionals. These stakeholders must collaborate to promote inclusive digital health literacy, integrate patients and caregivers into innovation processes, and ensure that digital health technologies are designed and implemented as part of a broader cultural and systemic transformation.

Context
The European Health Data Space (EHDS) is a strategic initiative designed to promote the cross-border sharing of health data, enabling more personalized diagnoses, innovative treatments, and the integration of digital health technologies. However, the success of this transformation depends not only on robust legal safeguards to ensure data security, but also on the digital competencies of those managing and using these technologies, patients, caregivers, and healthcare professionals. The Digital Competence Framework for Citizens (DigComp) provides a shared reference across the EU for defining digital skills, and its relevance to health contexts is increasingly recognized.

Italy, despite its economic and demographic weight, has untapped potential to contribute to the EU's Digital Decade targets. While recent investments under the Resilience and Recovery Plan (PNRR) have strengthened digital infrastructure, Italy continues to lag behind the EU average in digital skills and the digitalization of public services. Only 46% of the population possess basic digital skills, and the country faces shortages in IT graduates and gender disparities in the digital workforce. These gaps undermine citizens' ability to benefit from digital opportunities and exercise digital citizenship, particularly in healthcare.

Patients and caregivers are no longer passive recipients of care. They increasingly participate in healthcare decisions and administer therapies, expressing their right to self-determination not only as individuals to be protected, but as active agents of change. This shift demands new competencies, especially digital ones, for all involved, medical professionals, patients, caregivers, and citizens. The digital transformation in health requires a cultural shift that empowers all stakeholders through inclusive digital health literacy. Digital skills for health are not optional; they are essential for improving quality of life and ensuring equitable access to personalized care. In this spirit, the chapter calls for the creation of a Digital Health Literacy Space at both national and European levels, building on initiatives such as Italy's "All Digital Weeks."

Definition of the challenge
The digital transformation of healthcare introduces critical challenges in personalized medicine, particularly regarding the digital competencies of patients and caregivers. While the EHDS aims to facilitate data-driven innovation, its implementation risks excluding those who lack the necessary digital skills to engage with digital therapies and health platforms. The absence of adequate training and support can lead to increased cyber and physical risks, reduced usability of digital tools, and inequitable access to personalized care.

Italy's slow progress in digital skill development, especially among older adults and underrepresented groups, further exacerbates these challenges. The limited availability of training programs, low numbers of IT graduates, and gender disparities in the digital workforce

hinder the country's ability to meet its own healthcare digitization goals and contribute meaningfully to EU targets.

Moreover, the evolving role of patients and caregivers as co-decision-makers and care administrators requires a redefinition of healthcare relationships. Their experiential knowledge must be integrated with the scientific expertise of professionals to co-create effective therapies and care pathways. The lack of a shared lexicon across disciplines, fragmented governance, and insufficient organizational empowerment further complicate the adoption of digital health solutions. Without a cultural shift that empowers all stakeholders, including families, practitioners, and policymakers, digital health risks becoming a source of fragmentation rather than inclusion.

Proposed best practices

Digital health literacy must be recognized as a permanent structural requirement for a participatory health democracy. National and European strategies should promote inclusive training programs for patients, caregivers, and healthcare professionals, aligned with the DigComp 2.2 framework and supported by initiatives such as the National Training Plan in Digital Health and the Patient Expert in Digital Technologies for Health course. These programs should include accessible formats such as video tutorials, infographics, and coaching, and be integrated into telemedicine platforms.

Patients should be involved as partners in the design and development of digital health technologies. Participatory approaches that include patients in co-design processes from the outset can improve usability, relevance, and adoption. The Patient Learning Pathway (PLP) framework offers a flexible model for organizing competencies across the patient's life and integrating them into organizational processes.

Digital caregivers and support services should be designed with attention to personalization, human interaction, and ethical considerations. Policymakers should explore funding mechanisms to support digital services for informal caregivers, including subsidies and incentives. Video-based group education and disease-specific digital resources should be expanded and evaluated through high-quality research.

Patient-reported outcomes must be systematically integrated into performance measures to ensure that healthcare services reflect patient experiences and values. Governments should support user-friendly service interfaces and guide design flows to assist older users, while promoting digital health technologies as socially and culturally accepted tools for improving quality of life.

Constraints of the best practices

The lack of digital skills among patients, caregivers, and healthcare professionals remains a major barrier to the adoption of personalized digital health solutions. Italy's performance in digital literacy is below the EU average, with significant disparities across age groups and socioeconomic backgrounds. These gaps limit access to digital services and undermine the inclusiveness of healthcare systems.

Healthcare technologies are often developed through top-down, techno-centric approaches that exclude patients from meaningful participation. Without appropriate training paths, patients are rarely recognized as contributors to innovation. The absence of a shared data culture,

fragmented governance, and limited organizational empowerment further hinder the integration of digital health solutions.

Older adults face specific barriers to adopting e-health services, including low awareness of technological benefits and limited digital readiness. Their adoption is influenced more by social and environmental factors than by internal technological ones. The current design of e-health services often fails to accommodate these needs, requiring modifications to models such as UTAUT to improve accessibility and usability.

The efficacy and adoption of public e-health services remain uncertain. Without targeted interventions to improve equity, access, and digital competence, vulnerable populations risk being left behind. The digital divide, lack of interoperability, and insufficient integration of experiential knowledge into care models continue to challenge the sustainability and effectiveness of digital health systems.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Grigolo S, Galluppi RB and Sirtori M, 'New Perspectives in Research and Development for Patients and Caregivers: The Challenges of Digital Health Competencies in Europe for Personalized Medicine' in Francesca Casarosa, Federica Gennari and Arianna Rossi (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) https://doi.org/10.1007/978-3-031-99709-9_10

Year of publication: 2025.

(BP16) Best practices for enabling the sustainable and inclusive adoption of robotics in rehabilitation[180]

Authors: Irene Giovanna Aprile, Marco Germanotta, Maria Cristina Mauro & Alessio Fasano (IRCCS Fondazione Don Carlo Gnocchi ONLUS, Florence, Italy)

Re-elaborated by: Arianna Rossi

Addressees
Healthcare professionals and rehabilitation therapists; hospital administrators and clinical decision-makers; researchers and developers of robotic rehabilitation technologies; policymakers in health innovation and digital transformation; educators and training institutions in health technology; national health authorities and reimbursement bodies.

