Viewpoint

Rethinking Health Care Regulatory Sandboxes Under the European Union AI Act

María Flores Ceballos[1*]; Fernando Cagigas Villar[2*]; Alberto López Tejerina[2]

[1]Escuela Superior de Ingeniería y Tecnología (ESIT), Universidad Internacional de La Rioja, Logroño, Spain,
[2]AIR Institute (Fundación Instituto de Investigación Internacional en Inteligencia Artificial y Ciencias de la Computación), Salamanca, Spain
*these authors contributed equally

Corresponding Author:

María Flores Ceballos
Escuela Superior de Ingeniería y Tecnología (ESIT)
Universidad Internacional de La Rioja
Avenida de la Paz, 137
Logroño 26006
Spain
Email: maria.floresceballos@unir.net

Abstract

AI models have enabled substantial advances in health care. Since the entry into force of the European Union (EU) AI Act in 2024, these systems have become subject to a common regulatory framework in the EU intended to harmonize AI governance across all member states. One of the act's distinctive features is that it accommodates existing economic, technological, and regulatory differences among European countries by leaving certain implementation responsibilities to national authorities. This is evident in regulatory sandboxes, supervised environments for developing, testing, and validating AI systems before deployment. In this viewpoint, we provide a policy-oriented analysis of health care regulatory sandboxes under the EU AI Act and propose recommendations for their design. These recommendations address the need to document model compression techniques that may affect the classification of large high-risk systems above the 10^{25} floating point operations per second threshold, incorporate explainability into AI systems from the design stage, identify and document biases in health care datasets, clarify the treatment of federated learning parameter updates under the General Data Protection Regulation, and strengthen the oversight of AI agents in clinical environments. The implementation timeline makes this discussion particularly timely given the approaching deadline of August 2, 2027, by which all member states must have established at least one national regulatory sandbox.

JMIR AI 2026;5:e98653; doi: 10.2196/98653

Keywords: deep learning; delivery of health care; federated learning; European Union AI Act; EU AI Act; regulatory sandboxes; artificial intelligence; AI

Introduction

The integration of AI into European health care systems represents a major transformation in contemporary medicine, with significant implications for diagnostic precision, personalized treatment, and care efficiency [1]. These systems are currently regulated in the European Union (EU) under the AI Act, which establishes a common legal framework based on the risk profile of AI systems across all member states. However, the European landscape remains structurally heterogeneous as countries differ substantially in technological capacity, economic investment, and the maturity of their national AI-related regulatory environments. For this reason, although the AI Act provides a harmonized legal architecture, it leaves certain implementation responsibilities in the hands of individual member states, among which regulatory sandboxes occupy a relevant position.

This Viewpoint aims to provide a set of recommendations for the design of regulatory sandboxes in the health care sector, one of the most sensitive domains for the deployment of AI systems. The recommendations are informed by a policy-oriented analysis of selected challenges under the EU AI Act. They address both the governance of sensitive medical datasets and the treatment of recent AI models and techniques whose legal interpretation may prove challenging. Given that by August 2, 2027, each member state must ensure the existence of at least one regulatory sandbox, we argue that

this discussion is both timely and potentially useful for the design and operational development of such environments.

Regulation of AI in the EU Under the AI Act

The EU AI Act, the world's first comprehensive legal framework for the regulation of AI, entered into force on August 1, 2024 [2]. This milestone represents a pivotal moment in technology governance by establishing harmonized rules for the internal market (Article 1(a)); fostering the development and uptake of human-centric and trustworthy AI systems (Article 1(b)); and ensuring a high level of protection of health, safety, and fundamental rights (Article 1(c)).

The AI Act follows a phased implementation timeline (Table 1). The first provisions began to apply on February 2, 2025, followed by governance rules and obligations for general-purpose AI models on August 2, 2025. Although August 2, 2026, remains the regulation's general date of application, the requirements applicable to high-risk AI systems under Chapter III, Sections 1 to 3, have been postponed. Following Regulation (EU) 2026/1744, these requirements apply from December 2, 2027, to systems classified as high-risk under Article 6(2) and Annex III, which may include AI systems used in health care–related essential public services, and from August 2, 2028, to systems classified under Article 6(1) and Annex I, including AI systems that constitute or are safety components of regulated products, such as medical devices subject to third-party conformity assessment.

