Eva Dias Costa, Mónica Correia, Rui Nunes

Clinical AI requires living oversight: legal and ethical grounds for a new accountability framework

DOI: https://doi.org/10.34625/issn.2183-2705(39.2)2026.ic-5

Secção

Investigação Científica / Scientific Research*

* Os artigos presentes nesta secção foram sujeitos a processo de revisão segundo o método *blind peer review* / The articles in this section have undergone a blind peer review process.

Clinical AI requires living oversight: legal and ethical grounds for a new accountability framework

A IA em saúde exige uma supervisão contínua: fundamentos jurídicos e éticos para um novo quadro de responsabilização

Eva Dias Costa [1]
Mónica Correia [2]
Rui Nunes [3] [4]

Abstract: This article contends that the European frameworks governing clinical AI fail to sustain accountability once systems are in use, and that certification and documentation are not apt to capture distributed, evolving decision-making. Drawing on comparative doctrine, the paper diagnoses the gap between formal compliance and lived responsibility. It advances living oversight as a governance practice that is continuous and substantively human, reframes stewardship as a shared responsibility across developers, clinicians, institutions, and regulators, and sets out practical implications: dynamic consent, post-deployment auditing, and enforceable redress. Accountability has to be designed into operation – not verified after the event – if clinical AI is to remain compatible with the moral foundations of healthcare.
Keywords: Accountability; Living oversight; Stewardship; Informed consent; Human dignity; Clinical AI; Redress mechanisms

Resumo: Este artigo sustenta que os atuais instrumentos jurídicos que regem a inteligência artificial no sector da saúde na União Europeia são insuficientes para assegurar uma verdadeira responsabilidade ética e legal na fase pós-implementação dos sistemas. Argumenta que a arquitetura normativa vigente assenta em mecanismos estáticos de certificação e documentação, que não respondem à natureza distribuída e evolutiva da tomada de decisão clínica mediada por algoritmos. A partir de uma análise comparada e de uma reflexão normativa, identifica um desfasamento estrutural entre o cumprimento formal e a responsabilização efetiva. Defende que, para proteger a dignidade humana, a autonomia e para assegurar mecanismos efetivos de responsabilização, a supervisão deve tornar-se uma prática institucional viva: contínua, contextualizada, sempre com substancial envolvimento humano. O artigo desenvolve o conceito de responsabilidade por tutela (*stewardship*) como modelo de responsabilização partilhada e duradoura entre desenvolvedores, clínicos, instituições e reguladores. Apresenta implicações concretas para a governação da IA em saúde, como o consentimento dinâmico, auditorias pós-implementação e mecanismos de indemnização verdadeiramente executáveis. Conclui que a responsabilização deve estar

[1] Universidade Portucalense Infante D. Henrique, eva@upt.pt , https://orcid.org/0000-0003-0262-977X.
[2] Universidade do Porto, https://orcid.org/0000-0002-9034-7029.
[3] Universidade do Porto, https://orcid.org/0000-0002-1377-9899
[4] Author contributions: The article was conceived and drafted by Eva Dias Costa, who undertook the legal and ethical analysis. Mónica Correia contributed comments on the bioethical sections and reviewed the manuscript. Rui Nunes contributed to the conceptual framing of the study and reviewed the manuscript prior to submission.

incorporada no funcionamento efetivo dos sistemas, para que o uso de IA se mantenha compatível com os fundamentos morais dos cuidados de saúde.
Palavras-chave: Responsabilidade; Supervisão contínua; Tutela institucional; Consentimento informado; Dignidade humana; IA na saúde; Mecanismos de reparação.

1. Introduction

The introduction of Artificial Intelligence (AI) in healthcare has challenged fundamental assumptions about decision-making and responsibility in clinical settings, touching the universal principles that define the permissible scope of technological intervention in human life.

International instruments such as the Universal Declaration of Human Rights[5] and the UNESCO Universal Declaration on Bioethics and Human Rights[6] articulate dignity and autonomy as moral constants in the governance of science and medicine.

The Universal Declaration on Bioethics and Human Rights affirms that human dignity and individual autonomy should orient medical and scientific progress, and the UN General Assembly's Resolution 78/265[7] urges states to promote the development and use of artificial intelligence in ways that respect human rights, and to implement safeguards that ensure safety and accountability throughout the AI lifecycle. Likewise, the UN Human Rights Council's Resolution 47/23[8] on new and emerging digital technologies affirms that states and business actors must ensure that digital and algorithmic systems do not compromise human rights, and should implement safeguards to prevent discriminatory or rights-eroding practices.

Within Europe, these values find their first binding formulation in the Convention on Human Rights and Biomedicine[9] (the Oviedo Convention), adopted under the Council of Europe. The Convention operationalises human dignity and autonomy as

[5] UNITED NATIONS. *Universal Declaration of Human Rights* [online]. New York: United Nations, 1948. Available from: https://www.un.org/en/about-us/universal-declaration-of-human-rights [viewed 4 February 2026].
[6] UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. *Universal Declaration on Bioethics and Human Rights* [online]. Paris: UNESCO, 2005. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000146180 [viewed 4 February 2026].
[7] UNITED NATIONS GENERAL ASSEMBLY. *Resolution A/RES/78/265: Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development* [online]. New York: United Nations, 2024. Available from: https://docs.un.org/en/A/res/78/265 [viewed 4 February 2026].
[8] UNITED NATIONS HUMAN RIGHTS COUNCIL. *Resolution A/HRC/RES/47/23: New and emerging digital technologies and human rights* [online]. Geneva: United Nations, 2021. Available from: https://undocs.org/A/HRC/RES/47/23 [viewed 4 February 2026].
[9] COUNCIL OF EUROPE. *Convention for the Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Medicine (Oviedo Convention)* [online]. Strasbourg: Council of Europe, 1997. Available from: https://rm.coe.int/168007cf98 [viewed 4 February 2026].

justiciable duties. States have an obligation to ensure that the interests and welfare of the human being prevail over the sole interest of society or science.

The European Union's constitutional order integrates these same values, albeit at a different level. Through the Charter of Fundamental Rights of the European Union[10], dignity, integrity of the person, and data protection form part of EU's primary law. This alignment between the Council of Europe's bioethical framework and the EU's constitutional one forms a distinctive European model, in which moral norms are juridically codified and enforceable.

It is within this layered system – international soft law, binding Council of Europe bio-law, and EU constitutional and regulatory instruments – that the tension between human oversight and algorithmic autonomy emerges as a question of accountability under law.

This article examines the legal and ethical implications of that shift, and it focuses on how AI systems in healthcare blur traditional lines of responsibility and dignity. It considers the consequences of these transformations for individuals already situated in vulnerable positions within healthcare systems, and how the protections once anchored in direct human oversight are increasingly challenging to sustain.

By tracing these developments through international, European, and EU legal frameworks, as well as contemporary bioethical discourse, we seek to clarify how delegation of decision-making complicates accountability, and to ask whether, in this emerging order, the person can still remain visible within processes no longer centred on human judgement.

Indeed, clinical AI is not just a new tool in the field of healthcare; it is changing the very epistemic structure through which clinical decisions are made. This transformation is destabilising the normative foundations of medical practice. Values such as responsibility, autonomy, and human dignity require identifiable agents, intelligible reasoning, and meaningful interpersonal relationships. Static regulatory mechanisms are unable to adapt to the dynamic and adaptive nature of AI systems.

The central thesis of this article is therefore that accountability in clinical AI requires a model of living supervision, i.e., a form of continuous, relational, and

[10] Particularly in Articles 1, 3 and 8, CHARTER of Fundamental Rights of the European Union [online]. *Official Journal of the European Union*, C 326, 26.10.2012. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT [viewed 4 February 2026].

traceable governance in terms of rights, which integrates accountability throughout the AI lifecycle, rather than verifying it at a single point of certification. In this sense, the article adopts a unified perspective that treats ethical, legal, and operational concerns as interdependent dimensions of the same accountability challenge.

To frame the central argument, the article maps the main ethical and legal tensions (responsibility, consent, dignity, and transparency), before turning to liability architectures from an EU/US perspective, and concludes proposing a model of living oversight and stewardship. The method is doctrinal, engaging European and EU law alongside soft-law sources and, where they illuminate accountability, US doctrine. It aims to translate ethical commitments into enforceable duties through a model of living oversight.

2. The Epistemic Shift

As artificial intelligence takes hold in care, responsibility - both ethical and legal - comes under strain. Traditional bioethical models assume interpersonal transparency: that a patient can understand, evaluate, and consent to the actions of identifiable professionals.