Context

[180] Aprile IG, Germanotta M, Mauro MC and Fasano A, 'Bridging the Gap: Overcoming the Barriers to Using Robotics in Rehabilitation' in Casarosa F, Gennari F and Rossi A (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) <https://doi.org/10.1007/978-3-031-99709-9_13>

Robotics and digital technologies are increasingly recognized as transformative tools in rehabilitation, offering advanced and personalized treatments for individuals with motor, sensory, and cognitive impairments. These technologies have evolved significantly over the past decades, with growing scientific evidence supporting their efficacy. Robotic systems can increase the intensity and volume of therapy, standardize treatment protocols, and provide real-time sensory feedback that enhances brain plasticity and patient engagement. Digital platforms such as virtual and augmented reality allow patients to practice daily tasks in immersive environments, while wearable sensors and AI-driven analytics enable highly personalized rehabilitation plans that adapt to individual progress.

Recent innovations have also enabled home-based rehabilitation, allowing patients to continue therapy remotely with real-time feedback and monitoring. This continuity of care is particularly valuable during chronic phases of disability and contributes to long-term adherence and recovery. Despite these advances, the integration of robotics into clinical settings remains limited. Barriers such as practitioner resistance, economic constraints, and ethical and legal concerns persist. The rapid introduction of these technologies has not been matched by necessary adjustments in healthcare systems, particularly in organizational models, professional training, and regulatory frameworks. This chapter critically examines the potential of robotic and digital therapies, the barriers to their implementation, and strategies to overcome these challenges in rehabilitation and healthcare.

Definition of the challenge

The main challenge is to ensure that robotic technologies in rehabilitation are not only technically effective but also clinically viable, economically sustainable, and equitably accessible. Despite their therapeutic potential, several barriers continue to limit their integration into clinical settings. Economically, the high costs of acquisition, installation, maintenance, and training pose a significant burden for healthcare facilities. Reimbursement mechanisms are often unclear or absent, and even where robotic rehabilitation is formally recognized within national healthcare frameworks, the absence of specific guidelines and tariff structures creates uncertainty and limits institutional investment.

Resistance from healthcare practitioners continues to hinder adoption, particularly when training is insufficient or when technologies are perceived as overly complex or disruptive to established practices. Without structured professional development and leadership support, these attitudes may persist, slowing the pace of innovation.

Regulatory uncertainty adds another layer of complexity. Guidelines for the use and reimbursement of robotic systems are fragmented and inconsistent across regions and care settings. Accreditation requirements may vary, and economic policies are not always based on robust analysis. Ethical and privacy concerns further complicate implementation, especially in relation to data protection and patient consent. Compliance with regulations such as the GDPR requires robust security measures, transparent consent procedures, and ongoing training for healthcare personnel. Without institutional support and clear legal frameworks, organizations may struggle to navigate these obligations.

Together, these barriers form a multifaceted challenge that must be addressed through coordinated efforts across clinical, educational, regulatory, and policy domains.

Proposed best practices

To overcome the multifaceted barriers to integrating robotics and digital technologies into rehabilitation, a coordinated and evidence-based approach is required. From an economic standpoint, healthcare systems should adopt innovative organizational models that demonstrate both clinical effectiveness and cost-efficiency. Studies have shown that robot-assisted rehabilitation can be delivered sustainably through models where a single physiotherapist supervises multiple patients using robotic devices, without compromising treatment outcomes or patient satisfaction. These models should be supported by pragmatic clinical trials and cost-utility analyses to provide decision-makers with robust data on long-term value and resource optimization. Dedicated interdisciplinary teams should be established within healthcare facilities to manage the implementation of robotic systems, provide ongoing training, and ensure seamless integration into clinical workflows.

To address practitioner resistance, comprehensive training programs must be developed and embedded into professional education and continuing development pathways. These programs should cover both technical and clinical aspects of robotic technologies, enabling practitioners to confidently operate devices and incorporate them into personalized treatment plans. Leadership within healthcare organizations must actively promote the adoption of these technologies by highlighting evidence-based benefits and involving frontline staff in decision-making processes. Creating a culture of innovation and collaboration is essential to fostering acceptance and enthusiasm among clinical teams.

Regulatory frameworks must be strengthened to ensure equitable access to robotic rehabilitation. These frameworks should be informed by clinical trial data and cost-effectiveness analyses, and should guarantee that patients who may benefit from robotic therapies are entitled to use them and receive reimbursement. Integrating robotics into national healthcare subsidy systems with clear reimbursement policies will help standardize access and promote sustainability.

To address ethical and privacy concerns, comprehensive guidelines must be developed for the use of robotic and digital technologies in clinical settings. These guidelines should ensure full compliance with data protection regulations such as the GDPR, including robust encryption, transparent consent procedures, and regular security audits. Given the dynamic nature of robotic systems, regulatory frameworks must also accommodate frequent updates and modifications through fast-track approval mechanisms. Collaboration between regulatory bodies and healthcare institutions is essential to ensure that evolving technologies are safely and effectively integrated into practice, while maintaining public trust and safeguarding patient rights.

Constraints of the best practices

The integration of robotics and digital technologies into rehabilitation is constrained by several interrelated factors. Financial limitations remain a major obstacle, as the high costs of equipment, maintenance, and training are not always offset by existing reimbursement models. Even where robotic rehabilitation is formally recognized within national healthcare frameworks, the absence of specific tariff structures and economic guidelines creates uncertainty and limits institutional investment.

Resistance from healthcare practitioners continues to hinder adoption, particularly when training is insufficient or when technologies are perceived as overly complex or disruptive to established practices. Without structured professional development and leadership support, these attitudes may persist, slowing the pace of innovation.

Regulatory frameworks are often fragmented and lack the flexibility needed to accommodate rapidly evolving technologies. The absence of fast-track procedures for approving updates and modifications to robotic systems can delay their deployment and reduce their responsiveness to clinical needs. Ethical and privacy concerns further complicate implementation, especially in relation to data protection and patient consent. Compliance with regulations such as the GDPR requires robust institutional systems and ongoing staff training, which may be difficult to maintain without dedicated resources.

Finally, the lack of coordinated national strategies and stakeholder collaboration risks perpetuating disparities in access and undermines the potential for robotics to contribute meaningfully to the democratization of rehabilitation services.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Aprile IG, Germanotta M, Mauro MC and Fasano A, 'Bridging the Gap: Overcoming the Barriers to Using Robotics in Rehabilitation' in Casarosa F, Gennari F and Rossi A (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) https://doi.org/10.1007/978-3-031-99709-9_13

Year of publication: 2025.

