LIABILITY IN AI-DRIVEN HEALTHCARE

JOINT OPINION (PT/ES/IT)

The National Council of Ethics for the Life Sciences
(Conselho Nacional de Ética para as Ciências da Vida, Portugal)
The Spanish Bioethics Committee (Comité de Bioética de España)
The Italian Committee for Bioethics
(Comitato Nazionale per la Bioetica)

May 2026

CNECV – VI Mandato

Maria do Céu Patrão Neves,
Presidente
André Dias Pereira,
Vice-Presidente
Anália Torres
Carlos Cortes
Carlos Maurício Barbosa
Hélder Mota Filipe
Inês Fronteira
Inês Godinho
João Ramalho-Santos
José Manuel Pereira de Almeida
Luís Duarte Madeira
Maria de Lurdes Martins
Margarida Godinho Costa
Margarida Silvestre
Miguel Oliveira da Silva
Miguel Ricou
Paula Pinto de Freitas
Pedro Fevereiro
Rosalvo Almeida
Rui Nunes

Comité de Bioética de Espanha (CBE)

Juan Carlos Siurana Aparisi,
Presidente
Isolina Riaño Galán,
Vicepresidenta
Alberto Palomar Olmeda
Atia Cortés Martínez
Carme Borrell i Thio
Cecilia Gómez-Salvago Sánchez
Aurelio Luna Maldonado
Íñigo de Miguel Beriain
José Antonio Seoane Rodríguez
Leonor Ruiz Sicilia
Lydia Feito Grande
María Desirée Alemán Segura
María José Buitrago Serna (secretaria)

Comitato Nazionale per la Bioetica

Angelo Luigi Vescovi
Presidente
Claudia Navarini
Vice-Presidente
Maria Luisa Di Pietro
Vice-Presidente
Riccardo Di Segni
Vice-Presidente
Mauro Ronco
Vice-Presidente
Carlo Antonio Barone
Luisella Battaglia
Raffaele Calabrò
Stefano Canestrari
Tonino Cantelmi
Cinzia Caporale
Giuseppe Casale
Lorenzo d'Avack
Antonio Da Re
Maria Grazia De Marinis
Luisa De Renzis
Alberto Gambino
Silvio Garattini
Mariapia Garavaglia
Matilde Leonardi
Andrea Domenico Maria Manazza
Domenico Menorello
Maurizio Mori
Assunta Morresi
Alessandro Nanni Costa
Carlo Maria Petrini
Giovanna Razzano
Marcello Ricciuti
Giuliana Ruggieri
Luca Savarino
Lucetta Scaraffia
Stefano Semplici

The rapporteurs were: *Inês Godinho, Juan Carlos Siurana, Stefano Semplici, André Dias Pereira, Maria do Céu Patrão Neves*

Liability in AI-Driven Halthcare – Joint Opinion (PT/ES/IT)

CONTENTS

Preamble

The drafting of joint opinions by several National Ethics Councils (NECs) represents an excellent opportunity to identify relevant and current issues shared across different countries, to broaden reflection and encourage pluralism in debate, to build inclusive and robust consensus, and consequently to enhance representativeness and strengthen the validity of the issued opinion, both in the eyes of state authorities and civil society at large.

The challenges of collaborative work are naturally significant, ranging from differences in language among participants, to scheduling meetings with a large number of rapporteurs, and to producing a jointly written text, whose final version must still undergo scrutiny by the plenary bodies of the NECs involved. However, the experience is enriching for all participants and ultimately benefits its intended audience.

The National Council of Ethics for the Life Sciences (Portugal), the Spanish Bioethics Committee, and the Italian Committee for Bioethics considered that *Liability in AI-Driven Healthcare* was not only an important but also an urgent issue in the current accelerated process of digitalizing healthcare provision and therefore decided to prioritize it in a joint opinion.

The initiative for this joint reflection came from Stefano Semplici, of the Italian Committee for Bioethics, after he heard Inês Godinho, from the National Council of Ethics for the Life Sciences, deliver a lecture entitled "Liability in AI-Driven Healthcare," which ultimately served as the basis for the collaborative work of the three NECs. At the time, the National Council of Ethics for the Life Sciences and the Spanish Bioethics Committee – whose collaboration had begun in 2012 – were finalizing another joint opinion. Building on this long-standing partnership between the Portuguese and Spanish NECs, currently chaired by Maria do Céu Patrão Neves and by Juan Carlos Siurana, respectively, both agreed to move forward with a three-way collaboration, now including the Italian NEC.