Table 1. Milestones in the implementation timeline of the European Union AI Act relevant to health care regulatory sandboxes.

Key date	Implementation milestone
February 2, 2025	The first provisions of the AI Act became applicable, including prohibited AI practices and AI literacy requirements.
August 2, 2025	Governance rules and obligations for general-purpose AI models became applicable.
August 2, 2026	The AI Act reached its general date of application, although specific requirements for high-risk AI systems remain subject to later deadlines.
August 2, 2027	Member states must ensure that at least one national AI regulatory sandbox is operational.
December 2, 2027	Requirements for certain high-risk AI systems become applicable.
August 2, 2028	Requirements for high-risk AI systems embedded in or constituting regulated products become applicable.

The progressive implementation of the regulation constitutes a unique opportunity to identify and address, in a gradual and evidence-based manner, the difficulties that may arise during its application. The key date for this Viewpoint is August 2, 2027, by which member states must ensure that at least one national AI regulatory sandbox is operational. Originally set for August 2, 2026, this deadline was subsequently postponed, reflecting the practical challenges involved in designing regulatory sandboxes across member states.

Although it is frequently argued that the AI Act may hinder innovation, this interpretation reflects a conceptual conflation between research activities and market regulation. The AI Act does not restrict the development of or experimentation with AI systems. Instead, it regulates the conditions under which such systems may be placed on the market or put into service within the EU. Its scope is therefore focused on commercialization and high-risk deployment rather than on upstream scientific or technological innovation.

The AI Act establishes a risk-based classification framework for AI systems. Within this framework, "minimal-risk" systems are subject to very limited regulatory obligations and typically include applications such as ethically designed video games. Second, "limited-risk" systems must comply with specific transparency requirements, including the obligation to inform users when content has been generated or modified using AI. Third, "high-risk" systems are subject to significantly more stringent obligations, encompassing enhanced transparency, human oversight, risk management procedures, and postmarket monitoring, and include, among others, remote biometric identification technologies and certain autonomous vehicle applications. Finally, "prohibited" AI systems are those considered to pose an unacceptable risk to safety, human life, or fundamental rights and are therefore banned from being placed on the market or put into service within the EU.

Legal Uncertainty and Implementation Challenges of the AI Act in Health Care

The health care sector presents intrinsic characteristics—such as high-stakes clinical decision-making, the systematic processing of sensitive data, and the preservation of clinical autonomy—that complicate the implementation of AI governance. Accordingly, digital health AI systems may fall within the high-risk categories established by the AI Act (Annex III). At the same time, the regulation provides narrowly tailored exceptions for specific health care–related contexts, including those referred to in Recital 29 and Article 5(1)(f). Within this framework, the regulation relies predominantly on conformity assessments conducted by providers, with comparatively limited independent external oversight. Such self-assessment mechanisms may be vulnerable to technical misinterpretation or informational asymmetries. This concern becomes particularly salient when health care professionals or hospitals themselves qualify as AI providers [3,4] given their often limited organizational and technical capacity for AI governance.

Moreover, the AI Act uses a computational threshold of 10^{25} floating point operations per second (FLOPS) as a presumption of high-impact capabilities for general-purpose AI models and, consequently, as a criterion for their classification as models with systemic risk. However, computational scale alone may not fully capture the specific risks associated with health care applications built on such models [5,6].

Regulatory complexity is further amplified by the interaction between the AI Act and the General Data Protection Regulation (GDPR) [7]. In health care, compliance must satisfy AI-specific and data protection requirements. Previous analyses also note limited attention in the AI Act to transparency, international data transfers, and accountability [8-10].

Parallel legislative initiatives such as the AI Liability Directive have been proposed to reinforce patient protection and facilitate access to remedies [3,11]. Complementary proposals include digital information instruments such as the AI Passport [12] intended to enhance transparency and mitigate risks in clinical AI deployment.