Philosophical accounts of digital ethics have long warned of this opacity – or, to be precise, about opacity's effects on agency and answerability. As decision-making becomes computational, responsibility must evolve from an individualistic model to one of distributed moral agency, shared among designers, regulators, and practitioners[11].

Still, this call is far from realised, as empirical research found persistent gaps between ethical principles and institutional practice, particularly regarding accountability and stakeholder understanding and suggest that responsibility is not merely diffused, it is frequently displaced, creating spaces where no single actor feels answerable for outcomes[12].

Historically, medical decisions were grounded in personal expertise and relational dialogue between clinician and patient, but today, the same decisions are increasingly

[11] FLORIDI, Luciano. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford: Oxford University Press, 2023; BURRELL, Jenna. 'How the machine "thinks": understanding opacity in machine learning algorithms'. *Big Data & Society* [online]. 2016, 3(1), 2053951715622512. https://doi.org/10.1177/2053951715622512.

[12] BLEHER, Hannah; BRAUN, Matthias. 'Diffused responsibility: attributions of responsibility in the use of AI-driven clinical decision support systems'. *AI and Ethics* [online]. 2022, 2(4), pp. 747–761. https://doi.org/10.1007/s43681-022-00135-x; TANG, Lu; LI, Jinxu; FANTUS, Sophia. 'Medical artificial intelligence ethics: a systematic review of empirical studies'. *Digital Health* [online]. 2023, 9, 20552076231186064. https://doi.org/10.1177/20552076231186064.

mediated by algorithmic systems that synthesise data at a scale beyond human comprehension. This represents a huge epistemic shift: authority is moving from interpretive judgement to computational inference.

This migration of decision-making power profoundly alters the texture of accountability, since algorithmic healthcare is a multilayered web of responsibility that spans developers, institutions, and practitioners[13].

Within this network, no single actor can claim complete authorship of a decision, even though each contributes to its production. In practice, this means that when an AI-driven recommendation leads to harm – be it an incorrect diagnosis or a biased risk score – accountability becomes diffuse, spread across thin layers of design, implementation, and use, making redress procedurally complex and morally ambiguous.

The more AI systems are embedded in clinical workflows, the more clinicians will see themselves less as authors of decisions and more as users, and this perception will inevitably affect their ethical posture of care: where discretion once implied responsibility, automation now creates psychological and institutional distance. Clinicians can point to the algorithm, institutions to compliance standards, and developers to user error, each partially correct, all collectively unaccountable[14].

The potential benefits of AI in healthcare are, no doubt, substantial, but gains carry a high ethical cost. The patient is now diluted by mediation through data and code, stripped of context. The result is a form of decision-making that may be technically robust but is morally thinned, where the reasoning process may be auditable but is still completely alien to the patient's lived experience.

3. The Relational Conditions of Care

Informed consent - a cornerstone of bioethical practice - presupposes that patients can understand – and therefore meaningfully accept – the reasoning behind a proposed course of action. When medical decisions are shaped by opaque algorithmic systems, reasoning is increasingly inaccessible, sometimes even to the clinicians deploying the technology.

[13] KISELEVA, Anastasiya; KOTZINOS, Dimitris; DE HERT, Paul. 'Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations'. *Frontiers in Artificial Intelligence* [online]. 2022, 5, 879603. https://doi.org/10.3389/frai.2022.879603.
[14] TANG, Lu; LI, Jinxu; FANTUS, Sophia. 'Medical artificial intelligence ethics: a systematic review of empirical studies'. *Digital Health*, 2023.

This erosion of responsibility has direct implications for informed consent: it is obvious that if patients cannot access or comprehend the mechanisms shaping their care, their capacity for genuine consent becomes merely symbolic. This is a problem of algorithmic accountability: opacity undermines agency and erodes the legitimacy of the consent process itself. Consent depends on an intelligible account of decision-making; when such explanation becomes computationally or legally inaccessible, autonomy collapses into procedural formality[15].

Opacity undermines consent. Meaningful autonomy depends on grasping the context in which a decision takes shape – the informational landscape that surrounds it[16]. Many clinicians themselves struggle to interpret algorithmic outputs, raising the question of how informed consent can survive when even the intermediary human agents cannot explain the system's logic[17].

The challenge is therefore both epistemic and ethical – no tidy separation helps here. Opacity transforms the consent process from a deliberative exchange into an act of administrative compliance[18]; the patient's agency becomes formal, the signature replacing understanding. This shift erodes the reciprocity that once defined medical decision-making; it weakens the role of consent as a safeguard of dignity.

Emerging *dynamic consent* models can partially restore agency by treating consent as an iterative engagement across the AI lifecycle[19]. But patients and clinicians need to be able to actually interrogate how these systems work.

Consent is one of several lawful bases, under the GDPR, for processing personal data (Article 6). For sensitive health data, explicit consent sits alongside other grounds such as public health interests and scientific research with appropriate safeguards (Article 9); the right not to be subject solely to automated decisions (Article 22)

---

[15] YEUNG, Karen; LODGE, Martin (eds.). *Algorithmic Regulation*. Oxford: Oxford University Press, 2019.
[16] FLORIDI, Luciano, *The ethics of artificial intelligence*, 2023.
[17] TANG, Lu; LI, Jinxu; FANTUS, Sophia. 'Medical artificial intelligence ethics'. *Digital Health*, 2023.
[18] YEUNG; LODGE (eds.), *Algorithmic Regulation*, 2019; VILAS BOAS PINTO, João. '"I don't know where I'm going, I know I'm (not) going that way": some notes about the opportunities, challenges, and limits of artificial intelligence in the scope of administrative law'. *Revista Jurídica Portucalense* [online]. 2025, 38, pp. 286–307. https://doi.org/10.34625/ISSN.2183-2705(38)2025.IC-15.
[19] LAY, Winnie; GASPARINI, Loretta; SIERO, William; HUGHES, Elizabeth K. 'A rapid review of the benefits and challenges of dynamic consent'. *Research Ethics* [online]. 2025, 21(1), pp. 180–202. https://doi.org/10.1177/17470161241278064; LEE, Ah Ra; KOO, Dongjun; KIM, Il Kon; LEE, Eunjoo; YOO, Sooyoung; LEE, Ho-Young. 'Opportunities and challenges of a dynamic consent-based application: personalised options for personal health data sharing and utilisation'. *BMC Medical Ethics* [online]. 2024, 25(1), 92. https://doi.org/10.1186/s12910-024-01091-3.

reinforces the need for meaningful human involvement[20].

The European Health Data Space Regulation[21] takes a somewhat different tack, establishing a controlled access regime for secondary use, with permitted purposes adjudicated by national Health Data Access Bodies under strict procedural safeguards. For primary use, patients retain rights to restrict professional access – and, where Member States allow, to opt out of cross-border exchange. This regulatory architecture reflects a broader tension running through health data governance—between individual consent mechanisms and collective-benefit frameworks that rely on limitation of purpose, secure processing environments, and institutional oversight alongside (and sometimes in place of) granular patient permissions.

Whether these innovations amount to living consent or simply ease the terms of a collective-benefit model is a question regulators and patients will shape the answer to over the coming years.

Even so, dynamic consent alone cannot close the structural power gap between those who design and control AI systems and those who are subjected to them. The ethical requirement of autonomy presupposes direct human deliberation[22]. When AI intermediates the care relationship, autonomy must be reconceived as relational and collective – grounded in systems that remain intelligible and contestable at every stage of operation.

If informed consent is to remain credible, the distance between patient, clinician, and system has to close. Transparency must become operational — through models that explain themselves and oversight that means something. Consent, in the end, is inseparable from accountability: autonomy keeps its force only where responsibility remains visible and answerable.

As set out in the European framework recalled in the Introduction, dignity in healthcare has both ethical weight and binding legal force through the Oviedo Convention and the Charter of Fundamental Rights; it requires that the interests and

[20] SANTOS, Elisabete. 'Workers under the aegis of artificial intelligence: new risks and challenges in the light of data protection'. *Revista Jurídica Portucalense* [online]. 2025, 38, pp. 427–449. https://doi.org/10.34625/ISSN.2183-2705(38)2025.IC-22.

[21] EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. *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* [online]. *Official Journal of the European Union*, OJ L, 2025/327, 5.3.2025. Available from: https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng [viewed 4 February 2026].

[22] LE MOLI, Ginevra. 'Intelligence artificielle vs dignité humaine : quand la sous-performance humaine est légalement requise'. *Revue européenne du droit* [online]. 2022, 4(1), pp. 122–127. https://doi.org/10.3917/red.004.0122.

welfare of the human being prevail over the interests of science or society. With artificial intelligence mediating clinical decision-making, that requirement is placed under very substantive strain.