All three recognize that the digitalization of healthcare provision brings not only opportunities but also challenges that require prompt attention. Liability is certainly one of them. Although inherently legal in nature, its foundation is ethical – in that it serves as a guarantee for the protection of individuals involved, particularly the most vulnerable – and its framework must also remain ethical, respecting and promoting human rights and essential ethical principles. It is within this context that *Liability in AI-Driven Healthcare* was developed as the result of an intensive collaborative effort.

Introduction

Artificial Intelligence (AI) and Digital Health Information Technologies (DHITs) are transforming healthcare delivery, from diagnostics and digital therapeutics to clinical decision support and remote monitoring. While these innovations promise efficiency and improved outcomes, they also raise complex challenges. The pace of technological deployment has often shortened the time for ethical reflection on the requirements that should guide it and has also outstripped the development of clear regulatory guidance, creating uncertainties around responsibility (moral duty) and liability (legal obligation).

In this context, we observe that healthcare regulation increasingly embodies a paradox: frameworks designed to protect patients, such as the General Data Protection Regulation (GDPR), the Medical Device Regulation (MDR), and the Regulation (EU) 2024/1689 (AI Act), also risk fragmenting governance, chilling innovation, or creating gaps in enforcement and liability. Opaque algorithms and insufficiently robust consent models endanger patient autonomy, while safety is compromised by many risks, including bias (which can have different origins: minority bias, missing data bias, bubble bias, etc.), cybersecurity breaches, and unresolved questions of liability. The issue of liability, although inherently legal in nature, lies at the heart of the relationship between healthcare institutions and professionals, on the one hand, and patients and families, on the other, within the paradigm of care, and carries ethical implications due to the risk of exacerbating vulnerability and failing to uphold the principle of beneficence, which is foundational in the provision of healthcare. The complexity of this issue is also demonstrated by the fact that the European Commission after publishing a proposal for a directive on adapting non-contractual civil liability rules to AI in September 2022, decided to withdraw it in October 2025.

We will first define the scope of the problem. Then, we will move on to illustrate possible standard of care and liability models in AI healthcare. Finally, we will suggest some general and liability-based recommendations. The text focuses mainly on the use of AI in clinical practice, without delving into the regulatory aspects related to clinical trials necessary to develop AI tools for medical purposes.

1. Framework

The model of medical practice, though changing, has experienced exponential growth with the rise of AI. The digitalization of medicine, propelled by the increasing ubiquity of AI, has prompted a series of inquiries regarding its implications for the prevailing model of medical practice. The integration of technology within healthcare provision has been marked by ethical deliberation aimed at mitigating associated risks, namely by requiring respect for human rights and human dignity, and robust AI systems, transparent and explainable. Therefore, digital technologies have been introduced into healthcare through the establishment of boundaries and explicit delineation of exclusions. However, bioethics in postmodernity confronts an "irremediable plurality"[1].

One of the primary issues that has been brought to the spotlight is that of responsibility and liability, particularly in relation to trustworthy AI and its acceptance by both physicians and patients.

Indeed, in a study commissioned by the EU on AI in Healthcare (2022)[2], the implementation of prevailing principles and legislation within the domain of emerging AI applications in the medical sector raises several challenges (p. 26-27). Among these is the issue of the heterogeneity of actors, which thwarts the attribution of responsibilities to those involved in the process of developing, applying, and using medical algorithms and AI. In addition, there is the challenge of determining the exact origin of any medical error related to AI. Such an error may be attributable to the AI algorithm itself, to the data used to train the algorithm, or to the incorrect use or understanding in clinical practice.

The doctor-patient relationship, which has historically been at the center of issues relating to responsibility and liability – mainly due to malpractice issues –, is undergoing a paradigm shift with the progressive use of AI tools. The integration of AI into this dynamic introduces a new layer of complexity, with the addition of technological actors that mediate the interaction between healthcare providers and their patients[3], which makes responsibility more diffuse. The more difficult it is to establish the responsibility of the various actors in a given process, the greater the

[1] ENGELHARDT JR., TRISTAN, H. Foundations of Bioethics, 2nd ed., NY: OUP, 1996, p. 9.
The necessity of medical technology assessment is well established, and a variety of bodies have been established to fulfill this function. Examples include the US Office of Technology Assessment, among others. (TEN HAVE, H. / PATRÃO NEVES, M. C. P., Technology Assessment, in: *idem*, Dictionary of Global Bioethics, Springer, 2021, p. 997.)
[2] Available at: https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2022)729512.
[3] The issues regarding the safety requirements or data representativity requirements of AI technologies, considering the quest, at EU level particularly, for "trustworthy" AI, will not be discussed, as well as the relating product liability issues.

vulnerability of the most disadvantaged.