The AI Act incorporates concepts whose practical interpretation may remain uncertain, most notably "vulnerability." Article 5 addresses vulnerability in the context of exploitation based on factors such as age, disability, or economic status but does not provide detailed operational criteria for identifying vulnerable groups. This lack of specificity has attracted academic criticism. Furthermore, the obligation to report serious incidents referred to in Recital 96 may be affected by divergent national interpretations of what constitutes a "public service," particularly in health care. Such inconsistencies risk undermining uniform application across member states and have been widely discussed in the literature [13-15].

Finally, these structural and conceptual ambiguities intersect with ongoing debates on algorithmic transparency and nondiscrimination in medical AI. The AI Act's requirement of "appropriate transparency" for high-risk systems remains insufficient in clinical contexts, where explainability is not merely a regulatory formality but a prerequisite for informed medical decision-making [6]. The absence of explicit mandates regarding established interpretability techniques limits the practical enforceability of transparency obligations [14]. Similarly, scholarship has addressed unresolved questions concerning medical liability [16] and the amplification of gender biases in health care AI teams [17]. The regulation's reliance on broadly framed requirements to implement "appropriate measures" against discrimination lacks the technical granularity necessary to address complex bias patterns in medical datasets, particularly those affecting underrepresented demographic groups.

This section highlights the complexity of health care under the AI Act and the need for a nuanced regulatory approach. As noted in the work by Gilbert [18], no law can foresee all possible scenarios. Therefore, it is a shared responsibility to design adaptive solutions and contribute to the AI Act's evolution so that its intended objectives can be effectively achieved.

How the AI Act Combines Harmonization With Member State Autonomy

The AI Act establishes a harmonized regulatory framework across the EU while simultaneously embracing differences across member states. As a European regulation, it is directly applicable in all member states without requiring transposition into national law. This means that, from the date of application, national authorities, companies, health care institutions, and AI developers are legally obliged to comply with its provisions even in the absence of additional national legislation.

The regulation follows a progressive implementation timeline, with different provisions becoming applicable in stages until full applicability is reached. During this transitional period, a regulatory gap may emerge in which not all obligations are yet enforceable. In parallel, supervisory bodies begin to operate and issue guidance to clarify permissible practices, thereby reducing uncertainty and supporting early compliance. At the European level, the AI Office [19] within the European Commission and the European Artificial Intelligence Board coordinate the implementation of the regulation, contributing to the alignment of regulatory practices across member states. More broadly, this design allows the AI Act to evolve through practical experience, consistent with the regulation of emerging technologies, where innovation often outpaces legislation.

Although the AI Act does not require national transposition, member states retain the ability to adopt complementary measures to operationalize its provisions. This occurs for several reasons. First, the regulation requires the designation or creation of competent authorities, supervisory bodies, and market surveillance mechanisms at the national level. Second, member states may define procedural aspects such as inspection regimens, enforcement practices, and administrative sanctions. Third, they may introduce sector-specific rules to adapt the regulation to particular domains such as health care, including provisions related to clinical evaluation, hospital deployment, and the governance of medical data.

Within this framework, regulatory sandboxes play a central role by enabling the supervised testing of AI systems prior to their market deployment, particularly for high-risk applications such as those in health care. Article 57 of the AI Act requires member states to ensure that at least one AI regulatory sandbox is established at the national level and becomes operational before August 2, 2027. The regulation explicitly provides for this instrument to facilitate technological experimentation, enable early risk assessment, and support responsible innovation under public supervision. In practice, a regulatory sandbox allows a company, hospital, or research institution to submit an AI system for testing in a controlled environment authorized by the competent authority, where its performance and its potential impact on safety, fundamental

rights, data protection, transparency, and nondiscrimination are assessed.

However, the decentralized implementation of regulatory sandboxes may generate differences in evaluation methodologies, supervisory practices, and access conditions across member states. This may lead to regulatory fragmentation, unequal opportunities for experimentation, and divergent approaches to the evaluation of health care AI systems.

Member states with smaller health systems may therefore face greater difficulties in applying complex testing and documentation requirements. As a result, similar health care AI systems could be evaluated differently across national sandboxes. Rather than pursuing further binding legal harmonization, greater practical convergence could be promoted through common evaluation criteria, transparency standards, and supervisory guidance. Some of the criteria proposed in this Viewpoint are intended to contribute to this convergence in the health care context. European mechanisms such as EU regulatory sandboxes for AI could help translate these proposals into practice by providing shared methodological frameworks, technical guidance, capacity-building measures, and cooperation among national authorities.