Dignity turns on recognition: on someone seeing you. It presupposes that a person's needs and vulnerabilities, and a person's voice, are engaged by another human being within a relationship of care[23]. Opaque systems displace that relationship by substituting interpersonal deliberation with statistical inference. In practice, human particularity is translated into patterns and probabilities, and the patient risks being re-ontologised[24], treated as a data entity optimised for prediction. Clinical outcomes may even be accurate, but the pathway by which they are reached still failed to honour the person as a moral subject. Dignity has to inform system design, data governance, and clinical interaction from the outset, rather than appearing only after deployment as a matter of residual protection[25].

AI also reshapes the clinical encounter. The person-centred paradigm of the Oviedo Convention presupposes a setting in which understanding and empathy form part of the act of care. When systems intermediate diagnosis or treatment, clinicians can forget they should be agents engaged in dialogue and slide into the role of output interpreters; where probabilistic reasoning displaces that engagement, the ethical horizon narrows, and the patient is treated as an instance of a model instead of a subject of concern[26].

The concept of dignity illustrates particularly well the ethical issues associated with this transformation. As machine inference intervenes in the clinical relationship, the patient risks becoming epistemically invisible, that is, a statistical case rather than a moral subject. Human supervision is therefore not only a governance structure, but also a normative guarantee: it ensures the relational visibility of the person in algorithmic decision-making[27].

The ethical challenge is not merely whether AI can make correct or fair decisions,

---

[23] *Idem, ibidem.*
[24] FLORIDI, Luciano. *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford: Oxford University Press, 2014.
[25] MOLBÆK-STEENSIG, Helga; SCHEININ, Martin. 'Human rights and artificial intelligence in healthcare-related settings: a grammar of human rights approach'. *European Journal of Health Law* [online]. 2025, 32(2), pp. 139–164. https://doi.org/10.1163/15718093-bja10146.
[26] LE MOLI, Ginevra. 'Intelligence artificielle vs dignité humaine'. *Revue européenne du droit*, 2022.
[27] HERZOG, Christian; BRANFORD, Jason. 'Relational ethics and structural epistemic injustice of AI in medicine'. *Philosophy & Technology* [online]. 2025, vol. 38, no. 4, 160 [viewed 4 February 2026]. Available from: https://doi.org/10.1007/s13347-025-00987-1.

but whether such systems can sustain the moral texture of care that respects persons *as ends in themselves*. AI governance must therefore prioritise semantic transparency[28], making the logic of systems legible enough to preserve moral and epistemic reciprocity between patient and practitioner.

Preserving dignity in this landscape calls for relational accountability – human oversight and ethical attention that travel with the system across its lifecycle. Explainable methods and human in the loop arrangements assist only insofar as they make decisions intelligible to clinicians and patients, and sustain the reciprocity on which trust depends[29]. The aim here is recognition that endures within data-driven practice, and ensures that persons remain visible as moral subjects. Respecting dignity in digital systems means making them understandable – and interruptible.

4. Liability and its Limits

As AI systems become embedded in clinical settings, the issue of legal liability becomes even more nuanced.

In traditional medical settings, if something goes wrong, the responsibility can be traced back to an individual or institution.

If an AI system makes an incorrect diagnosis or recommendation, the chain of liability is fractured. Who is responsible for the harm caused by the malfunctioning AI system: the clinician who used the tool, the healthcare institution that implemented it, the developers who built the system?

These questions are further complicated by the fact that AI is often seen as a support tool, but AI systems play a much more active role in making decisions – or even replacing human decision-making. We must confront the reality that liability isn't always clearly defined.

The AI Act[30] lays out a framework for high-risk systems, but legal accountability

[28] Semantic transparency denotes explainability that preserves the patient–clinician reciprocity of care by rendering salient inputs, model rationales, and decision pathways intelligible in professional language and actionable within clinical workflows. It goes beyond disclosure to support contestability – PIERCE, Robin L.; VAN BIESEN, Wim; VAN CAUWENBERGE, Daan; DECRUYENAERE, Johan; STERCKX, Sigrid. 'Explainability in medicine in an era of AI-based clinical decision support systems'. *Frontiers in Genetics* [online]. 2022, 13, 903600. https://doi.org/10.3389/fgene.2022.903600.

[29] ZHANG, Jie; ZHANG, Zong-ming. 'Ethics and governance of trustworthy medical artificial intelligence'. *BMC Medical Informatics and Decision Making* [online]. 2023, 23(1), 7. https://doi.org/10.1186/s12911-023-02103-9.

[30] EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised*

is still a grey area. Existing frameworks focus primarily on individual human agents, which does not adequately address the collective nature of decision-making in AI-assisted care[31].

The current regulatory framework represents real progress, and shows a growing awareness that algorithmic systems affect the very structure of clinical decision-making.

Existing regulations introduce documentation duties, enshrine rights to human review, and affirm dignity as a legal principle – all important advances. But their reach is still limited because the procedures they create tend to induce formal compliance without ensuring that responsibility endures in practice.

Article 22 of the GDPR affirms an individual's right not to be subject solely to automated decision-making that produces significant effects, a principle that rests on the idea that human oversight will function as a safeguard. The problem is, in practice, most clinical AI operates in hybrid human-in-the-loop configurations, where the boundary between automation and human judgement is opaque, and the GDPR's binary distinction between human and automated decisions fails to capture these layered forms of interaction, leaving patients uncertain about who or what has shaped their care[32].

Recent analyses of the AI Act and its interface with the GDPR confirm these structural blind spots: there are persistent regulatory gaps in post-market accountability, particularly in healthcare, where multiple actors contribute to outcomes, but none are individually liable, and harm often arises from distributed decision-chains[33].

The combined effect is a regime of documented conformity that leaves

---

rules on artificial intelligence (Artificial Intelligence Act) [online]. *Official Journal of the European Union*, OJ L, 2024/1689, 12.7.2024. Available from: https://eur-lex.europa.eu/eli/reg/2024/1689/oj [viewed 4 February 2026].

[31] GROTE, Thomas; BERENS, Philipp. 'On the ethics of algorithmic decision-making in healthcare'. *Journal of Medical Ethics* [online]. 2020, 46(3), pp. 205–211. https://doi.org/10.1136/medethics-2019-105586.

[32] KOLFSCHOOTEN, Hannah B. van. 'A health-conformant reading of the GDPR's right not to be subject to automated decision-making'. *Medical Law Review* [online]. 2024, 32(3), pp. 373–391. https://doi.org/10.1093/medlaw/fwae029.

[33] SCHMIDT, Jelena; SCHUTTE, Nienke M.; BUTTIGIEG, Stefan; et al. 'Mapping the regulatory landscape for artificial intelligence in health within the European Union'. *npj Digital Medicine* [online]. 2024, 7(1), 229. https://doi.org/10.1038/s41746-024-01221-6.; DUFFOURC, Mindy; GERKE, Sara. 'Decoding U.S. tort liability in healthcare's black-box AI era: lessons from the European Union'. *Stanford Technology Law Review* [online]. 2024, 27(1), pp. 1–58. Available from: https://law.stanford.edu/wp-content/uploads/2024/02/Publish_27-STLR-1-2024_Decoding-U.S.-Tort-Liability-in-Healthcares-Black-Box-AI-Era.pdf [viewed 4 February 2026].

substantive responsibility unaccounted for: every actor documents their conformity, yet accountability disappears into the system itself[34].

Even developers cannot fully explain an AI's reasoning, and the absence of clear liability allocation undermines redress for patients and, some argue, discourages innovation, as institutions and manufacturers fear unpredictable exposure to claims[35]. These structural features of AI-mediated care may demand hybrid compensation models in which responsibility is shared and patients receive accessible pathways to redress, including no-fault schemes that already operate alongside traditional liability systems[36].

Such approaches would complement the GDPR's procedural safeguards by ensuring that accountability is distributed and enforceable. They need to draw on more than one legal tradition.

In the United States, tort accountability is adjudicated *ex post*: the plaintiff must prove factual and legal causation after harm has occurred, typically through expert evidence that the court admits under the Daubert standard[37].

The EU's preventive architecture contrasts with U.S. ex-post liability under FDA 21 CFR Part 820[38] and common-law negligence.