Nonetheless, the attention to responsibility and liability related to malpractice should not overlook the more positive approach to the doctor-patient relationship grounded in trust, which underpins good practices.

Responsibility, as an ethically based perspective and relational concept regarding human action, does not necessarily entail the (legal) obligation to compensate for negative consequences (*e.g.*, harm, damages) of actions, as does the concept of liability[4], which includes a sanctioning nature. Hence, the perspective will be that of liability within an ethical framework.

On the one hand, there are Several Ethical Guidelines (*e.g.*: UNESCO, WHO, EU), where a convergence in the main values and principles can be observed regarding the following principles:

- Human supervision (through HITL/HOTL/HIC models)
- Human responsibility and accountability
- Privacy, security and non-discrimination
- Fairness, equity and inclusiveness
- Autonomy and safety.

On the other hand, there are liability models and rules in healthcare that are not yet fully adapted to AI healthcare. Under the WHO: "Updated liability rules for the use of AI in clinical care and medicine should at least include the same standards and damages already applied to health care"[5].
What is important is that AI in healthcare cannot create a responsibility – and liability – gap.

2. Standard of care and AI models

AI in Healthcare has fueled discussions on the standard of care. There are some general recommendations, but one that stands out is that it is imperative to acknowledge that AI systems should be regarded primarily as complementary tools, rather than as substitutes.

[4] Furthermore, accountability, as the quality of answering for results of actions, although retrospective – similarly to liability – is not necessarily legally imposed.
[5] WHO, Ethics and Governance of Artificial Intelligence for Health, p. 76, available at https://iris.who.int/server/api/core/bitstreams/f780d926-4ae3-42ce-a6d6-e898a5562621/content.

Even when, in specific and well-defined contexts, highly reliable systems may operate with some degree of autonomy, human oversight must always remain a constant. This concept is highlighted in particular by the recent World Medical Association Statement on Artificial and Augmented Intelligence in Medical Care, and also in the Council of Europe's Steering Committee for Human Rights in the fields of Biomedicine and Health (CDBIO) extensive work on the impact of AI implementation on the patient-doctor relationship—from the perspective of both doctors and patients—and on safeguarding professional standards in that relationship and human rights[6]. The term *augmented* "signals a human-centered approach to AI—one that reinforces the physician's role as the final decision-maker"[7].

In fact, in the ethical guidelines put forward by WHO (2021), regarding the promotion of responsibility and accountability, responsibility can be ensured through the application of the "human warranty" (WHO, 2021:28), *i.e.*, the evaluation by physicians, patients, Ethics Committees, and Competent Authorities for the development of AI technologies. This evaluation can be simplified through the application of oversight with human intervention governance mechanisms (HITL). In terms of liability, when medical decisions involving AI technologies have as a collateral consequence harm to individuals, it is imperative that responsibility and liability processes clearly identify the relative roles of manufacturers and clinical users in causing the harm.

In this regard, the definition of liability models covering all actors involved in the design, development, implementation and use of AI technologies is a very sensitive issue. Targeting clinical users (doctors) as well as manufacturers (who are already subject to strict liability under the EU Directive on defective products) with strict civil liability, for example, could discourage the use of AI in medicine.

Liability models and rules are mostly related or dependent upon the existing AI standard of care models. Hence, if general liability models are in place, it is important to intertwine them with specific liability models in AI-driven healthcare and only then can these combined (general + specific) liability models be put together with the standard of care to ascertain the resulting liability scenario.

---

[6] Council of Europe, Steering Committee for Human Rights in the fields of Biomedicine and Health (CDBIO), The Report on the application of AI in healthcare and Its Impact on the 'patient-doctor' relationship (2024). https://www.coe.int/en/web/human-rights-and-biomedicine/ai-in-healthcare

[7] World Medical Association. WMA Statement on Artificial and Augmented Intelligence in Medical Care Adopted by the 76th WMA General Assembly, Porto, Portugal, October 2025. https://www.wma.net/policies-post/wma-statement-on-artificial-and-augmented-intelligence-in-medical-care/

As such, it is paramount to set an AI standard of care, to ascertain its implications regarding the possible (combined) liability models. There are two pathways regarding the standard of care – both advanced under the scope of the aforementioned guidelines or premises: AI Integration and AI Acceptance.