Recommendations for the Design and Implementation of AI Regulatory Sandboxes in Health Care

Overview

The preceding discussion highlights several regulatory uncertainties that may affect the implementation of AI regulatory sandboxes in health care. Building on these challenges, the following subsections present practical proposals for AI regulatory sandboxes in health care under the EU AI Act, focusing on technical and legal difficulties that may emerge when medical AI systems are tested under regulatory supervision. These recommendations focus on model assessment, explainability, dataset governance, federated learning (FL), and AI agents. They address what, in our view, are among the main emerging regulatory and technical challenges in the design and implementation of health care AI regulatory sandboxes.

Document Model Compression Methods That May Affect Risk Classification

As previously discussed, the AI Act establishes a computational threshold of 10^{25} FLOPS as a presumption of high-impact capabilities for general-purpose AI models and, consequently, as a criterion for identifying models with systemic risk. FLOPS quantify the amount of computation performed during model training. In general, models with larger numbers of parameters tend to require greater computational resources during training and, consequently, a higher number of FLOPS. In neural network architectures, these parameters typically correspond to the weights and biases learned during training.

However, recent advances in model design have led to the emergence of architectures such as tensor neural networks [20], which enable a substantial reduction in the number of parameters while preserving comparable functional performance. When such compression techniques are incorporated before or during training, they may reduce the computational resources required to train the model. Consequently, an architecture that would otherwise require training compute above the FLOPS threshold may remain below it when trained using these techniques.

For example, a large language model exceeding the 10^{25} FLOPS threshold could be integrated into a hospital information system to summarize patient records. Compression methods could subsequently produce a smaller model that preserves comparable performance for this task while falling below the computational threshold. The model would therefore no longer automatically meet the compute-based presumption of systemic risk even though clinically relevant risks could remain, such as omitting an allergy, medication, or previous diagnosis from a patient summary. A similar situation could arise with compressed multimodal models used to analyze medical images, as well as with other AI-based medical tools that retain clinically significant capabilities despite substantial reductions in computational complexity.

In light of this, we recommend that developers explicitly disclose whether a given model relies on tensor-based architectures or other forms of parameter compression. Where such techniques are used, developers should also report both the original number of parameters and the reduced parameter count after compression. In a health care regulatory sandbox, this information could be considered together with clinically relevant performance to determine whether compression has altered the model's practical capabilities or risks before deployment.

Promote Explainability Approaches That Reflect Dataset Characteristics

According to Article 13 of the AI Act, "High-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured with a view to achieving compliance with the relevant obligations of the provider and deployer" [2]. The regulation does not impose specific technical requirements for explainability, such as the use of commonly cited methods in the literature, including Shapley additive explanations [21] or local interpretable model-agnostic explanations [22]. Instead, it adopts a functional approach centered on ensuring interpretability of outputs in context.

In this regard, we recommend selecting explainability approaches according to the clinical question and the characteristics of the underlying dataset. Local interpretable model-agnostic explanations is primarily used to explain individual predictions (eg, why a model assigns a high risk score to a particular patient), whereas Shapley additive

explanations can provide both individual and aggregated feature importance across the patient population (eg, which clinical variables are generally most influential in predicting disease progression).

However, when the dataset is heterogeneous and the aim is to determine whether the relevance of particular variables changes across different regions of the patient population, complementary sensitivity analysis approaches may be more informative (eg, identifying a variable with moderate overall importance that becomes highly influential in a specific subgroup of patients).

The sensitivity analysis method developed in the study by Pizarroso et al [23] illustrates this by revealing variables with moderate average influence that become highly influential in specific regions of the dataset. In its application to Parkinson disease progression, sex showed the highest global importance and a comparatively consistent influence across patients, whereas jitter (absolute) and pitch period entropy became particularly influential in specific patient cohorts, a pattern that the authors suggested could be related to different Parkinson disease subtypes.