3.4.Liability and product safety

(PR10) Changing the draft Article 7 of the new Product Liability Directive Update. A few suggestions

Main author: Francesca Gennari (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

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)[181], 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[182]). The Court of Justice had to

[181] 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.
[182] See the case C-127/04. *Declan O'Byrne v Sanofi Pasteur MSD Ltd and Sanofi Pasteur SA* ECLI:EU:C:2006:9.

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)[183], 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[184], 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 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*"[185]. 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[186].

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[187]. 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.

[183] Product Liability Directive Proposal (n19).
[184] PLDU Explanatory Memorandum (within the proposal see footnote 2) p. 2.
[185] Article 7(5) PLDU
[186] Article 7(6) PLDU.
[187] Article 14(1) PLDU.

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[188]. 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 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[189]. 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

[188] 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>.
[189] Francesca Gennari, "A tale of two cities? Fennia v Philips and Article 7 of the Product Liability Directive Update", Forthcoming EuCML December issue 2003

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[190]. 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 private international law. Specifically, there should be an explicit reference to Article 5 of the Rome II regulation[191], 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.

References

For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:

[190] 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.

[191] 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.

Gennari, F. (2023). A Tale of Two Cities? Fennia v Philips and Article 7 of the Product Liability Directive Update. *Journal of European Consumer and Market Law*, *12*(6), 267–274.

Year of publication: 2023.

(PR11) The Notion of Manufacturer's Control in the new PLD. The design implications for advanced technological manufacturers

Author: Francesca Gennari (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

Addressee:
Manufacturers and designers of advanced medical devices, such as Biorobotic prostheses for upper and lower limbs

Context/history of the problem
This is partly an old and a new problem as far as EU product liability rules are concerned. As far as the old rules the old Product Liability Directive, PLD, (COUNCIL DIRECTIVE of 25 July 1985 on the Approximation of the Laws, Regulations and Administrative Provisions of the Member States Concerning Liability for Defective Products (85/374/EEC), 1985) it was important to establish the liability of producer (now called manufacturer) especially in complex value chains (Gennari, 2023). As far as the new problems are concerned, they are caused mainly the interconnection of software within consumer products and the inclusion of AI (interconnected or standalone) in the new definition of product at Article 4(1) of the new PLD (Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on Liability for Defective Products and Repealing Council Directive 85/374/EEC (Text with EEA Relevance) PE/7/2024/REV/1 OJ L, 2024/2853, 18.11.2024, n.d.). Sometimes software is created by the manufacturer (former producer), but other times it is third parties that provide either standalone or interconnecting software.

Definition of the problem
Because of Article 5 of the New PLD that states that consumer must find a way to recover damages in the EU, irrespective of the complexity of technological objects. This includes also bioprosthesis as the new PLD considers software, in all its forms, a product. According to Article 8 the person who is responsible for the malfunctioning and damage caused by the product is the manufacturer. In order to simplify recovering damages for consumer a new concept, called manufacturer's control is now defined at Article 4(5) NPLD. This concept means " *(a) the manufacturer of a product performs or, with regard to actions of a third party,* ***authorises or consents*** *to: (i)* ***the integration, inter-connection or supply of a component, including software updates or upgrades****; or (ii) the modification of the product, including substantial modifications; (b) the manufacturer of a product has the ability to supply software updates or upgrades, themselves or via a third party*."

The importance of this concept is also testified by the fact that it precedes the definition of manufacturer at Article 4(10) NPLD. This means that if a third-party software concurs or causes a kind of damage that is compensated under the NPLD (see article 6 NPLD), then the consumer can sue the manufacturer of the final product even if they did not create the software nor had any business in updating or modifying it. This could be a problem as a small manufacturer

oftentimes cannot address all the software issues and can be more cost-effective to delegate someone else to have software to be interconnected. With the notion of manufacturer's control problems connected to software that is of a third party but interconnected to the hardaware product will be a source of liability. This is in principle a good thing for consumers but for small manufacturers of advanced medical devices it can be a huge source of liability and a disincentive to innovation. It is true that Article 14 NPLD states that each Member State must actually implement their own discipline of their right to recourse. For instance, if I am a manufacturer in control of third-party's faulty software then the consumer is entitled to ask compensation to me according to the PLDU. Then if the cause of damage was the faulty software, Member States allow me as a manufacturer to sue the third party software developer. Nevertheless, this second passage is not harmonized for all the EU and it can become difficult for a manufacturer to recover compensation from the actor who actually caused the defect of their product across the EU.

Proposed policy recommendation aimed at solving the problem

The suggestions could be several

- On a case by case basis, to evaluate whether to create software in-house or thanks to assess the risk of the damage that could be created by the use of the software.
- If the risk that something might happen is high and the damage could be serious, the manufacturer might want to retain a better control on software and decide that that creating software can be more demanding but safer for them to control
- Otherwise, they might decide to delegate the creation of the software to a third party
- In the latter case, it is mandatory to pre-vet a series of alternatives and prefer companies that are certified for cybersecurity, privacy risk management and other related issues. In the future it might be possible to have AI systems certifications. Hence, if an AI system is needed for the interconnection within a product, those can be indicators of trustworthiness.
- Within the contract setting the basis for the collaboration between the manufacturer and the third party, a clause might be reserved for the manufacturer to send second party auditors to the third party at regular intervals, to be cartain that the level of quality and safety of the software is monitored and checked regularly
- Within the contract setting the basis for the collaboration between the manufacturer and the third party, it must be written that the code produced must respect the coding good practices and other controls of Annex I of standard ISO 27001/2022 or the Annex of standard ISO 42001/2023 concerning the Information technology , Artificial intelligence , Management system

Constraints of the policy recommendation

This are recommendations that are general and do not take into consideration real life cases as the NPLD will be applicable from 9 December 2026. Apart from very general ISO standards there are not yet any EU harmonized standards specialized for AI systems quality and safety. These will need to be considered as soon as they will be published in the EU official journal.

References

For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Gennari, F. (2023). A Tale of Two Cities? Fennia v Philips and Article 7 of the Product Liability Directive Update. *Journal of European Consumer and Market Law*, *12*(6), 267–274.

Year of publication: 2025.