AI Integration means the assumption of an AI-driven healthcare, including AI tools in the standard of care. It can be set as partial – necessarily including AI tools in some healthcare services – or total, making AI healthcare delivery a norm, meaning that due care should include AI tools. Nonetheless, even with full integration, this does not mean the substitution of human healthcare; rather, it means that humans are to work with AI tools in a human-AI tool (human-machine) model.

AI Acceptance is a different standard. It does not necessarily include AI tools, but instead accepts their use in healthcare delivery, under the idea of the unavoidability of their use, in a model of human-driven healthcare. Here, the acceptance may be restricted to certain areas of practice, with use(r) guidelines or even with mere admissibility of use, all with respect of the applicable AI tool-maintenance requirements.

Regardless of the standard of care models of AI-healthcare, the main thing is not to lose sight of who the healthcare giver is. It is also necessary to address the phenomenon of *automation bias, i.e.,* tendency to over-rely on an automated system. Evidence shows that practitioners may over-rely on AI recommendations, even when conflicting evidence is present, thereby reducing human oversight to a merely formal safeguard. Recognizing and actively mitigating automation bias—through training, procedural safeguards, and institutional awareness—must therefore become part of the AI standard of care. Automation bias should be explicitly recognized as a systemic risk requiring both individual and institutional responses. For this reason, governance should evolve from a simple Human-in-the-Loop model to a more robust Human-in-the-Reasoning approach. In this framework, the practitioner is not merely a final approver of AI outputs, but is equipped with sufficient information, transparency, and tools to critically assess and, where necessary, challenge the system's recommendation and its underlying assumptions. If an Artificial Intelligence System (AIS) only provides an output in the form of a diagnosis (*e.g.*, signaling a neoplasm or an ongoing heart attack), and the AIS is certified as highly accurate, the human practitioner will face a relevant incentive to always agree with the system's output. By departing from the system's conclusion, practitioners would expose themselves to potential liability, should they be wrong, with no possibility of shifting the duty to compensate to the manufacturer. On the contrary, if the AIS were "explainable", providing an appropriate level of explainability or interpretability proportionate to the clinical context and the level of risk involved, practitioners could be held responsible

when they failed to spot a relevant error in the AIS's reasoning and conclusion in the same way as they would be held responsible if they failed to do so *vis-à-vis* an opinion rendered by a colleague.

Only under these conditions – training, procedural safeguards, institutional awareness and Human-in-the Reasoning approach – can human responsibility remain substantive and liability attribution remain coherent with its human-centered foundations, this shift from Human-in-the-Loop to Human-in-the-Reasoning being understood as a qualitative evolution ensuring that human oversight remains substantive rather than merely formal.

3. Liability models

The law establishes normatively regulated attribution schemes that direct the results of actions. A process, determined by normatively understood causality, exists for establishing a correspondence between conduct, result, and liability. In other words, depending on whether the result of the agent's conduct is damage or a breach of duty, the imputation of the action for the purposes of assigning liability is carried out according to certain pre-established standards. Liability models are not neutral: they shape critical behaviour, levels of trust, and the adoption of AI systems in practice.

It is imperative to acknowledge that the legal constructs of responsibility and liability, mostly in criminal law, are human-centered in nature, grounded as they are in an understanding of what ought to be, beyond pure empirical phenomenology, which implies the assumption of values and meaning, but also, foreseeability, in the sense of knowing for what one is responsible and liable for.

As civil law countries, Portugal, Spain and Italy follow the contractual and tort models in civil law and the harm model in criminal law. In civil law, strict liability is an exception. In criminal law there is never strict liability. Liability frameworks must strike a balance between avoiding excessive burdens on clinicians and preventing *de facto* immunity when relying on AI systems.

Regarding AI regulation, there are different legal frameworks:

3.1. Portugal

In Portugal, aside from the European legal framework, there is still no national law on AI complementing said framework. However, Portugal has national laws complementing other relevant European regulation: i) the Basic Health Law (Law

95/2019) mandates information system interoperability, NHS digitalization, and enhanced protection of health data; ii) medical device regulation, while centered on EU instruments, preserves complementary national frameworks through the legal regime for medical devices (DL 145/2009); iii) data protection is structured not only by the GDPR and its national implementing law (Law 58/2019), but also by specific legislation governing genetic information and health data (Law 12/2005).
Nevertheless, regarding AI, the National Artificial Intelligence Agenda was recently approved (Resolution nr. 2/2026). The National Artificial Intelligence Agenda (ANIA) is governed by six guiding principles and structured around four areas of action with clear objectives. The principles are as follows:

1. Responsible innovation, meaning that the development and adoption of AI must be carried out responsibly, *i.e.*, transparently, ethically, safely, and in line with European values.