These are two very different moral geographies of accountability, and each system exposes different vulnerabilities: while Europe risks procedural rigidity that

---

[34] CESTONARO, Clara; DELICATI, Arianna; MARCANTE, Beatrice; CAENAZZO, Luciana; TOZZO, Pamela. 'Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review'. *Frontiers in Medicine* [online]. 2023, 10, 1305756. https://doi.org/10.3389/fmed.2023.1305756; PALANIAPPAN, Kavitha; LIN, Elaine Yan Ting; VOGEL, Silke; LIM, John C. W. 'Gaps in the global regulatory frameworks for the use of artificial intelligence (AI) in the healthcare services sector and key recommendations'. *Healthcare* [online]. 2024, 12(17), 1730. https://doi.org/10.3390/healthcare12171730.

[35] PRICE, W. Nicholson II. 'Medical malpractice and black-box medicine'. In: COHEN, I. Glenn; FERNANDEZ LYNCH, Holly; VAYENA, Effy; GASSER, Urs (eds.). *Big Data, Health Law, and Bioethics*. Cambridge: Cambridge University Press, 2018. https://doi.org/10.1017/9781108147972.027.

[36] MALIHA, George; GERKE, Sara; COHEN, I. Glenn; PARIKH, Ravi B. 'Artificial intelligence and liability in medicine: balancing safety and innovation'. *Milbank Quarterly* [online]. 2021, 99(3), pp. 629–647. https://doi.org/10.1111/1468-0009.12504.

[37] *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993); KEETON, W. Page; DOBBS, Dan B.; KEETON, Robert E.; OWEN, David G. *Prosser and Keeton on the Law of Torts*. 5th ed. St Paul, MN: West Publishing Co., 1984; FEDERAL JUDICIAL CENTER; NATIONAL RESEARCH COUNCIL. *Reference Manual on Scientific Evidence* [online]. 3rd ed. Washington, DC: National Academies Press, 2011. Available from: https://nap.nationalacademies.org/catalog/13163/reference-manual-on-scientific-evidence-third-edition [viewed 4 February 2026]; AMERICAN LAW INSTITUTE. *Restatement (Third) of Torts: Liability for Physical and Emotional Harm*. St Paul, MN: American Law Institute, 2010.

[38] UNITED STATES. Food and Drug Administration. *Quality System Regulation* [online]. Code of Federal Regulations, Title 21, Part 820. Washington, DC: US Government Publishing Office, 2024. Available from: https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-820 [viewed 4 February 2026].

obscures accountability, the U.S. risks reactive litigation that encourages defensive design and fragmented adoption.

Cross-fertilisation between these models could perhaps yield hybrid mechanisms, combining ex-ante oversight with accessible ex-post compensation, to better align accountability with patient protection[39].

In clinical AI, where decisions are dynamic and distributed, neither reactive fault-finding nor static conformity can secure the continuity of moral responsibility that healthcare requires.

The EU's evolving approach, especially under the AI Act and product-liability reform proposals, suggests an effort to close this gap – transforming documentation from a bureaucratic formality into a living instrument of responsibility,

Building from that, the calls for reform increasingly converge on the need to embed accountability directly into data-protection governance. It is absolutely necessary to apply continuous monitoring and post-deployment auditing in to ensure that responsibility remains traceable across the AI lifecycle[40]. Such mechanisms would give practical effect to the GDPR's principle of "data protection by design and by default" (Article 25), translating it from a compliance requirement into an ongoing ethical obligation.

Transparency is the watchword of AI governance, but its practical meaning in healthcare is at best, ambiguous.

Under both the GDPR and the AI Act, developers and deployers of "high-risk" AI systems must ensure that their systems are explainable and auditable. Still, transparency in healthcare operates as a multilayered system of accountabilities[41] – and one that is also constrained by technical limits – that can leave post-market responsibility extremely thin.

This tension between formal transparency and substantive intelligibility is not new, and it did not begin with AI. The Court of Justice of the European Union, in Google

---

[39] DUFFOURC, Mindy; GERKE, Sara. 'Decoding U.S. tort liability in healthcare's black-box AI era: lessons from the European Union'. *Stanford Technology Law Review*, 2024.

[40] MEDTECH EUROPE. *Liability challenges in AI medical technologies: key considerations for policymakers and industry* [online]. MedTech Europe White Paper. Brussels: MedTech Europe, 2022. Available from: https://www.medtecheurope.org/resource-library/liability-challenges-in-ai-medical-technologies/ [viewed 4 February 2026].
MedTech Europe is the leading European industry association for the medical technology sector. Its 2022 White Paper frames liability primarily from the perspective of legal certainty and market stability for manufacturers and insurers.

[41] KISELEVA, Anastasiya; KOTZINOS, Dimitris; DE HERT, Paul. 'Transparency of AI in healthcare as a multilayered system of accountabilities'. *Frontiers in Artificial Intelligence*, 2022.

Spain (C-131/12)[42], had already affirmed that data protection under Articles 7 and 8 of the Charter requires individuals to remain visible within digital processes, recognising a right to trace and contest the handling of their personal information, including the right to be forgotten (Article 17), i.e., the right to delete their data[43].

The gap between regulatory 'transparency' and epistemic opacity has become a defining challenge of algorithmic medicine. It has introduced a semantic drift: machine reasoning gradually displaces the interpretive dialogue that once anchored the moral encounter between clinician and patient.

The stakes go beyond administration. In clinical settings, transparency underpins consent, accountability, and ultimately trust. When patients cannot understand how decisions about their health are made – and clinicians rely on outputs they cannot interrogate – the mutual recognition at the heart of medical ethics begins to fray.

The doctor–patient relationship evolves into a triadic dynamic in which the physician is no longer acts a fully autonomous fiduciary, and is reduced to the role of mediator between the patient and a technical system[44].

This algorithmic mediation reconfigures the physician's role and redistributes moral and legal accountability across a diffuse sociotechnical landscape. Such transformations challenge traditional expectations of clinical responsibility, requiring new frameworks that can address system-level accountability[45].

---

[42] *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*, C-131/12, ECLI:EU:C:2014:317, Court of Justice of the European Union [online]. 2014. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62012CJ0131 [viewed 4 February 2026].

[43] The ethical challenges of a 'right to be forgotten' in healthcare lie beyond the scope of this article. A body of recent literature has examined these issues in relation to genetic data, gender transition, pandemic contexts, and gamete donation; see, for example: CORREIA, Mónica; RÊGO, Guilhermina; NUNES, Rui. 'The right to be forgotten regarding genetic data: a legal and ethical analysis'. *Acta Bioethica* [online]. 2024, 30(2), pp. 231–243. Available from: https://actabioethica.uchile.cl/index.php/AB/article/view/76137 [viewed 4 February 2026]; CORREIA, Mónica; RÊGO, Guilhermina; NUNES, Rui. 'Gender transition: is there a right to be forgotten?' *Health Care Analysis* [online]. 2021, 29(4), pp. 283–300. https://doi.org/10.1007/s10728-021-00433-1; CORREIA, Mónica; RÊGO, Guilhermina; NUNES, Rui. 'The right to be forgotten and COVID-19: privacy versus public interest'. *Acta Bioethica* [online]. 2021, 27(1), pp. 59–67. Available from: https://actabioethica.uchile.cl/index.php/AB/article/view/63954 [viewed 4 February 2026].; CORREIA, Mónica; RÊGO, Guilhermina; NUNES, Rui. 'The right to be forgotten versus the right to disclosure of gamete donors' ID: ethical and legal considerations'. *Acta Bioethica* [online]. 2021, 27(1), pp. 69–78. Available from: https://actabioethica.uchile.cl/index.php/AB/article/view/63955 [viewed 4 February 2026].

[44] LORENZINI, Giorgia; ARBELAEZ OSSA, Laura; SHAW, David Martin; ELGER, Bernice Simone. 'Artificial intelligence and the doctor–patient relationship expanding the paradigm of shared decision making'. *Bioethics* [online]. 2023, 37(5), pp. 424–429. https://doi.org/10.1111/bioe.13158.

[45] ALLEN, Matthew R.; WEBB, Sophie; MANDVI, Ammar; FRIEDEN, Marshall; TAI-SEALE, Ming; KALLENBERG, Gene. 'Navigating the doctor-patient-AI relationship – a mixed-methods study of

The truth is that legal and institutional frameworks have not fully caught up. Transparency is often framed as a tool for fostering innovation and maintaining public trust[46], but it reduces accountability to compliance checklists: audits and documentation requirements, a proceduralism that risks shifting responsibility upward to institutions and regulators while still leaving patients without meaningful recourse.