(PR12) Robotics and biorobotics in the law of personal injuries compensation and rehabilitation

Main author: Maria Gagliardi (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

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:
There is uncertainty about the application of existing rules[192] 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.

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

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.

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.

Year of publication: 2023.

(BP17) Best practices for managing emerging risks and liability in AI-powered medical devices under the revised Product Liability Directive

Author: Andrea Parziale (LIDER-Lab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy; Health Science Interdisciplinary Center, Sant'Anna School of Advanced Studies, Pisa, Italy)

Re-elaborated by: Arianna Rossi

Addressees

Legal experts in health technology and product liability; policymakers involved in AI and medical device regulation; manufacturers and developers of AI-powered medical devices; compliance officers and risk managers in digital health companies; scholars in tort law and regulatory innovation; representatives of EU institutions working on the implementation of the AI Act and the revised Product Liability Directive.

Context

The integration of artificial intelligence into medical devices is transforming healthcare by enhancing diagnostic precision, treatment personalization, and operational efficiency. AI-powered systems are increasingly assisting physicians in clinical decision-making, shifting the traditional boundaries of human-led healthcare. These technologies leverage vast datasets, including genomic, biochemical, physiological, and behavioral information, to tailor interventions to individual patients. Innovations such as pharmacogenomic assays, wearable biosensors, and implantable devices equipped with autonomous AI software are enabling real-time monitoring and adaptive treatment strategies across a wide range of conditions.

AI techniques, including machine learning and neural networks, are essential for identifying clinically relevant patterns in complex datasets and delivering actionable insights. As a result, AI-powered medical devices are becoming central to personalized medicine, offering new paradigms for prevention, diagnosis, and therapy. However, the dynamic and evolving nature of these systems raises significant ethical and legal concerns, particularly regarding patient safety and accountability.

To address these concerns, Regulation (EU) 2017/745 (Medical Device Regulation, MDR) and Regulation (EU) 2024/1689 (Artificial Intelligence Act, AI Act) establish an ex-ante regulatory framework aimed at ensuring that only safe and effective AI-powered medical devices are placed on the market. The MDR defines medical devices broadly, including software intended for medical purposes, and applies even when software functions as an accessory to a physical device. It requires manufacturers to implement a lifecycle-based risk management system, document foreseeable hazards, and ensure that residual risks are acceptable and clearly communicated. Software-specific provisions mandate repeatability, reliability, and performance aligned with intended use, including considerations for mobile platforms.

Classification under the MDR determines the level of regulatory scrutiny and post-market obligations. Rule 11 of Annex VIII assigns software to Class IIa, IIb, or III depending on the

severity of potential harm resulting from its use. Software that drives or influences a device inherits the classification of the device itself. According to MDCG 2019-11 guidance, AI-powered devices for personalized medicine are likely to fall into medium-to-high risk categories, requiring certification by a Notified Body. This triggers a series of post-market obligations including Clinical Evaluation Reports, Post-market Surveillance Plans, Periodic Safety Update Reports, Post-market Clinical Follow-up Plans and Reports, and Summaries of Safety and Clinical Performance.

Under the AI Act, AI-powered medical devices are likely to be classified as high-risk AI systems. These systems must undergo third-party conformity assessment and comply with a comprehensive set of obligations, including the establishment of a continuous and iterative risk management system throughout the lifecycle of the AI system. Providers must also prepare unified technical documentation demonstrating compliance with both the AI Act and MDR, ensure transparency in system outputs, and maintain a quality management system that includes testing, validation, and incident reporting procedures.

Directive (EU) 2024/2853 (Revised Product Liability Directive, RPLD) complements these safety regulations by clarifying that software, including autonomous AI systems, is a product subject to liability. The chapter assesses how the RPLD interacts with existing frameworks and whether it provides meaningful incentives for manufacturers to monitor and address risks in AI-powered medical devices. Through a hypothetical scenario involving an initially safe device that later causes harm, the analysis explores the limitations of the ‘later defect’ and ‘development risk’ exemptions and considers whether strict liability should be applied.

Definition of the challenge

The central challenge lies in ensuring that manufacturers of AI-powered medical devices are incentivized to proactively manage risks that emerge after deployment, while also providing fair compensation to patients harmed by unforeseen failures. The RPLD introduces important clarifications, including the explicit recognition of software as a product subject to liability. However, the burden remains on the claimant to prove defectiveness, damage, and causality. The definition of defectiveness now includes a broader set of circumstances, such as the product’s ability to learn after deployment and the specific needs of its intended user group. While this expansion is welcome, it introduces ambiguity. It is unclear whether the ability to self-learn should raise or lower safety expectations, especially when such learning may lead to errors.

Although the RPLD limits the applicability of the ‘later defect’ exemption in the case of software-related failures, the continued availability of the ‘development risk’ defense introduces a problematic loophole. Manufacturers may attempt to argue that a defect, though manifesting later, was present from the outset but undiscoverable due to the state of scientific knowledge. This line of reasoning conflates the existence of a defect with its cause and may incentivize manufacturers not to fully investigate or expand the state of the art, in order to preserve the option of invoking this defense. The situation is particularly concerning for the first victims of emerging risks, who may face greater difficulty in obtaining redress. This raises the question of whether Member States should consider derogating from the development risk defense under Article 18 of the RPLD, especially for high-risk AI-powered medical devices.

Proposed best practice

To address the legal and technical challenges posed by AI-powered medical devices, the RPLD should be interpreted and implemented in a way that reinforces safety expectations and strengthens accountability. The ability of a product to learn after deployment must be treated as a factor that increases, not decreases, its expected safety. Manufacturers should be required to design AI systems that minimize the risk of harmful learning outcomes and to maintain robust post-market surveillance mechanisms capable of detecting and correcting emerging risks.

Legal frameworks should discourage vague disclaimers and ensure that product instructions do not serve as a shield against liability. Instead, they should promote transparency and clarity, especially when devices are used in high-stakes clinical environments. The RPLD's provisions for easing the burden of proof, such as presumptions of causality and consequences for withholding evidence, should be actively applied to support claimants facing technical or scientific barriers.