Adopt Inclusive and Bias-Aware Documentation for Health Care Datasets

The AI Act imposes explicit data governance obligations on high-risk AI systems that rely on training data. Under Article 10(1), such systems must be developed based on training, validation, and testing datasets that comply with the quality requirements set out in Articles 10(2) to 10(5). More specifically, Article 10(2)(f) requires an examination of possible biases that may affect health and safety, negatively impact fundamental rights, or lead to unlawful discrimination, while Article 10(2)(g) requires appropriate measures to detect, prevent, and mitigate those biases. In addition, Article 10(3) requires datasets to be relevant; sufficiently representative; and, to the extent possible, free of errors and complete in light of their intended purpose. Although these requirements are especially important in health care, their deliberately general formulation creates uncertainty when medical AI systems are trained on complex or potentially unbalanced datasets.

In this context, we recommend that dataset documentation and auditing explicitly identify potential sources of bias, with particular attention to representation bias in clinically relevant subgroups [24]. Where adequate representation cannot be achieved, this limitation should be clearly justified in the technical documentation as otherwise it may distort downstream clinical decision-making [25]. We also recommend the use of inclusive and precise language in both dataset and model documentation, especially regarding sex and gender variables, to avoid forms of exclusion that may affect the lesbian, gay, bisexual, transgender, and queer community and undermine both representational fairness and clinical validity.

Classify FL Parameter Updates as Personal Data Under the GDPR

Since its original proposal in 2016 [26], FL has gained increasing relevance in AI development. Its main advantage lies in allowing a model to be trained across local datasets without transferring those datasets between institutions. This is particularly valuable in health care, where data are often large and highly sensitive [27,28]. At the same time, because the local datasets remain inaccessible, assessing potential sources of bias may become more difficult. This limitation raises transparency and explainability concerns and supports the need for documentation on how each local dataset was obtained [29].

However, FL also presents important limitations [30]. In addition to the risk of malicious local modifications that may introduce bias, FL is vulnerable to privacy attacks aimed at inferring information about local datasets from the parameter updates transmitted to the central server. This raises a legal question under Article 4(1) of the GDPR, which defines personal data as “any information relating to an identified or identifiable natural person.” In this context, the relevant issue is whether information contained in local datasets can be reconstructed from trained local models or from the exchanged parameter updates.

In this respect, the TechDispatch report on FL states that there is a risk that part of the personal data used in the training could be extracted from the resulting models. Thus, an assessment should be performed on a case-by-case basis to determine whether models need to be considered as personal data [30]. Similarly, the European Data Protection Board [31] notes that the required documentation should include “information on the technical and organisational measures taken while designing the AI model to reduce the likelihood of identification, including the threat model and risk assessments on which these measures are based,” as well as “the documentation demonstrating the AI model’s theoretical resistance to re-identification techniques” and “metrics on the likelihood of re-identification based on the current state-of-the-art.”

Because this assessment depends on a constantly evolving state of the art, the risk of identification can rarely be considered entirely absent. This may only be achievable when the original dataset is anonymous from the outset or has been effectively anonymized beforehand, although it may reduce model performance in some cases. For regulatory sandbox purposes, we therefore propose a precautionary approach. Whenever the dataset held at each local node is not anonymous, the parameter updates exchanged within the FL architecture should be treated as personal data even if the design phase tests indicate only a low level of risk.

For example, several hospitals could use FL to jointly train a model for predicting clinical deterioration. If the exchanged parameter updates are treated as personal data, their transmission, storage, aggregation, and access would also fall within the GDPR governance framework. During sandbox assessment, participating institutions could therefore

define the purposes and responsibilities associated with these operations and assess appropriate safeguards before deployment.

Ensure Traceable Access to Clinical Records by AI Agents in Health Care

The AI Act was originally conceived for comparatively static AI systems, yet recent years have seen the rapid emergence of AI agents. In this discussion, we use this term to refer to systems capable of executing multistep tasks, accessing external tools, and autonomously performing certain actions. These systems are becoming increasingly relevant in health care [32-34] and have shown potential performance gains over state-of-the-art models [35]. This evolution raises important regulatory challenges under Article 14 of the AI Act, which requires human oversight measures proportionate to the risks, level of autonomy, and context of use of high-risk systems. At the same time, the regulatory treatment of AI agents remains at a preliminary stage as current European Commission guidance acknowledges that the concept is still evolving and that the relevant regulatory considerations are not yet fully settled.