By contrast, empirical research on explainable AI (XAI) in clinical diagnostics suggests that transparency must be conceived as interpretability in context rather than disclosure in principle[47]. Interpretability that aligns with clinical workflows is more likely to sustain accountability than the abstract promise of transparent algorithms is.

The path forward lies in translating transparency into an operational ethos. We need "ethically grounded compliance" that embeds transparency and explainability throughout the AI lifecycle, from data collection to deployment. Implementing this would mean designing systems where accountability is a structural feature, not an afterthought – traceable and contestable at every stage.

This would close the gap between what law requires and what ethics truly demands.

5. From Compliance to Stewardship

Legal frameworks protect dignity only when they embody the moral commitments they were designed to express. Law and ethics are not separate domains here; they are continuous. The task is to translate ethical principles into enforceable institutional practice.

At a deeper normative level, algorithmic discrimination is a violation of dignity itself. When AI systems classify individuals as statistical profiles, they risk treating

physician attitudes toward artificial intelligence in primary care'. *BMC Primary Care* [online]. 2024, 25(1), 42. https://doi.org/10.1186/s12875-024-02282-y.

[46] ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. *Recommendation of the Council on Artificial Intelligence* [online]. Paris: OECD, 2019 (amended 2024). OECD/LEGAL/0449. Available from: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [viewed 4 February 2026].

[47] THE PRECISE4Q CONSORTIUM; AMANN, Julia; BLASIMME, Alessandro; VAYENA, Effy; FREY, Dietmar; MADAI, Vince I. 'Explainability for artificial intelligence in healthcare: a multidisciplinary perspective'. *BMC Medical Informatics and Decision Making* [online]. 2020, 20(1), 310. https://doi.org/10.1186/s12911-020-01332-6.; SUBRAMANIAN, Harishankar V.; CANFIELD, Casey; SHANK, Daniel B. 'Designing explainable AI to improve human-AI team performance: a medical stakeholder-driven scoping review'. *Artificial Intelligence in Medicine* [online]. 2024, 149, 102780. https://doi.org/10.1016/j.artmed.2024.102780; TONEKABONI, Sana; JOSHI, Shalmali; McCRADDEN, Melissa D.; GOLDENBERG, Anna. 'What clinicians want: contextualising explainable machine learning for clinical end use' [online]. *Preprint, arXiv*, 2019. https://doi.org/10.48550/ARXIV.1905.05134.

persons no longer as subjects of recognition but as objects to be managed[48].

In healthcare, vulnerability is the starting condition. When patients are treated as objects, the moral consequences are acute: they become invisible within the very systems meant to serve them. Dignity has to function as a limit on instrumentalisation, and this aligns with bioethical traditions that put responsiveness before efficiency.

From a moral-theological perspective, dignity is not an abstract status, but a lived reality upheld in concrete relationships of care[49]. This resonates, albeit from a different tradition, with a human-rights-based model of governance in which Articles 1 and 3 of the EU Charter – dignity and bodily integrity – provide normative guidance for AI regulation[50]. The key obligation is the same: algorithmic decision-making must remain compatible with empathy and personal accountability.

Translating these insights into governance practice demands that we shift from procedural box-ticking to real ethical architecture. Dignity cannot be retrofitted into AI; it must be designed in.

This is achieved by embedding human oversight, interpretability, and contestability at every stage of the AI lifecycle – from data collection to clinical deployment – and guaranteeing that each act of delegation preserves a traceable line of moral responsibility.

To maintain trust and justice in this evolving environment, accountability must be treated as an active practice, designed into the system through audit trails, explainable interfaces, and sustained human oversight at the points where decisions are enacted. The challenge is to ensure that it does not sever the moral continuity of care[51].

Affirming accountability in AI-driven healthcare requires a re-conceptualisation of responsibility itself. European human-rights law offers a crucial normative anchor: Article 1 of the EU Charter frames human dignity as the inviolable foundation of all rights, while Article 3 safeguards bodily integrity[52]. These provisions imply that accountability for AI decisions in healthcare must be evaluated not only through

[48] ORWAT, Carsten. 'Algorithmic discrimination from the perspective of human dignity'. *Social Inclusion* [online]. 2024, 12, 7160. https://doi.org/10.1745/si.7160.
[49] GOZUM, Ivan Efreaim A.; FLAKE, Chastene Christopher D. 'Human dignity and artificial intelligence in healthcare: a basis for a Catholic ethics on AI'. *Journal of Religion and Health* [online]. 2024. Available from: https://doi.org/10.1007/s10943-024-02206-1 [viewed 4 February 2026].
[50] MOLBÆK-STEENSIG, Helga; SCHEININ, Martin. 'Human rights and artificial intelligence in healthcare-related settings'. *European Journal of Health Law*, 2025.
[51] VILLEGAS-GALAVIZ, Carolina; MARTIN, Kirsten. 'Moral distance, AI, and the ethics of care'. *AI & SOCIETY* [online]. 2024, 39(4), pp. 1695–1706. https://doi.org/10.1007/s00146-023-01642-z.
[52] MOLBÆK-STEENSIG, Helga; SCHEININ, Martin. 'Human rights and artificial intelligence in healthcare-related settings'. *European Journal of Health Law*, 2025.

causation or fault, but through the extent to which systems respect and preserve the person's status as a moral agent.

However, current legal regimes remain fragmented. The AI Act (2024) introduces ex-ante obligations for high-risk systems – risk assessments and documentation, but it provides little guidance for post-market accountability when harm occurs. This procedural approach ends up protecting institutions more than individuals: compliance is verified before deployment, and failures afterward are treated as technical anomalies, not as the ethical breaches they actually are. Similarly, the Article 22 right to human oversight is often satisfied by nominal, often perfunctory, review.

There are calls for hybrid frameworks, ones that combine regulatory supervision with accessible remedies, proportionate liability models balancing innovation and patient protection, so that accountability shifts from merely ceremonial to truly operative. Developers and healthcare institutions would then share responsibility for algorithmic harm, while patients could obtain redress through no-fault compensation schemes.

In practice, this would mean embedding liability within the governance lifecycle of AI systems, continuous post-market monitoring as a tool for both safety and trust[53].

The truth is, though, that industry self-regulation cannot replace legal enforceability. The key reform here is the creation of an accountability framework that recognises responsibility as shared over shifted. It is true that obligations are distributed, but they remain interdependent.

Developers must design systems that are transparent in their operation and also actively mitigated against bias.

Institutions have a duty to validate and document the tools they adopt, maintaining genuine human-in-the-loop oversight, and not limit themselves to symbolic checks.

As for clinicians, they do not lose their moral or professional responsibility simply because an algorithm is involved; they remain answerable for how they apply these systems in their practice.

Regulators, for their part, must ensure that auditing does not become optional and that patients have real, enforceable avenues of redress when harm occurs.

This shift parallels the emerging notion of collective responsibility in bioethics,

[53] MEDTECH EUROPE. *Liability challenges in AI medical technologies: key considerations for policymakers and industry*. 2022.

where moral agency is traceable even when it is shared.

Distributed responsibility does not mean diluted responsibility[54]; it calls for a transparent architecture that records how decisions are co-produced and who retains authority at each stage. Implementing such a system would make accountability visible and therefore enforceable in real time.

A promising direction lies in linking these reforms to the EU's liability-directive proposals currently under negotiation. Harmonising liability standards for AI-related harm would give patients a foundation for equitable redress across Member States. But to align with human-rights obligations, the Directive's criteria should treat dignity and autonomy as substantive elements of due diligence, not as ideals invoked from outside. The law would then integrate what bioethics has long understood: responsibility in healthcare cannot be separated from recognition of the person.

Ultimately, strengthening accountability through legal reform requires moving beyond the notion of liability as an after-the-fact remedy, otherwise AI in healthcare serves innovation without justice.

As the World Health Organization stresses, regulation must follow the life of the system[55], making sure that AI technologies remain safe, equitable, and aligned with the fundamental values of medical ethics throughout their deployment, and not merely at the moment of approval.

The current European framework (mostly built around the AI Act and the GDPR), provides a foundation but still lacks mechanisms for living oversight. Both instruments still rely primarily on ex-ante conformity assessment and documentation, assuming that accountability can be verified once, at certification, but the fact is such static models are ill-suited to dynamic systems that learn and evolve in response to new data[56].