The exemptions under Article 11 of the RPLD must be narrowly interpreted. The 'later defect' exemption should not apply to software-related failures, and the 'development risk' defense should not be used to avoid liability for defects that emerge through post-deployment learning. Member States should consider using their discretion under Article 18 to derogate from the development risk defense for high-risk AI-powered medical devices, particularly where public interest and patient safety are at stake. Additionally, Member States may explore the introduction of strict liability regimes for these products, either by derogation or through separate contractual or extracontractual frameworks. Models such as Article 5:101 of the Principles of European Tort Law or Article 2050 of the Italian Civil Code offer useful reference points for designing liability systems that reflect the inherent risks of autonomous AI systems. The EU should promote comparative legal research and stakeholder engagement to refine liability models and ensure that the RPLD delivers consistent and fair outcomes across Member States. This includes exploring the potential for strict liability in cases involving autonomous AI systems, where the complexity of the technology may otherwise prevent injured parties from obtaining redress.

Constraints

The implementation of effective liability frameworks for AI-powered medical devices is constrained by legal uncertainty, technological complexity, and the coexistence of multiple liability regimes across Member States. The ambiguity surrounding the interpretation of defectiveness in learning systems may weaken safety expectations and complicate the evidentiary burden for claimants. While the RPLD excludes the later defect exemption for software, the continued availability of the development risk defense may incentivize manufacturers to avoid expanding the state of the art, thereby undermining proactive safety efforts.

The lack of harmonized guidance on how to apply these exemptions in practice creates uneven legal protection across jurisdictions. Moreover, the coexistence of fault-based, product, and strict liability regimes within the EU adds complexity to enforcement and may result in inconsistent outcomes for injured parties. The absence of a unified approach to strict liability

for high-risk AI systems limits the ability of legal frameworks to fully address the risks posed by autonomous learning technologies.

Finally, the need for empirical and comparative legal research remains urgent. Without robust data on the socio-economic impact of different liability models, policymakers may struggle to design effective reforms. The European legal framework offers a promising testbed for such research, but further coordination and investment are needed to translate legal theory into actionable policy.

Year of publication: 2025.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Parziale A, 'AI-powered Medical Devices and the Development Risk Defense Under the Revised Product Liability Directive' in Casarosa F, Gennari F and Rossi A (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) <https://doi.org/10.1007/978-3-031-99709-9_14>

(BP18) Designing transparent and accountable AI systems to support clinical decision-making and liability standards

Authors: Andrea Blatti (LIDER-Lab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy) & Stefano Tramacere (LIDER-Lab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy; Department of Computer Science, University of Pisa, Pisa, Italy)

Re-elaborated by: Arianna Rossi

Addressees
Healthcare technology developers; AI system providers; regulatory bodies responsible for medical devices and AI compliance; hospital administrators and clinical governance teams; legal experts in health law and liability; policymakers involved in AI and health regulation; medical educators and professional associations; patient advocacy groups; researchers conducting AI-based clinical trials; and ethics committees overseeing the deployment of AI in healthcare settings.

Context
AI systems in healthcare are increasingly recognized for their predictive accuracy and potential to enhance medical performance. Their capacity to operate autonomously and adaptively, as defined in Article 3(1) of the EU AI Act, introduces transformative dynamics into clinical decision-making. These systems analyze vast datasets to support diagnoses and treatment recommendations, often surpassing traditional diagnostic tools. This shift challenges the conventional doctor-patient relationship and introduces new actors, such as AI developers and providers, into the clinical workflow. The novelty of AI lies in its ability to perform tasks once reserved for human expertise, relying on statistical correlations rather than causal reasoning. As a result, AI is accelerating the move toward personalized medicine while raising complex ethical, legal, and social questions, particularly concerning transparency, accountability, and the protection of fundamental rights.

Transparency plays a strategic role in this transformation. In high-risk domains like medicine, opacity in AI systems can create new forms of information asymmetry, undermining the constitutional right to free and informed consent. The AI Act and national case law increasingly recognize the need for algorithmic transparency to support patient autonomy and legal accountability. Moreover, the integration of AI into clinical workflows is reshaping the legal concept of the standard of care, which traditionally relies on evidence-based medicine, professional guidelines, and consensus within the medical community. This evolution raises questions about how AI-generated recommendations fit within existing liability frameworks and whether they redefine what constitutes negligence or fault.

Definition of the challenge

The deployment of AI-based medical systems presents two interrelated challenges: transparency and the evolving standard of care in medical liability.

First, many AI systems operate as black boxes, making it difficult for deployers, such as physicians and healthcare facilities, to understand how diagnoses or treatment recommendations are generated. This opacity limits traceability, complicates reliability assessments, and hinders the detection of errors or biases. It also raises concerns about automation bias, translational bias, and the legal attribution of fault in cases of harm. Without clear insight into how an AI system reaches its conclusions, it becomes difficult to establish causation, an essential element in liability regimes.

Second, the integration of AI into clinical decision-making may alter the definition of the standard of care. Traditionally, this standard is derived from medical guidelines, best practices, and evidence-based medicine. Evidence-based frameworks prioritize decisions grounded in high-quality clinical research, such as randomized controlled trials, systematic reviews, and meta-analyses, over anecdotal or opinion-based approaches. These hierarchies of evidence are designed to minimize bias and confounding, thereby enhancing transparency and reliability.

However, AI systems often rely on statistical correlations and data-driven proxies that may not align with established clinical evidence. Even when AI systems demonstrate superior diagnostic accuracy, their black-box nature prevents full inspection of their decision-making processes. This raises a legal dilemma: should physicians be held liable for not using a more accurate AI system, even if its reasoning is opaque? Or does the lack of explainability mean that such systems cannot yet redefine the standard of care?

Critically, evidence-based medicine itself has been challenged for relying on probabilistic reasoning and post-hoc rationalizations, rather than true causal mechanisms. Randomized controlled trials, while considered the gold standard, are not immune to confounding and methodological flaws. This suggests that clinical decisions, even when grounded in evidence, may be just as opaque as AI outputs. Scholars argue that requiring full explainability from AI systems while accepting statistical opacity in traditional medicine may constitute a double standard. In fact, AI-supported clinical decision support systems are increasingly being tested in randomized trials, with emerging guidelines aiming to align AI research with evidence-based reporting standards.

Together, these challenges demand a rethinking of both technical design and legal accountability in the deployment of AI systems in healthcare. They also call for stronger alignment between AI development practices and the principles of evidence-based medicine to ensure that AI outputs are not only technically robust but also legally defensible.