2. Focus on strategic objectives, meaning that resources should be concentrated in a few areas with real transformative potential, avoiding dispersion and maximizing impact.

3. The state as a catalyst, leading by example.

4. More than technology, recognizing that the transformation and adoption of AI are, for the most part, an organizational and human challenge.

5. Product-oriented AI. This principle is based on the need to develop initiatives focused on concrete, useful, and scalable products that deliver real solutions for society.

6. Building on what already works and ensuring continuous evaluation: reinforcing what already works, promoting centers, infrastructures, and projects that have demonstrated impact, and avoiding duplication and dispersion of resources.

The Agenda is organized into four areas of action**: I.** Infrastructure and Data – robust computing capacity and data economy; **II.** Innovation and Adoption – accelerating adoption in the economy, with a focus on SMEs and Public Administration; **III.** Talent and Skills – training, attracting, and retaining talent in AI; and **IV.** Responsibility and Ethics – protecting citizens and promoting responsible innovation. Regarding this latter area of action, we emphasize the assertion that "responsibility should not be seen as a brake, but as a driver of value", in the sense that it reduces uncertainty and legal risk, while also increasing citizens' trust. The Agenda also foresees a national AI Regulation with fundamental coordinates: the definition of competent authorities; the coordination model and the penalty framework. Aside from the assessment of existing

resources and identification of critical needs for strengthening competencies, this regulation should also ensure alignment with AI regulatory sandboxes.

From a specifically ethical standpoint, and in summary, the Agenda advocates for the ethical use of AI, *i.e.,* one that is safe, transparent, trustworthy, and aligned with European and public values.

3.2. Spain

In Spain, there is still no legal framework to complement general European regulations, such as the AI Law and the GDPR; the framework that is further supplemented by the European Health Data Area (EHDS) Regulation.

There is a draft bill for the proper use and governance of AI, but it has not yet been approved.

The Spanish Agency for the Supervision of Artificial Intelligence (AESA) has been created (established by Royal Decree 729/2023, of August 22) and will be the entity responsible for monitoring compliance with and the practical application of risk level requirements in Spain.

A draft bill on Digital Health is currently under consideration. This bill adapts Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025, on the European Health Data Area, to Spanish law. It also amends Directive 2011/24/EU and Regulation (EU) 2024/284 and regulates the national interoperable digital health record and the use of digital technologies in healthcare. The draft bill is currently in the public consultation phase and is still in its preliminary stages.

A strategy for implementing AI in the National Health System (NHS) was established in September 2025. Among its key principles are:

1. The adoption of AI in the NHS must be based on essential principles such as safety, equity, transparency, and human oversight.

2. The State aims to promote this progress towards the widespread adoption and cultural acceptance of AI in a coordinated manner and to support health services in addressing the many and varied challenges it presents.

3. AI acts as a support tool, allowing healthcare professionals to focus on more complex and higher value-added tasks, such as critical clinical decision-making and

direct interaction with patients. This redesign does not imply a replacement but rather an adaptation of their functions, as well as the need for training in these skills, the emergence of new roles within organizations, and robust governance frameworks that ensure ethical, safe, and patient-centered implementation.

4. The adoption and effective integration of AI in the NHS requires a coordinated roadmap that considers the particularities of the Spanish health system, the ethical and legal aspects inherent in the use of sensitive health data, the need for interoperability and security of systems, the training and education of its professionals, as well as the active participation of both patients and professionals, and of technology providers.

5. The aim is to integrate AI into the NHS in an ethical, equitable and coordinated manner, with the objective of supporting the health care of the population, empowering patients and professionals through its use and optimizing the efficiency of the system.

6. The strategy is based on four transformative pillars:

Pillar One: Reliability. Citizen and professional trust will be at its highest when ethical, regulatory, and transparency principles are rigorously applied, guaranteeing security and privacy from the design stage and throughout the entire AI lifecycle.

Pillar Two: Utility. Healthcare professionals will work with the support of reliable AI tools that provide them with value, optimizing their time and updating their clinical knowledge to enable higher quality and more personalized patient care.