By contrast, the Medical Devices Regulation[57] and the In Vitro Diagnostic

[54] FLORIDI, Luciano, *The ethics of artificial intelligence*, 2023.
[55] WORLD HEALTH ORGANIZATION. *Regulatory considerations on artificial intelligence for health* [online]. Geneva: WHO, 2023. Available from: https://www.who.int/news/item/19-10-2023-who-outlines-considerations-for-regulation-of-artificial-intelligence-for-health [viewed 4 February 2026].
[56] BOUDERHEM, Rabaï. 'Shaping the future of AI in healthcare through ethics and governance'. *Humanities and Social Sciences Communications* [online]. 2024, 11(1), 416. https://doi.org/10.1057/s41599-024-02894-w.
[57] EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. *Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices* [online]. *Official Journal of the European Union*, L 117, 5.5.2017. Available from: https://eur-lex.europa.eu/eli/reg/2017/745/oj [viewed 4 February 2026].

Regulation[58] already impose post-market surveillance and vigilance duties that could serve as a model for genuine continuous oversight.

Embedding similar, real-time monitoring and auditability in AI governance would transform conformity from a paperwork exercise into a living duty of attention, shared by regulators and healthcare providers.

There is a further weakness and it lies in cybersecurity. Both the AI Act[59] and the NIS 2 Directive[60] formally require healthcare providers and AI developers to adopt risk-management and incident-reporting procedures.

These provisions remain procedural, focused on documenting policies rather than testing resilience. The boundary between safety and security is porous: corrupted data or model drift can degrade clinical reliability silently, long before a breach comes to light. Real-time penetration testing, red-team evaluation, and transparent disclosure of incidents would bring cybersecurity into line with the principle of continuous accountability that the current framework still lacks.

At the level of the Council of Europe, the Commissioner for Human Rights has recently expanded this perspective in a speech[61] by proposing human rights oversight as a formal requirement in all high-risk AI contexts, including healthcare. This calls for mandatory Human Rights Impact Assessments (HRIAs) at key points in the system's lifecycle – design, deployment, and retraining – making certain that dignity, autonomy, and non-discrimination are explicitly evaluated alongside safety and accuracy. This reframes oversight as something more than procedure. It becomes an inquiry into how AI systems affect the lived experience and rights of those subject to their decisions.

International coordination matters more than we usually admit, and global and regional organisations have already outlined normative expectations that can

---

[58] EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. *Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices* [online]. *Official Journal of the European Union*, L 117, 5.5.2017. Available from: https://eur-lex.europa.eu/eli/reg/2017/746/oj [viewed 4 February 2026].

[59] EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. Regulation (EU) 2024/1689 (Artificial Intelligence Act), arts 9 and 73.

[60] EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. *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 (NIS 2 Directive)* [online]. *Official Journal of the European Union*, L 333, 27.12.2022. Available from: https://eur-lex.europa.eu/eli/dir/2022/2555/oj [viewed 4 February 2026].

[61] O'FLAHERTY, Michael. *Human rights oversight of artificial intelligence* [online]. Strasbourg: Council of Europe, Commissioner for Human Rights, 2025. Available from: https://www.coe.int/en/web/commissioner/-/human-rights-oversight-of-artificial-intelligence [viewed 4 February 2026].

complement binding EU law. The OECD AI Principles[62] and the UNESCO Recommendation on the Ethics of AI[63] both call for risk-proportionate oversight, transparency, and accountability mechanisms that operate throughout the lifecycle of AI systems.

The WHO Guidance on Ethics and Governance of AI for Health[64] translates these values into the clinical domain, insisting that autonomy and equity remain central across all stages of design and deployment. The 2020 Council of Europe Recommendation on Algorithmic Systems[65] adds a specifically rights-based dimension, urging states to integrate human-rights impact assessments and continuous supervision of algorithmic systems.

Together, these instruments articulate a shared international vocabulary of responsibility that reaches beyond the EU's procedural focus, emphasising that genuine oversight depends on transparency and redress that travel with the technology rather than stopping at national borders.

These converging frameworks share this conviction: effective governance of medical AI must treat accountability as a shared ethical practice rather than a formal regulatory outcome.

Ultimately, holistic oversight entails a shift in regulatory philosophy, from compliance to stewardship. Compliance measures whether an AI system meets a checklist of rules; stewardship demands that every actor within the system – developer, clinician, institution, and regulator – actively sustain the moral and legal integrity of decision-making. This requires a reflexive governance model, one capable of revising itself in light of new knowledge and societal expectations[66].

Oversight becomes an ongoing dialogue among institutions and individuals, anchored in the recognition that the legitimacy of algorithmic medicine depends not on

---

[62] ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449).

[63] UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. *Recommendation on the Ethics of Artificial Intelligence* [online]. Paris: UNESCO, 2021. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000381137 [viewed 4 February 2026].

[64] WORLD HEALTH ORGANIZATION. *WHO guidance on ethics and governance of artificial intelligence for health* [online]. Geneva: WHO, 2021. Available from: https://www.who.int/publications/i/item/9789240029200 [viewed 4 February 2026].

[65] COUNCIL OF EUROPE. Committee of Ministers. *Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems* [online]. Strasbourg: Council of Europe, 2020. Available from: https://rm.coe.int/09000016809e1154 [viewed 4 February 2026].

[66] O'FLAHERTY, Michael. *Human rights oversight of artificial intelligence*, 2025.

automation's efficiency but on its capacity to remain answerable to those it serves.

By embedding legal oversight within this continuous, reflexive model, healthcare systems can preserve the moral core of medicine while embracing the transformative potential of AI. A truly holistic approach ensures that accountability is not outsourced to procedures or devices, that it is a shared and sustained human enterprise, one that is transparent, and responsive to the dignity of every patient.

6. Conclusion

AI in healthcare reshapes how clinical decisions are made and who is seen as responsible for them. Where once there was a direct chain of accountability between the clinician and the patient, we now encounter hybrid systems where decisions are mediated by algorithms and institutional routines. This shift challenges the legal and ethical infrastructure of medical responsibility.

Frameworks such as the AI Act, the GDPR, and the Oviedo Convention (alongside complementary instruments like the MDR and NIS 2) signal a growing commitment to embedding rights and dignity in the governance of AI. The Oviedo Convention articulates a bioethical foundation for protecting autonomy in clinical contexts; the AI Act introduces testing and documentation requirements; the GDPR affirms rights to human review and data protection. These are not superficial interventions. They matter, but they remain insufficient to address the deeper ethical and institutional transformations underway. In practice, human oversight is too often displaced by procedural proxies; responsibility is fragmented across multiple actors; and patients are left with limited insight into – or recourse against – the systems that shape their care. The central challenge is no longer whether AI should be integrated into healthcare, it is how such integration can preserve the moral conditions of care: autonomy, intelligibility, proximity, and redress. This is the essence of stewardship: an ethics of maintenance, where technology remains answerable to the human lives it serves.

To meet this challenge, regulatory efforts must move beyond compliance rituals toward structural responsibility.

Accountability has to be actionable. Right now, liability spreads thin – everyone documents, no one owns the outcome. Legal frameworks need to name who answers for what, patients need somewhere to go when things feel wrong, early enough that systems can still respond. Accountability that only appears in litigation

arrives too late to change anything.

Oversight must be continuous. Certifying a system once isn't enough. These systems learn, drift, and surprise. Ethics has to travel with the technology: present in daily use, and be revisited when things change.

Consent has to be reconfigured. The traditional model depends on explanation, and algorithms mostly can't provide it. Patients need to have ongoing ways to understand their care, to question it, to refuse it. Consent has to remain in conversation.

Opacity has to give way to intelligibility. If clinicians themselves cannot audit the system, the patients have no chance.

Human oversight must be preserved as a condition of care. Human-in-the-loop has to mean more than a signature. Clinicians need room to question, override, to take responsibility.

The trajectory of AI in healthcare is not predetermined: the ethical and legal frameworks can still shape how responsibility is understood and enacted. The AI Act, the GDPR, and the Oviedo Convention offer a foundation, but the future depends on whether we treat these frameworks as the ceiling or as the floor.

Ethical responsibility cannot be bolted on, it has to be built in, and this requires shifting from formalised compliance to *situated accountability*: the ability to act, explain, and justify decisions in context, as they unfold.

Ultimately, the multiplication of regulatory measures does not by itself address the accountability gap. This analysis points to the need for governance practices that operate continuously and remain attentive to the evolving character of AI systems. Accountability has to accompany these systems in use, not appear intermittently. A dynamic model of supervision helps to sustain the moral and legal coherence of healthcare at a time when adaptive algorithms are increasingly involved in clinical decision-making.