Building on the human oversight principles established in Article 14, we propose operational safeguards for testing AI agent access to clinical records within health care regulatory sandboxes. Each access should be recorded as an auditable event and linked to a defined purpose, an identifiable nonhuman identity, and a responsible human or organizational authority. For example, an AI agent used to coordinate follow-up care after hospital discharge could access appointment schedules, referral information, and pending test results. During sandbox testing, it could be verified that the agent accesses only the information required for this task, that every access is logged, and that responsibility remains attributable to a human supervisor. Access to unrelated records or unnecessary information should trigger review and accountability mechanisms equivalent to those triggered by unjustified access by human professionals.

The recommendations developed in the preceding sections are consolidated into a practical overview (Table 2) for the design and implementation of health care AI regulatory sandboxes.

Table 2. Principal regulatory and technical challenges of the European Union (EU) AI Act in health care and proposed measures for regulatory sandboxes.

EU AI Act limitations	Proposed recommendations
Model compression may reduce computational complexity without reducing clinical impact	Require developers to disclose the use of parameter compression techniques (including tensor-based architectures)
Lack of specific technical requirements for explainability	Select explainability approaches according to the clinical question and the characteristics of the datasets
Health care datasets may contain representation biases	Request dataset documentation and auditing to identify bias, particularly underrepresentation of clinically relevant groups, and justify unavoidable limitations
Federated learning parameter updates may allow information about patients to be inferred	Treat federated learning parameters as personal data whenever local datasets are not anonymous, even when the risk of reidentification is low
AI agents may access clinical records without sufficient traceability and accountability	Ensure that AI agent access to clinical records is logged as auditable events linked to a defined purpose and responsible authority

Conclusions

The EU AI Act provides a general legal framework for the governance of AI in health care, but its practical effectiveness will depend on how its requirements are operationalized in concrete testing and deployment settings. In this regard, the recommendations proposed in this Viewpoint are intended to support the design of health care regulatory sandboxes that are more robust, transparent, and responsive to the specific challenges posed by medical AI systems. In particular, they address issues that are likely to become increasingly relevant in practice, including model compression, explainability, dataset bias, FL, and human oversight of AI agents. Taken together, these proposals aim to facilitate earlier risk identification, improve legal and technical clarity, and strengthen the supervised development of trustworthy AI in clinical contexts.

Acknowledgments

The authors would like to express their gratitude to Ángel Canal Alonso and Federico Bueno de Mata for their valuable comments and insights, which contributed to improving the quality of this paper.

Generative AI Use

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the Generative AI Delegation Taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: proofreading and editing and translation. The GenAI tool used was ChatGPT-Auto (OpenAI). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

Financial support from the Tecnovitae Foundation is gratefully acknowledged (Beca CTS 2024).

Authors' Contributions

Conceptualization: MFC (lead), FCV (equal), ALT (supporting)
Formal analysis: MFC (lead), FCV (equal)
Funding acquisition: FCV (lead), MFC (supporting), ALT (supporting)
Methodology: MFC (lead), FCV (equal), ALT (supporting)
Project administration: FCV (lead), MFC (supporting)
Supervision: MFC (lead), FCV (equal)
Writing—original draft: MFC (lead)
Writing—review and editing: MFC (lead), FCV (supporting)

Conflicts of Interest

None declared.