Pillar Three: Humanism. Each user will be able to receive personalized, predictive, preventive, and participatory healthcare thanks to the use of AI, which will allow them to assume a more active and responsible role in their own health care.

Pillar Four: Universality. AI will actively contribute to reducing health inequalities, helping to guarantee equitable access to diagnoses and innovative treatments, regardless of location or socioeconomic status, with sustainable solutions.

3.3. Italy

Italy has established a specific national legal framework to govern the use of AI, which complements overarching European regulations such as the AI Act and the GDPR. The cornerstone of the Italian approach is the Law of 23 September 2025, no. 132, which sets out provisions and delegates authority to the Government regarding AI.

With specific regard to the healthcare sector, this law establishes several key principles:

- Auxiliary Role of AI: AI systems are defined as tools to support the processes of prevention, diagnosis, treatment, and therapeutic choice. Their function is to improve the healthcare system and patient care.

- Primacy of Human Decision-Making: The law unequivocally states that the final decision is always reserved for the healthcare professional, who retains ultimate responsibility for clinical decisions, including in contexts where AI systems operate with a degree of autonomy. AI systems can provide non-binding suggestions, but they cannot replace the physician's judgment.

- Reliability and Verification: AI systems used in healthcare, along with the data they employ, must be reliable, periodically verified, and updated to minimize the risk of errors and enhance patient safety.

- Transparency and Non-Discrimination: Patients have the right to be informed about the use of AI technologies in their care. Furthermore, AI systems cannot be used to select or condition access to healthcare services based on discriminatory criteria.

At the institutional level, the law has provided for the creation of a national AI platform for healthcare, managed by the National Agency for Regional Health Services (AGENAS), designed to support healthcare professionals. It has also designated the national competent authorities for the supervision and implementation of AI regulations: the National Cybersecurity Agency (ACN) is responsible for surveillance and enforcement, while the Agency for Digital Italy (AgID) is tasked with promoting innovation and development. These bodies operate alongside other authorities with specific competencies, such as the Data Protection Authority (*Garante per la protezione dei dati personali*) and AGENAS itself for the health sector.

This framework is grounded in the broader data protection principles of the GDPR, particularly concerning the processing of health data. The processing of such data for reasons of significant public interest, such as healthcare, is permitted under Italian law, provided it is proportionate and includes specific measures to protect the fundamental rights of the data subject. The Italian Data Protection Authority has also been proactive in this area, issuing a "Decalogue for the realization of national health services through AI systems" to provide guidance.

Given the convergence on the principle of human responsibility, these national approaches illustrate the diversity of regulatory environments, which may influence both the allocation of liability and the definition of standards of care.

In this context, three different models are conceivable: the reliance model, the deviance model and the neutrality model.

As for the reliance model, there is an encouragement for practitioners to rely upon AI regarding the information and confirmation of their clinical judgement. Here, practitioners (doctors) are not liable for relying on trustworthy, responsible AI tools.

The deviance model is, one could say, the complete opposite. It holds practitioners fully liable for every decision, making reliance decisions in AI accountable as well.

As it becomes clear, these models lead to very different approaches to AI in healthcare from practitioners. In the first one, there is a risk of overreliance, but it allows for the full potential of the AI tool to be deployed; in the second one, there is a risk of not exploring this full potential, also considering its added requirements and possible liability.

The third model – the neutral model (WHO, 2021: 78-79) – is a no-fault liability regime. It takes into account the evolution of AI – and the fact that not even developers are fully in control of the evolution process – to consider no-fault/no-liability compensation funds, particularly in those contexts where harm arises from systemic or highly complex AI interactions. Here, developers or companies funding the development of AI (for healthcare) would take on some risk and have to obtain insurance or pay into an insurance fund in order to make up no-fault/no-liability compensation funds.

In sum, all models have different aims and all of them are possible as such or even in some combinations, as they are not equivalent and generate distinct incentives for clinical behaviour; therefore, context-sensitive and risk-based combinations should be considered.

It is crucial to recognize that these models represent archetypes rather than mutually exclusive options. Evolving legal systems are more likely to adopt hybrid solutions. For instance, a "reliance model" could be tempered by a duty for the professional to exercise "professional reasonableness", requiring intervention in the face of patently erroneous or anomalous AI outputs. Conversely, a "deviance model" might be mitigated by acknowledging the shared responsibility of manufacturers and healthcare institutions, especially in cases of systemic or design-related failures. This nuanced approach aligns with the principle of "Human-in-the-Reasoning" also advocated in this document.