Proposed best practices

Ensure compliance with the transparency and human oversight obligations outlined in the EU AI Act, particularly Articles 13 and 14, for all high-risk AI systems used in healthcare. AI systems should be designed to be sufficiently transparent so that deployers can interpret and appropriately use the system's output. Instructions for use must detail the system's technical capabilities and explainability features. XAI techniques should be implemented to support post-hoc interpretability of black-box models, enabling deployers to understand and justify decisions. Interdisciplinary collaboration between AI developers, providers, and healthcare professionals should be promoted to align design choices with clinical needs and legal standards. AI literacy among healthcare staff must be enhanced, as required by Article 4 of the AI Act, to ensure informed deployment and oversight. Transparency measures should be applied throughout the AI lifecycle to support accountability and contestability. Article 86.1 of the AI Act should be interpreted in alignment with GDPR principles to strengthen the patient's right to explanation and legal protection-by-design.

Design choices must be documented and made transparent, in line with Article 11 and Annex IV of the AI Act, to support regulatory oversight and traceability. Data governance principles from Article 10 of the AI Act must be embedded, including measures to prevent bias and ensure representativeness of training datasets. Developers should clarify and communicate the selection of ground truths, proxies, and features used in model training, recognizing their normative impact on outcomes. They should disclose the statistical distribution of training data and ensure that all relevant subgroups are adequately represented. A culture of accountability must be fostered by recognizing that AI systems are not neutral, but shaped by intentional design choices that affect patient care.

Randomized controlled trials tailored to AI systems should be supported, following CONSORT-AI, SPIRIT-AI, and FUTURE-AI guidelines, to validate AI-CDSS and establish new standards of care. Clinicians should be encouraged to justify their reliance on or deviation from AI recommendations, especially when such systems are supported by robust clinical evidence. If AI systems demonstrate higher diagnostic accuracy through validated trials, they may contribute to a new standard of care, provided their design choices and limitations are transparently reported.

Constraints

The limited applicability of Article 86.1 AI Act to medical devices may restrict patients' rights to explanation in certain scenarios. Technical challenges in implementing transparency for complex models, especially deep learning systems, persist. Automation bias and overreliance on AI outputs by clinicians remain risks. Translational bias may occur when models are applied to populations different from their training data. Legal uncertainty in attributing liability arises when transparency is insufficient to establish causation or fault. Both AI and traditional evidence-based medicine rely on statistical associations rather than causal mechanisms, creating epistemological limitations. Clinicians may feel pressured to adopt AI systems they do not fully understand, facing liability risks whether they follow or reject AI recommendations.

There is no consensus on whether black-box AI systems can set new standards of care, despite superior accuracy in some domains. Incomplete reporting standards for AI-based clinical trials may hinder the generalizability and legal defensibility of AI-CDSS outcomes. The evolving

nature of clinical guidelines and the variability of national liability regimes across EU Member States further complicate the integration of AI into standard-of-care assessments.

Year of publication: 2025.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Blatti, A., and Tramacere, S., 'The Transparency and Liability Issues Associated with AI-Based Medical Systems' in Casarosa, F., Gennari, F., and Rossi, A. (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) <https://doi.org/10.1007/978-3-031-99709-9_15>

(BP19-PR13) Navigating liability for AI-powered medical IoT: best practices and policy recommendations for active prostheses under the revised EU Product Liability Directive

Author: Francesca Gennari (LIDER-Lab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

Re-elaborated by: Arianna Rossi

Addressees
Medical device manufacturers and software developers; legal and regulatory teams in health technology companies; EU and national policymakers responsible for product liability, AI regulation, and medical device law; insurance providers and risk assessors; standard-setting bodies and technical committees; national courts and judicial training institutions; consumer protection agencies; and patient advocacy organizations.

Context
The EU Product Liability Directive (PLD), originally adopted in 1985, remained largely unchanged for decades despite ongoing scholarly debate and limited litigation at the EU level. However, the rapid development of emerging technologies, particularly AI, IoT, and robotics, has placed increasing pressure on the adequacy of the PLD. This culminated in the adoption of the Product Liability Directive Update (PLDU) in November 2024, which will become effective on 9 December 2026. The PLDU introduces a significant shift by recognizing software, including AI, as a product, regardless of its connection to hardware. This change has profound implications for the medical sector, particularly for complex, interconnected devices such as active prostheses.

Active prostheses, which integrate AI systems to control functions like gait movement, lie at the intersection of liability law, health regulation, and multi-layered technologies. The PLDU's compensation rationale, articulated in Article 5, ensures that any natural person suffering damage from a defective product is entitled to compensation. This principle reflects a broader EU legal rationale aimed at harmonizing liability across Member States, despite the fact that liability itself remains a national competence. Historically, compensation under the old PLD was difficult to obtain, particularly in cases involving chronic conditions or complex causality. The PLDU now becomes the standard for compensation mechanisms in product liability claims across the EU.

The Directive expands the definition of "product" to include software, manufacturing files, and interconnected systems, and redefines "defectiveness" to account for adaptive, data-driven technologies. It also introduces new procedural rules, including extended limitation periods, disclosure obligations, and legal presumptions of defectiveness and causality. These rules are intended to support the Directive's compensation rationale but may vary significantly across Member States, creating uncertainty for manufacturers operating in multiple jurisdictions. The involvement of private insurance companies, the role of transparency under Article 19, and the cascading liability structure outlined in Article 8 further complicate the landscape. Manufacturers must now anticipate liability not only for the physical device but also for its software components, updates, and interactions with third-party systems.

Definition of the challenge

The revised PLDU creates a complex liability regime for manufacturers of AI-powered medical devices, particularly active prostheses. The Directive's expanded definitions of product and defectiveness mean that software components, including AI systems, are now subject to liability. Manufacturers may be held liable for defects originating in third-party software or cloud services, even if they did not develop or directly manage those components. The presumption of control remains unless the manufacturer can prove otherwise, placing a heavy burden on those producing connected medical devices.

Procedural rules further complicate the liability landscape. Claimants may request access to the product's functional core, including AI systems, and invoke legal presumptions of defectiveness and causality when technical complexity impedes proof. These presumptions shift the burden of rebuttal to manufacturers, who must demonstrate that the product was safe and compliant. Liability exemptions under Article 11 offer limited relief. For example, manufacturers may not be held liable if the defect did not exist at the time of market placement, unless the defect arose from software updates, services, or modifications under the manufacturer's control. Other exemptions, such as compliance with legal requirements or the risk development defense, are narrowly interpreted and subject to national discretion.