References

1. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization; 2021. URL: https://iris.who.int/server/api/core/bitstreams/f780d926-4ae3-42ce-a6d6-e898a5562621/content [Accessed 2026-10-26]
2. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (text with EEA relevance). European Union. 2024. URL: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng [Accessed 2026-10-17]
3. Solaiman B, Malik A. Regulating algorithmic care in the European Union: evolving doctor-patient models through the Artificial Intelligence Act (AI-Act) and the liability directives. Med Law Rev. Jan 4, 2025;33(1):fwae033. [doi: 10.1093/medlaw/fwae033] [Medline: 39257157]
4. Ranjbar A, Mork EW, Ravn J, et al. Managing risk and quality of AI in healthcare: are hospitals ready for implementation? Risk Manag Healthc Policy. 2024;17:877-882. [doi: 10.2147/RMHP.S452337] [Medline: 38617593]
5. Hooker S. On the limitations of compute thresholds as a governance strategy. arXiv. Preprint posted online on Jul 8, 2024. [doi: 10.48550/arXiv.2407.05694]
6. Panigutti C, Hamon R, Hupont I, et al. The role of explainable AI in the context of the AI Act. In: FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery; 2023:1139-1150. [doi: 10.1145/3593013.3594069]
7. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (text with EEA relevance). European Union. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [Accessed 2026-03-31]
8. Cascella M, Shariff MN, Viswanath O, Leoni ML, Varrassi G. Ethical considerations in the use of artificial intelligence in pain medicine. Curr Pain Headache Rep. Jan 6, 2025;29(1):10. [doi: 10.1007/s11916-024-01330-7] [Medline: 39760779]
9. Szeftel D, Jouhet V, Duluc G, et al. Health data: regionalised access is a priority challenge for building a secure, transparent and innovative national health data repository. Therapies. 2025;80(1):125-134. [doi: 10.1016/j.therap.2024.12.008]
10. Van Roijen D. The European digital identity wallet: a healthcare perspective. Blockchain Healthc Today. 2024;7. [doi: 10.30953/bhty.v7.344] [Medline: 39649413]
11. van Kolfschooten H, van Oirschot J. The EU Artificial Intelligence Act (2024): implications for healthcare. Health Policy. Nov 2024;149:105152. [doi: 10.1016/j.healthpol.2024.105152] [Medline: 39244818]
12. García-Gómez JM, Blanes-Selva V, Doñate-Martínez A. Proposing an AI passport as a mitigating action of risk associated to artificial intelligence in healthcare. Stud Health Technol Inform. Aug 22, 2024;316:547-551. [doi: 10.3233/SHTI240472] [Medline: 39176800]
13. Schmidt J, Schutte NM, Buttigieg S, et al. Mapping the regulatory landscape for artificial intelligence in health within the European Union. NPJ Digit Med. Aug 27, 2024;7(1):229. [doi: 10.1038/s41746-024-01221-6] [Medline: 39191937]
14. Busch F, Kather JN, Johner C, et al. Navigating the European Union Artificial Intelligence Act for healthcare. NPJ Digit Med. Aug 12, 2024;7(1):210. [doi: 10.1038/s41746-024-01213-6] [Medline: 39134637]
15. Galli F, Novelli C. The many meanings of vulnerability in the AI Act and the one missing. BioLaw J. 2024;1S:53-72. [doi: 10.15168/2284-4503-3302]