4. Liability sandbox for AI-driven healthcare

Even though international ethical guidelines allow for a standard that can showcase some instances of non-compliance with ethical parameters that will result in legal liability, general guidelines have the limitation of not being adaptable to different AI standards of care models.

Although the AI Act aims at fostering regulatory sandboxes (Art. 57 AI Act), it is also undisputed that sandboxes provide a controlled environment for testing and experimentation with normative models of liability attribution, in order to ensure an ethical liability model for AI-driven healthcare.

It is possible to identify instances where liability is clear: the use of AI systems that do not meet the requirements of trustworthy AI; the dereliction of duty regarding the adherence to guidelines on human oversight and the primacy of the human; the failure to provide sufficient elucidation on the use of AI systems for the purpose of attaining informed consent; or even the failure to obtain adequate training and information prior to the implementation of AI systems, or the refusal of said training and information.

Nonetheless, these general guidelines do not amount to the necessary specificity required by the intertwinement between standard of care and liability foreseeability.

The optimal liability regime will depend heavily on what is construed as the "problem". On the one hand, if there is concern that the deployment of AI-driven healthcare is associated with a high risk for patients being harmed, aside from the maintenance of the mostly human-centered criminal law model, the road will be to keep tort law as it functionally is supposed to be, acting on (1) deterrence and (2) compensation of the victims. On the other hand, if there is a belief that reliance on AI promotes patient health, then it is problematic to have AI-resistant practitioners.

Thus, it is our suggestion that, with the different emerging models in relation with AI (standard of care and liability), a workgroup for a "liability sandbox" is undertaken in order to analyze, under current EU models, which is the one that meets the correct standard, considering the different ethical principles and guidelines: are current models adequate if the standard of care changes or shifts? Is there a need for an active policy for a liability model change that promotes using AI to its full potential in terms of healthcare? This sandbox has to be consistent with an updated standard of care.

What must be avoided is both a liability gap and a liability overflow: the main objective is to have clear rules that allow both practitioners and patients to trust (also) in the liability model in place: no good would come of having scenarios with uncertain outcomes.

5. Recommendations

The National Council of Ethics for the Life Sciences (Conselho Nacional de Ética para as Ciências da Vida, Portugal), the Spanish Bioethics Committee (Comité de Bioética de España), and the Italian Committee for Bioethics (Comitato Nazionale per la Bioetica), considering that

a) the use of artificial intelligence (AI) in healthcare should

- respect the ethical principles governing the healthcare relationship, particularly the doctor-patient relationship, including beneficence, non-maleficence, autonomy and justice

- comply with the established ethical guidelines of trustworthy AI, particularly human supervision, transparency, inclusiveness and safety

b) the applicable standard of care in the use of AI entails implications for the attribution of liability

recommend:

1. At a general level, regarding the applicable standard of care

1.1. In relation to AI applications, it is necessary to regulate prediction, detection, and treatment solutions for healthcare by ensuring that

i) digital health and AI applications do not lead to the dehumanisation or diminution of the doctor–patient relationship, but instead function as tools to support and enhance it

ii) the assessment and definition of the role of AI applications duly take into account the trust-based nature of the doctor-clinical relationship

iii) the standard of care regarding the use of AI implies that physicians and other healthcare professionals should only use in clinical practice AI systems that have been approved by competent regulatory bodies.

iv) the provision of care, as well as final decisions regarding diagnosis and treatment, are always made by human professionals, while recognising that AI systems may assist in such processes

v) healthcare professionals shall fulfil their duty to inform patients of the inherent benefits and risks associated with AI-based technologies, thereby enabling them to choose alternative courses of action and safeguarding their right to self-determination, particularly in complex or high-risk medical procedures

1.2. In relation to AI applications, the decision-maker should

i) use only AI systems that meet the requirements for trustworthy AI, including adequate validation based on high-quality data prior to any application, as recent evidence indicates that this requirement is not always met

ii) exercise appropriate human oversight and ensure a human-centered approach, without abdicating responsibility in favour of, or being substituted by, AI systems

iii) receive adequate training and obtain all necessary information prior to implementing or using AI systems or applications

iv) take into account that not all AI systems or medical robots fulfil the same function in determining the *lex artis ad hoc* (standard of care), and that the level of responsibility varies according to the degree of autonomy and the level of risk associated with the system; where human intervention is minimal, the associated risk may be correspondingly lower, whereas where the system plays an active and critical role in clinical procedures, the level of risk—and, consequently, its relevance for liability—increases significantly.