The risk development exemption, in particular, raises concerns about legal fragmentation. Member States may enact more stringent rules, provided they are proportionate and justified by public interest objectives. This creates uncertainty for manufacturers operating across borders and underscores the need for harmonized implementation and clear guidance. The Directive's consumer-oriented procedural rules, including disclosure obligations and presumptions, will be interpreted differently across Member States, depending on judicial traditions and national consumer protection standards. Historical case law from countries such as France, Germany, Denmark, Finland, and Spain may serve as proxies for anticipating judicial behavior in future PLDU cases.

Proposed best practice

Manufacturers of active prostheses should adopt proactive risk management strategies that integrate legal foresight into product design and development. This includes identifying components most likely to cause defects, anticipating types of damage such as physical injury or psychological harm, and preparing country-specific protocols for post-damage procedures. Insurance coverage should be tailored to the types of damage the product may cause, and manufacturers should consider launching products first in jurisdictions with clearer or more favorable compensation frameworks.

Manufacturers should ensure that all software components, including updates and third-party integrations, are documented and assessed for compliance with the PLDU, the MDR, and the AI Act. They should avoid interoperability with commercial platforms unless liability risks are clearly understood and contractually managed. Instructions for use should be complete and accessible, potentially including multimedia formats such as video tutorials, to mitigate liability risks associated with user error.

Manufacturers should also prepare for the PLDU's procedural rules by maintaining detailed records of product design, safety testing, and post-market surveillance. They should be ready to respond to disclosure requests and to rebut legal presumptions with technical evidence. Collaboration with insurers and legal experts during the prototype phase can help anticipate litigation risks and ensure that products are defensible under the new liability regime.

Policy recommendation
Member States should adopt harmonized procedural standards for interpreting key concepts such as defectiveness, causality, and manufacturer's control. This includes establishing clear judicial guidelines for the disclosure of evidence, balancing transparency with the protection of intellectual property, and defining thresholds for invoking legal presumptions in cases involving complex technologies like AI-powered prostheses.

Member States should coordinate with the EU Commission to ensure that national implementations of the risk development exemption under Article 18 are proportionate, transparent, and limited to clearly defined product categories and public interest objectives. This coordination should include timely notification and justification of any stricter national measures, as well as participation in the EU-wide database of liability decisions to promote consistency and predictability across jurisdictions.

The involvement of private insurance companies should be encouraged, particularly in light of Article 19 PLDU's transparency provisions. Insurers can help standardize damage quantification and guide manufacturers in calibrating R&D investments and risk mitigation strategies. Precedents such as the Allianz IARD case suggest that insurance coverage for advanced devices like software-driven prosthetics may become more viable, provided that cross-border discrimination is avoided.

Regulators should clarify the legal status of technical standards and their role in liability exemptions, ensuring that compliance with state-of-the-art norms is appropriately recognized without undermining consumer protection. Finally, the PLDU should be seen not only as a regulatory challenge but also as an opportunity to align product development with the broader goals of the MDR and the AI Act, fostering a safer and more accountable medical technology ecosystem.

Constraints
The PLDU introduces legal uncertainty due to the variability of national implementations, particularly regarding procedural rules, liability exemptions, and the interpretation of defectiveness and causality. The presumption of manufacturer's control over third-party software and services increases exposure, especially in interconnected environments. Disclosure obligations and legal presumptions shift the burden of proof to manufacturers, requiring extensive documentation and technical rebuttals. Liability exemptions are narrowly construed and may be overridden by national laws. The risk development defense is subject to

national discretion and may be limited by public interest objectives. Insurance coverage may not be uniformly available across Member States, and technical standards may not be legally binding unless incorporated into national law. These constraints require manufacturers to adopt a cautious and well-informed approach to product design, market strategy, and legal compliance.

Year of publication: 2025.

References:
For a comprehensive analysis and detailed discussion, readers are referred to the full open access article:
Gennari, F., '(Product) Liability in the Medical Internet of Things. What Now?' in Casarosa, F., Gennari, F., and Rossi, A. (eds), *Enabling and Safeguarding Personalized Medicine. Data Science, Machine Intelligence, and Law*, vol 7 (Springer, Cham 2025) <https://doi.org/10.1007/978-3-031-99709-9_16>

3.5.Cybersecurity compliance and policy design

(PR14) Enhancing the participation of ENISA in the definition of cybersecurity requirements

Main author: Federica Casarosa (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

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[193] (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[194] 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 requirements defined in Articles 8-15, that are relevant to any AI system developer and

[193] AI Act Proposal (n14).
[194] Amendments to the AI Act (n76).

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.[195] 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:
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.

Year of publication: 2023.

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

(PR15) Reducing the risks of outdated cybersecurity requirements in European standardisation

Main author: Federica Casarosa (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

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[196] (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[197] 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 to set up their own self-defined standards, but also be able to self-assess their own compliance to the standards.

[196] AI Act Proposal (n14).
[197] Amendments to the AI Act (n76).

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.

Year of publication: 2023.

(BP20) The interplay between the Cyber Resilience Act and the Updated PLD – the scope of exemptions in case of damages to consumers

Main author: Federica Casarosa (LIDERLab, DIRPOLIS, Sant'Anna School of Advanced Studies, Pisa, Italy)

Addressee:
Manufacturers and designers of advanced wellness devices and applications that do not fall into the definition of medical devices.

Context/history of the problem
The recently adopted Cyber Resilience Act (Regulation 2024/2847 on horizontal cybersecurity requirements for products with digital elements - CRA) has moved a step forward towards the harmonization of cybersecurity requirements across the European market. Its scope does not cover the medical devices design and development, however it may affect those applications and devices that are excluded from the definition of Art. 2 (1) Medical Device Regulation, but still can provide a health function. The examples may be the social care robots that are designed to provide company to elderly people, regardless of the existence of a disease. According to CRA, the manufacturers are required to design, develop and put products on the market that have no known vulnerabilities. However, cybersecurity is not fault proof and any product available on the market can be subject to malicious attacks due to newly discovered vulnerabilities. Such attacks may result in damages for the consumers which are subject to the European and national rules on liability. In this context, it is crucial to verify if and how the recently updated Product Liability Directive (Directive 2024/2853 on Liability for Defective Products, UPLD) applies to the cases of vulnerability exploitation.