16. Martín-Noguerol T, López-Úbeda P, Luna A. AI in radiology: legal responsibilities and the car paradox. Eur J Radiol. Jun 2024;175:111462. [doi: 10.1016/j.ejrad.2024.111462] [Medline: 38608500]
17. Piva L. Vulnerability in the age of artificial intelligence: addressing gender bias in healthcare. BioLaw J. 2024;1:169-178. [doi: 10.15168/2284-4503-3310]
18. Gilbert S. The EU passes the AI Act and its implications for digital medicine are unclear. NPJ Digit Med. May 22, 2024;7(1):135. [doi: 10.1038/s41746-024-01116-6] [Medline: 38778162]
19. Commission decision establishing the European AI Office. European Commission. 2024. URL: https://digital-strategy.ec.europa.eu/en/library/commission-decision-establishing-european-ai-office [Accessed 2026-03-31]
20. Socher R, Chen D, Manning CD, Ng AY. Reasoning with neural tensor networks for knowledge base completion. In: NIPS'13: Proceedings of the 27th International Conference on Neural Information Processing Systems. Curran Associates Inc; 2013:926-934. [doi: 10.5555/2999611.2999715]
21. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc; 2017:4768-4777. [doi: 10.5555/3295222.3295230]
22. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. In: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery; 2016:1135-1144. [doi: 10.1145/2939672.2939778]
23. Pizarroso J, Alfaya D, Portela J, Muñoz A. Metric tools for sensitivity analysis with applications to neural networks. Appl Soft Comput. Aug 2025;180(C). [doi: 10.1016/j.asoc.2025.113300]
24. Hasanzadeh F, Josephson CB, Waters G, Adedinsewo D, Azizi Z, White JA. Bias recognition and mitigation strategies in artificial intelligence healthcare applications. NPJ Digit Med. Mar 11, 2025;8(1):154. [doi: 10.1038/s41746-025-01503-7] [Medline: 40069303]
25. Shi Y, Ma J, Li S, et al. Sex difference in human diseases: mechanistic insights and clinical implications. Signal Transduct Target Ther. Sep 10, 2024;9(1):238. [doi: 10.1038/s41392-024-01929-7] [Medline: 39256355]
26. McMahan HB, Moore E, Ramage D, Aguera y Arcas B. Communication-efficient learning of deep networks from decentralized data. Proc Mach Learn Res. 2017;54:1273-1282. URL: https://proceedings.mlr.press/v54/mcmahan17a.html [Accessed 2026-10-17]
27. Brauneck A, Schmalhorst L, Kazemi Majdabadi MM, et al. Federated machine learning, privacy-enhancing technologies, and data protection laws in medical research: scoping review. J Med Internet Res. Mar 30, 2023;25:e41588. [doi: 10.2196/41588] [Medline: 36995759]
28. Abbas SR, Abbas Z, Zahir A, Lee SW. Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration. Healthcare (Basel). 2024;12(24):2587. [doi: 10.3390/healthcare12242587] [Medline: 39766014]
29. Chik WB, Gamper F. Can federated learning solve AI's data privacy problem?: A legal analysis. Rutgers Law J. 2024;52:1-40. URL: https://ink.library.smu.edu.sg/sol_research/4517/ [Accessed 2026-10-17]
30. TechDispatch: federated learning. Publications Office of the European Union; 2025. URL: https://www.edps.europa.eu/system/files/2025-06/techdispatch_federated-learning_en.pdf [Accessed 2026-10-26]
31. Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models. European Data Protection Board. Dec 18, 2024. URL: https://www.edpb.europa.eu/documents/opinion-of-the-board-art-64/opinion-282024-on-certain-data-protection-aspects-related-to_en [Accessed 2026-04-02]
32. Hinostroza Fuentes VG, Karim HA, Tan MJ, AlDahoul N. AI with agency: a vision for adaptive, efficient, and ethical healthcare. Front Digit Health. 2025;7:1600216. [doi: 10.3389/fdgth.2025.1600216] [Medline: 40400543]
33. Prakash C, Sisodia A, Lind M. A systematic review of agentic AI in healthcare: an evidence-informed seven-principle framework. J Comput Theor Appl. 2026;4(1):43-73. [doi: 10.62411/jcta.16171]
34. Ferber D, El Nahhas OS, Wölflein G, et al. Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology. Nat Cancer. Aug 2025;6(8):1337-1349. [doi: 10.1038/s43018-025-00991-6] [Medline: 40481323]
35. Gorenshtein A, Omar M, Glicksberg BS, Nadkarni GN, Klang E. AI agents in clinical medicine: a systematic review. medRxiv. Aug 26, 2025:2025.08.22.25334232. [doi: 10.1101/2025.08.22.25334232] [Medline: 40909853]

Abbreviations

EU: European Union
FL: federated learning
FLOPS: floating point operations per second
GDPR: General Data Protection Regulation

Edited by Fida Dankar; peer-reviewed by Meng-Hsun Tsai, Valentina Palama; submitted 18.Apr.2026; final revised version received 29.Aug.2026; accepted 08.Sep.2026; published 30.Oct.2026

Please cite as:
Flores Ceballos M, Cagigas Villar F, López Tejerina A
Rethinking Health Care Regulatory Sandboxes Under the European Union AI Act
JMIR AI 2026;5:e98653
URL: https://ai.jmir.org/2026/1/e98653
doi: 10.2196/98653

© María Flores Ceballos, Fernando Cagigas Villar, Alberto López Tejerina. Originally published in JMIR AI (https://ai.jmir.org), 30.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on https://www.ai.jmir.org/, as well as this copyright and license information must be included.