AI can improve healthcare, and it already does. The question is on what terms – and whether trust, justice, and care survive the improvement[67]. For that, accountability cannot be audited after the fact, it must be embedded in the design. Only then can AI

[67] RUSTAMZADE, Aykhan; BAGIROV, Anar Ramiz; HUSEYNOV, Sahil Zahir. 'Justice and artificial intelligence'. *Revista Jurídica Portucalense* [online]. 2024, 36, pp. 163–184. https://doi.org/10.34625/ISSN.2183-2705(36)2024.IC-7.

serve medicine without unravelling the human responsibilities that define it.

REFERENCES

Normative Instruments

CHARTER of Fundamental Rights of the European Union [online]. *Official Journal of the European Union*, C 326, 26.10.2012. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT [viewed 4 February 2026].

COUNCIL OF EUROPE. *Convention for the Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Medicine (Oviedo Convention)* [online]. Strasbourg: Council of Europe, 1997. Available from: https://rm.coe.int/168007cf98 [viewed 4 February 2026].

COUNCIL OF EUROPE. Committee of Ministers. *Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems* [online]. Strasbourg: Council of Europe, 2020. Available from: https://rm.coe.int/09000016809e1154 [viewed 4 February 2026].

EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. *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 (NIS 2 Directive)* [online]. *Official Journal of the European Union*, L 333, 27.12.2022. Available from: https://eur-lex.europa.eu/eli/dir/2022/2555/oj [viewed 4 February 2026].

EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. *Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices* [online]. *Official Journal of the European Union*, L 117, 5.5.2017. Available from: https://eur-lex.europa.eu/eli/reg/2017/745/oj [viewed 4 February 2026].

EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. *Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices* [online]. *Official Journal of the European Union*, L 117, 5.5.2017. Available from: https://eur-lex.europa.eu/eli/reg/2017/746/oj [viewed 4 February 2026].

EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)* [online]. *Official Journal of the European Union*, OJ L, 2024/1689, 12.7.2024. Available from: https://eur-lex.europa.eu/eli/reg/2024/1689/oj [viewed 4 February 2026].

EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION. *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* [online]. *Official Journal of the European Union*, OJ L, 2025/327, 5.3.2025. Available from: https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng [viewed 4 February 2026].

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. *Recommendation of the Council on Artificial Intelligence* [online]. Paris: OECD, 2019 (amended 2024). OECD/LEGAL/0449. Available from: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [viewed 4 February 2026].

UNITED NATIONS. *Universal Declaration of Human Rights* [online]. New York: United Nations, 1948. Available from: https://www.un.org/en/about-us/universal-declaration-of-human-rights [viewed 4 February 2026].

UNITED NATIONS GENERAL ASSEMBLY. *Resolution A/RES/78/265: Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development* [online]. New York: United Nations, 2024. Available from: https://docs.un.org/en/A/res/78/265 [viewed 4 February 2026].

UNITED NATIONS HUMAN RIGHTS COUNCIL. *Resolution A/HRC/RES/47/23: New and emerging digital technologies and human rights* [online]. Geneva: United Nations, 2021. Available from: https://undocs.org/A/HRC/RES/47/23 [viewed 4 February 2026].

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. *Universal Declaration on Bioethics and Human Rights* [online]. Paris: UNESCO, 2005. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000146180 [viewed 4 February 2026].

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. *Recommendation on the Ethics of Artificial Intelligence* [online]. Paris: UNESCO, 2021. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000381137 [viewed 4 February 2026].

UNITED STATES. Food and Drug Administration. *Quality System Regulation* [online]. Code of Federal Regulations, Title 21, Part 820. Washington, DC: US Government Publishing Office, 2024. Available from: https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-820 [viewed 4 February 2026].

WORLD HEALTH ORGANIZATION. *WHO guidance on ethics and governance of artificial intelligence for health* [online]. Geneva: WHO, 2021. Available from: https://www.who.int/publications/i/item/9789240029200 [viewed 4 February 2026].

WORLD HEALTH ORGANIZATION. *Regulatory considerations on artificial intelligence for health* [online]. Geneva: WHO, 2023. Available from: https://www.who.int/news/item/19-10-2023-who-outlines-considerations-for-regulation-of-artificial-intelligence-for-health [viewed 4 February 2026].

Jurisprudence

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, C-131/12, ECLI:EU:C:2014:317, Court of Justice of the European Union [online]. 2014. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62012CJ0131 [viewed 4 February 2026].

Bibliographic References

ALLEN, Matthew R.; WEBB, Sophie; MANDVI, Ammar; FRIEDEN, Marshall; TAI-SEALE, Ming; KALLENBERG, Gene. 'Navigating the doctor-patient-AI relationship – a mixed-methods study of physician attitudes toward artificial intelligence in primary care'. *BMC Primary Care* [online]. 2024, 25(1), 42. https://doi.org/10.1186/s12875-024-02282-y.

AMERICAN LAW INSTITUTE. *Restatement (Third) of Torts: Liability for Physical and Emotional Harm*. St Paul, MN: American Law Institute, 2010.

BLEHER, Hannah; BRAUN, Matthias. 'Diffused responsibility: attributions of responsibility in the use of AI-driven clinical decision support systems'. *AI and Ethics* [online]. 2022, 2(4), pp. 747–761. https://doi.org/10.1007/s43681-022-00135-x.

BOUDERHEM, Rabaï. 'Shaping the future of AI in healthcare through ethics and governance'. *Humanities and Social Sciences Communications* [online]. 2024, 11(1), 416. https://doi.org/10.1057/s41599-024-02894-w.

BURRELL, Jenna. 'How the machine "thinks": understanding opacity in machine learning algorithms'. *Big Data & Society* [online]. 2016, 3(1), 2053951715622512. https://doi.org/10.1177/2053951715622512.

CESTONARO, Clara; DELICATI, Arianna; MARCANTE, Beatrice; CAENAZZO, Luciana; TOZZO, Pamela. 'Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review'. *Frontiers in Medicine* [online]. 2023, 10, 1305756. https://doi.org/10.3389/fmed.2023.1305756.

CORREIA, Mónica; RÊGO, Guilhermina; NUNES, Rui. 'Gender transition: is there a right to be forgotten?' *Health Care Analysis* [online]. 2021, 29(4), pp. 283–300. https://doi.org/10.1007/s10728-021-00433-1.

CORREIA, Mónica; RÊGO, Guilhermina; NUNES, Rui. 'The right to be forgotten and COVID-19: privacy versus public interest'. *Acta Bioethica* [online]. 2021, 27(1), pp. 59–67. Available from: https://actabioethica.uchile.cl/index.php/AB/article/view/63954 [viewed 4 February 2026].

CORREIA, Mónica; RÊGO, Guilhermina; NUNES, Rui. 'The right to be forgotten versus the right to disclosure of gamete donors' ID: ethical and legal considerations'. *Acta Bioethica* [online]. 2021, 27(1), pp. 69–78. Available from: https://actabioethica.uchile.cl/index.php/AB/article/view/63955 [viewed 4 February 2026].

CORREIA, Mónica; RÊGO, Guilhermina; NUNES, Rui. 'The right to be forgotten regarding genetic data: a legal and ethical analysis'. *Acta Bioethica* [online]. 2024, 30(2), pp. 231–243. Available from: https://actabioethica.uchile.cl/index.php/AB/article/view/76137 [viewed 4 February 2026].

DUFFOURC, Mindy; GERKE, Sara. 'Decoding U.S. tort liability in healthcare's black-box AI era: lessons from the European Union'. *Stanford Technology Law Review* [online]. 2024, 27(1), pp. 1–58. Available from: https://law.stanford.edu/wp-content/uploads/2024/02/Publish_27-STLR-1-2024_Decoding-U.S.-Tort-Liability-in-Healthcares-Black-Box-AI-Era.pdf [viewed 4 February 2026].

FEDERAL JUDICIAL CENTER; NATIONAL RESEARCH COUNCIL. *Reference Manual on Scientific Evidence* [online]. 3rd ed. Washington, DC: National Academies Press, 2011. Available from:

https://nap.nationalacademies.org/catalog/13163/reference-manual-on-scientific-evidence-third-edition [viewed 4 February 2026].

FLORIDI, Luciano. *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford: Oxford University Press, 2014.

FLORIDI, Luciano. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford: Oxford University Press, 2023.

GOZUM, Ivan Efreaim A.; FLAKE, Chastene Christopher D. 'Human dignity and artificial intelligence in healthcare: a basis for a Catholic ethics on AI'. *Journal of Religion and Health* [online]. 2024. Available from: https://doi.org/10.1007/s10943-024-02206-1 [viewed 4 February 2026].