Definition of the problem
Art. 3 n. 40 CRA defines 'vulnerability' as "a weakness, susceptibility or flaw of a product with digital elements that can be exploited by a cyber threat". The concept can be break down into three essential elements: (1) the existence of a flaw or weakness (e.g., misconfiguration, human error); (2) the capacity of attackers to exploit this flaw; (3) the resulting compromise of information security. Even if the manufacturer is 100% sure that the product is void of vulnerabilities, this does not mean that the latter will not be discovered in the future.
In order to mitigate the risks emerging from the discovery of vulnerability, the CRA requires manufacturers to adopt vulnerability handling mechanisms that require multiple controls and checks across the so-called supporting period (minimum five years after the product is put on the market). Therefore, the manufacturer should adopt a system that envisages whenever a new vulnerability is discovered to initiate a process that aims at mitigating the vulnerability, eventually informing the user/consumer as regards the mitigation/solving measures to be adopted.

If the exploited vulnerability causes damage to the consumer or to the user of the device, the manufacturer can be held liable for such damage. This is due to the fact that the UPLD includes among the concept of defective product also the lack of relevant product safety requirements, including safety-relevant cybersecurity requirements (Art. 7 (2) lett. f).
In a strict interpretation of the rule, the mere discovery of a vulnerability may lead to the qualification of the product as defective, and in case of exploitation of the vulnerability this will trigger the liability of the manufacturer. Art. 11 (2) lett. e) UPLD provides for an exemption to the rule: which affirms that the manufacturer is exempted from liability if "the objective **state**

of scientific and technical knowledge at the time the product was placed on the market or put into service or during the period in which the product was within the manufacturer's control was not such that the defectiveness could be discovered".

Within the process of vulnerability handling, is this exemption applicable? In particular can the exemption apply when the exploitation of vulnerability

(1) before discovery from the manufacturer?
(2) before the mitigating measures are defined by the manufacturer?
(3) after the manufacturer disclosed the mitigation measures to the consumer?

Proposed best practice aimed at solving the problem
According to the case law of the Court of Justice and the interpretation of the UPLD provisions, the following approaches can solve the previous questions.

In case (1), the exemption provided in art. 11 (2) lett. e) is applicable if the manufacturer can provide proofs that the vulnerability was not known or discovered in the design and development phase. In this case, the reference to existing vulnerability databases and to penetration testing carried out can provide evidence of the objective state of the art regarding the existence of the vulnerability.

In case (2), the exemption mentioned above may be challenged by the consumer as the manufacturer has already discovered the vulnerability: it may be affirmed that the manufacturer has reached the technical knowledge regarding the existence of the defect. In this case, the exemption may still be applicable if the manufacturer can provide the evidence of the state of the art mentioned above. As a matter of fact, the decision of the ECJ in Case C-300/95 Commission v UK affirmed, since the early days of the Product Liability Directive, the exemption should be interpreted whenever the state of the art and technical knowledge is *objective*, not subjective, i.e. when the manufacture is individually capable.

In case (3), the exploitation of the vulnerability may not require the application of the mentioned exemption rule. As a matter of fact, if the manufacturer can provide evidence regarding the provision of the mitigation measures to the consumer, for instance, through security updates, the manufacturer can rely on art. 11 (2) lett. c). The provision affirms the manufacturer's liability where the defectiveness of a product is due to "a lack of software updates or upgrades necessary to maintain safety". Therefore, the evidence regarding the provision of the mitigation measure is sufficient to devoid from the liability.

References
Judgment of the Court (Fifth Chamber) of 29 May 1997. ECJ, C-300/95, Commission of the European Communities v United Kingdom of Great Britain and Northern Ireland, ECR *1997 I-02649*, ECLI:EU:C:1997:255.

FUTURE WORK

The policy recommendations and proposed best practices developed by the LaPoH offer a valuable, interdisciplinary contribution to the BioRobotics research ecosystem. Designed to translate complex regulatory frameworks into policy recommendations and practical guidance, these outputs support researchers, developers, and policymakers in navigating legal and ethical

challenges while fostering research and innovation. The future refinement, comparison, and validation of the best practices will be essential to ensure their relevance and effectiveness across diverse use cases. Likewise, the policy recommendations will need to evolve in step with the dynamic legal landscape, particularly in light of new legal instruments such as the AI Act, the EHDS, and the Cyber Resilience Act.

As outlined in "D7.8 Sustainability Plan", the LaPoH intends to continue this work by producing tailored resources (such as handbooks, guidelines, and interactive tools) and by exploring mechanisms like regulatory sandboxes and the greenhouse model to support both bioengineering innovators, policy-makers and regulators. These efforts will be grounded in a theoretical-empirical methodology and supported by interdisciplinary collaboration, strategic outreach, and targeted training. Importantly, while the current report has identified key aspects of regulatory enablers and gaps, it was not feasible to exhaustively address each and every legal domain covered in the cross-field analysis presented in D7.5. Future work will therefore aim to expand the scope of inquiry, deepening the understanding of underexplored gaps and ensuring that the policy outputs remain comprehensive and responsive to emerging challenges. Nevertheless, thanks to the interdisciplinary nature of LaPoH and its collaborations across engineering, medicine, economics, and design, this report has already extended beyond the scope of the initial cross-field regulatory analysis, offering a richer and more integrated perspective on the challenges and opportunities facing BioRobotics innovation.

A key area for future development is the structured inclusion of civil society, particularly patients, caregivers, and the broader public, as active contributors to research and innovation. Building on international models and recent recommendations,[198] LaPoH aims to promote public and patient engagement not only as a matter of transparency, but as a source of values and insights that can shape more inclusive, sustainable, and socially responsive innovation. This may include the introduction of dedicated roles, such as a Patient and Public Engagement Manager, and the development of health and digital literacy pathways to support meaningful participation. Such engagement is also essential to improving data quality and ensuring responsible technology use.

The long-term goal is to integrate LaPoH's outputs into future institutional frameworks and national and international projects, ensuring that legal foresight and ethical awareness remain embedded in technological development. Through a comprehensive sustainability strategy encompassing human resources, financial sustainability, and synergies with ongoing initiatives, LaPoH is well-positioned to scale its impact and continue serving as a reference hub for responsible innovation in biorobotics.

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 D7.5, as well as from the stakeholders' needs illustrated in D7.2 and those gathered through collaborative meetings with the technologists and the researchers of the other WPs.

[198] E.g., https://htai.org/engage-with-us/working-groups/htai-eshta/

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