2. At the liability level, regarding the key aspects to comply with

2.1. In relation to general liability requirements

i) the liability model must be clear in both civil and criminal law, safeguarding practitioners' autonomy and patient safety, as well as ensuring adequate compensation in the event of harm.

ii) developers and companies involved in the development of AI systems play a significant role in their design and deployment and should therefore be taken into account within the liability framework

iii) it is important to assess the implications of different standards of care in relation to the various liability models

iv) a robust, comprehensive system of insurance-based compensation for harm should be implemented to foster confidence among both practitioners and patients, to reduce the practice of defensive medicine, as well as to increase error notification systems

v) The specific implications for criminal law should be thoroughly explored, particularly concerning the configuration of offenses such as culpable homicide or personal injury. This includes defining the contours of negligence for a healthcare professional who uncritically relies on an AI system (automation bias), and for the

producer who markets a system with known or knowable design flaws that contribute to the harmful event.

2.2. In relation to specific liability rules and allocation

i) civil liability arising from medical services provided with the assistance of AI systems (including diagnostic, therapeutic, surgical, and rehabilitative functions) should safeguard both the autonomy of physicians and the patient's right to fair compensation

ii) institutional liability should be the default rule in civil liability. In cases of gross negligence or willful misconduct (*dolus*), however, a right of recourse against the individual healthcare professional should be available, taking into account responsibility for the decision taken

iii) In the case of self-employed professionals in private practice, the law should likewise safeguard the injured patient's right to compensation and the healthcare professional's autonomy in decision-making

iv) Liability frameworks must reflect the differentiated role of AI systems, ensuring a proportionate allocation of responsibility, including in the context of criminal law

3. At the producers' and institutional liability level

i) Producers' liability should be further clarified and, where appropriate, simplified. In light of the emerging European regulatory framework, particularly following the reform of the Product Liability Directive, the notion of defectiveness—traditionally understood as lack of safety—should, in the context of medical AI (especially diagnostic systems), also encompass lack of performance**.** In particular, where an AI system fails to detect a lesion or condition, the issue may not be one of safety but rather of failure to meet the expected standard of accuracy, which should be reflected in liability assessments. It should be acknowledged that this extension would represent a significant evolution of the Product Liability Directive, shifting its focus from a tool to protect against unsafe products to a framework that also guarantees a certain level of product efficacy and accuracy

ii) The liability of hospitals and healthcare institutions should be primarily channeled to them, reflecting the allocation of business risk inherent in their activity. This may be configured, depending on the legal system, as a form of strict or presumed liability, covering prima facie all cases where harm is dependent on the functioning of the AIS adopted. This reflects the allocation of business risk inherent in their activity. Healthcare institutions should retain the possibility of recourse against manufacturers, thereby simplifying the position of the claimant without placing an excessive burden on medical practitioners. Indeed, shielding the practitioner from excessive litigation associated with the use of AIS, in particular in diagnostics, is essential to ensure adoption and use of the technology to the advantage of the patient

4. At the regulatory development level

i) clear, predictable and coherent rules should be established to avoid both liability gaps and liability overload, ensuring that the integration of AI enables clinicians and patients to clearly identify and understand the allocation of responsibility across different AI-enabled clinical scenarios

ii) the establishment of AI liability sandboxes should be considered, allowing policymakers, regulators, developers and clinicians to test liability frameworks in controlled environments before adoption at scale

Lisbon, May 19, 2026

The President of the CNECV	The President of the CBE	The President of the CNB
Maria do Céu Patrão Neves	*Juan Carlos Siurana*	*Angelo Luigi Vescovi*

The members of the Conselho Nacional de Ética para as Ciências da Vida (Portugal) have unanimously approved the final text of this opinion at the plenary session of May 15, 2026.

The members of the Comité de Bioética de España (Spain) have unanimously approved the final text of this opinion at the plenary session of May 11, 2026.

The regular members of the Comitato Nazionale per la Bioetica (Italy) have unanimously approved the final text of this opinion, a first version of which had been discussed and approved at the plenary session of April 27, 2026 through an online consultation, which took place between April 29 and May 8.

On May 19, the three National Ethics Committees held a joint meeting to together reaffirm their support for the Joint Opinion "Liability in AI-driven Healthcare".