GROTE, Thomas; BERENS, Philipp. 'On the ethics of algorithmic decision-making in healthcare'. *Journal of Medical Ethics* [online]. 2020, 46(3), pp. 205–211. https://doi.org/10.1136/medethics-2019-105586.

HERZOG, Christian; BRANFORD, Jason. 'Relational ethics and structural epistemic injustice of AI in medicine'. *Philosophy & Technology* [online]. 2025, vol. 38, no. 4, 160 [viewed 4 February 2026]. Available from: https://doi.org/10.1007/s13347-025-00987-1

KEETON, W. Page; DOBBS, Dan B.; KEETON, Robert E.; OWEN, David G. *Prosser and Keeton on the Law of Torts*. 5th ed. St Paul, MN: West Publishing Co., 1984.

KISELEVA, Anastasiya; KOTZINOS, Dimitris; DE HERT, Paul. 'Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations'. *Frontiers in Artificial Intelligence* [online]. 2022, 5, 879603. https://doi.org/10.3389/frai.2022.879603.

KOLFSCHOOTEN, Hannah B. van. 'A health-conformant reading of the GDPR's right not to be subject to automated decision-making'. *Medical Law Review* [online]. 2024, 32(3), pp. 373–391. https://doi.org/10.1093/medlaw/fwae029.

LAY, Winnie; GASPARINI, Loretta; SIERO, William; HUGHES, Elizabeth K. 'A rapid review of the benefits and challenges of dynamic consent'. *Research Ethics* [online]. 2025, 21(1), pp. 180–202. https://doi.org/10.1177/17470161241278064.

LE MOLI, Ginevra. 'Intelligence artificielle vs dignité humaine : quand la sous-performance humaine est légalement requise'. *Revue européenne du droit* [online]. 2022, 4(1), pp. 122–127. https://doi.org/10.3917/red.004.0122.

LEE, Ah Ra; KOO, Dongjun; KIM, Il Kon; LEE, Eunjoo; YOO, Sooyoung; LEE, Ho-Young. 'Opportunities and challenges of a dynamic consent-based application: personalised options for personal health data sharing and utilisation'. *BMC Medical Ethics* [online]. 2024, 25(1), 92. https://doi.org/10.1186/s12910-024-01091-3.

LORENZINI, Giorgia; ARBELAEZ OSSA, Laura; SHAW, David Martin; ELGER, Bernice Simone. 'Artificial intelligence and the doctor–patient relationship expanding the paradigm of shared decision making'. *Bioethics* [online]. 2023, 37(5), pp. 424–429. https://doi.org/10.1111/bioe.13158.

MALIHA, George; GERKE, Sara; COHEN, I. Glenn; PARIKH, Ravi B. 'Artificial intelligence and liability in medicine: balancing safety and innovation'. *Milbank Quarterly* [online]. 2021, 99(3), pp. 629–647. https://doi.org/10.1111/1468-0009.12504.

MEDTECH EUROPE. *Liability challenges in AI medical technologies: key considerations for policymakers and industry* [online]. MedTech Europe White Paper. Brussels: MedTech Europe, 2022. Available from: https://www.medtecheurope.org/resource-library/liability-challenges-in-ai-medical-technologies/ [viewed 4 February 2026].

MOLBÆK-STEENSIG, Helga; SCHEININ, Martin. 'Human rights and artificial intelligence in healthcare-related settings: a grammar of human rights approach'. *European Journal of Health Law* [online]. 2025, 32(2), pp. 139–164. https://doi.org/10.1163/15718093-bja10146.

O'FLAHERTY, Michael. *Human rights oversight of artificial intelligence* [online]. Strasbourg: Council of Europe, Commissioner for Human Rights, 2025. Available from: https://www.coe.int/en/web/commissioner/-/human-rights-oversight-of-artificial-intelligence [viewed 4 February 2026].

ORWAT, Carsten. 'Algorithmic discrimination from the perspective of human dignity'. *Social Inclusion* [online]. 2024, 12, 7160. https://doi.org/10.1745/si.7160.

PALANIAPPAN, Kavitha; LIN, Elaine Yan Ting; VOGEL, Silke; LIM, John C. W. 'Gaps in the global regulatory frameworks for the use of artificial intelligence (AI) in the healthcare services sector and key recommendations'. *Healthcare* [online]. 2024, 12(17), 1730. https://doi.org/10.3390/healthcare12171730.

PIERCE, Robin L.; VAN BIESEN, Wim; VAN CAUWENBERGE, Daan; DECRUYENAERE, Johan; STERCKX, Sigrid. 'Explainability in medicine in an era of AI-based clinical decision support

systems'. *Frontiers in Genetics* [online]. 2022, 13, 903600. https://doi.org/10.3389/fgene.2022.903600.

PRICE, W. Nicholson II. 'Medical malpractice and black-box medicine'. In: COHEN, I. Glenn; FERNANDEZ LYNCH, Holly; VAYENA, Effy; GASSER, Urs (eds.). *Big Data, Health Law, and Bioethics*. Cambridge: Cambridge University Press, 2018. https://doi.org/10.1017/9781108147972.027.

RUSTAMZADE, Aykhan; BAGIROV, Anar Ramiz; HUSEYNOV, Sahil Zahir. 'Justice and artificial intelligence'. *Revista Jurídica Portucalense* [online]. 2024, 36, pp. 163–184. https://doi.org/10.34625/ISSN.2183-2705(36)2024.IC-7.

SANTOS, Elisabete. 'Workers under the aegis of artificial intelligence: new risks and challenges in the light of data protection'. *Revista Jurídica Portucalense* [online]. 2025, 38, pp. 427–449. https://doi.org/10.34625/ISSN.2183-2705(38)2025.IC-22.

SCHMIDT, Jelena; SCHUTTE, Nienke M.; BUTTIGIEG, Stefan; et al. 'Mapping the regulatory landscape for artificial intelligence in health within the European Union'. *npj Digital Medicine* [online]. 2024, 7(1), 229. https://doi.org/10.1038/s41746-024-01221-6.

SUBRAMANIAN, Harishankar V.; CANFIELD, Casey; SHANK, Daniel B. 'Designing explainable AI to improve human-AI team performance: a medical stakeholder-driven scoping review'. *Artificial Intelligence in Medicine* [online]. 2024, 149, 102780. https://doi.org/10.1016/j.artmed.2024.102780.

TANG, Lu; LI, Jinxu; FANTUS, Sophia. 'Medical artificial intelligence ethics: a systematic review of empirical studies'. *Digital Health* [online]. 2023, 9, 20552076231186064. https://doi.org/10.1177/20552076231186064.

THE PRECISE4Q CONSORTIUM; AMANN, Julia; BLASIMME, Alessandro; VAYENA, Effy; FREY, Dietmar; MADAI, Vince I. 'Explainability for artificial intelligence in healthcare: a multidisciplinary perspective'. *BMC Medical Informatics and Decision Making* [online]. 2020, 20(1), 310. https://doi.org/10.1186/s12911-020-01332-6.

TONEKABONI, Sana; JOSHI, Shalmali; McCRADDEN, Melissa D.; GOLDENBERG, Anna. 'What clinicians want: contextualising explainable machine learning for clinical end use' [online]. *Preprint, arXiv*, 2019. https://doi.org/10.48550/ARXIV.1905.05134.

VILAS BOAS PINTO, João. '"I don't know where I'm going, I know I'm (not) going that way": some notes about the opportunities, challenges, and limits of artificial intelligence in the scope of administrative law'. *Revista Jurídica Portucalense* [online]. 2025, 38, pp. 286–307. https://doi.org/10.34625/ISSN.2183-2705(38)2025.IC-15.

VILLEGAS-GALAVIZ, Carolina; MARTIN, Kirsten. 'Moral distance, AI, and the ethics of care'. *AI & SOCIETY* [online]. 2024, 39(4), pp. 1695–1706. https://doi.org/10.1007/s00146-023-01642-z.

YEUNG, Karen; LODGE, Martin (eds.). *Algorithmic Regulation*. Oxford: Oxford University Press, 2019.

ZHANG, Jie; ZHANG, Zong-ming. 'Ethics and governance of trustworthy medical artificial intelligence'. *BMC Medical Informatics and Decision Making* [online]. 2023, 23(1), 7. https://doi.org/10.1186/s12911-023-02103-9.

Data de submissão do artigo: 08/12/2025

Data de aprovação do artigo: 05/02/2026

Edição e propriedade:
Universidade Portucalense Cooperativa de Ensino Superior, CRL

Rua Dr. António Bernardino de Almeida, 541 - 4200-072 Porto

Email: upt@upt.pt