ACTA MÉDICA PORTUGUESA
A Revista Científica da Ordem dos Médicos

Reusing Patient Data to Train Artificial Intelligence for Medical Devices in the European Union

A Reutilização dos Dados dos Doentes para Treinar Inteligência Artificial para Dispositivos Médicos na União Europeia

Yuri LADEIA[1,2], Nuno SOUSA[2,3,4]
Acta Med Port 2025 May;38(5):285-287 ▪ https://doi.org/10.20344/amp.22509

Keywords: Artificial Intelligence; Confidentiality/legislation & jurisprudence; European Union; Health Records, Personal
Palavras-chave: Inteligência Artificial; Confidencialidade/legislação e jurisprudência; Registos de Saúde Pessoal; União Europeia

Health data have enabled significant breakthroughs in medicine. A recent endorsement is the Nobel Prize in Chemistry awarded to the AlphaFold project, the artificial intelligence (AI) tool developed by Google DeepMind to predict the structure of proteins, thus enhancing advances in deep learning and experimental approaches.[1]

Consequently, the diffusion of AI – "machine-based systems with different levels of autonomy and post-deployment adaptation" – in the life sciences has rendered the processing of training, validation and testing health data (hereafter collectively 'training data') increasingly vital, according to the definitions provided in article 3(1)(29), (30), (31), (32) and (37) of the European Union (EU) regulation 2024/1689, Artificial Intelligence Act (AIA),[2] which lays down harmonized rules on AI.

Provided IA is intended to be placed on the market, either as a safety component or itself as a product, it is a high-risk system. This is the case of AI as a medical device (AIaMD), which generally relies on the processing of high-quality and representative training data to safeguard performance and safety.[3] In this regard, the challenge is harmonizing AI and MD requirements with privacy rights, as per article 6(1) subparagraphs (a) and (b), and section A (11) of annex I to the AIA.[4]

Data protection is enshrined as a fundamental right in several international treaties in Europe and the EU, including in the European Convention on Human Rights, and enhanced mainly by Regulation (EU) 2016/679 – General Data Protection Regulation (GDPR)[5] – which safeguards the privacy of data subjects.[6] The GDPR provides agnostic and strict rules for personal data processing, regardless of sector, including healthcare. The safety of MD is affected by the requirements of Regulation (EU) 2017/745 — the Medical Devices Regulation (MDR)[7] — and Regulation (EU) 2017/746[8] — *in vitro* Diagnostic Medical Devices Regulation (IVDR). Member States' laws complement these EU Regulations.

To understand the impact in practice, for the sake of clarity, it is essential to bear in mind the following definitions: personal data (any directly or indirectly identifiable information relating to an individual person, including pseudonymized non-anonymous data – e.g., encrypted clinical episodes, user numbers, notes and more); data processing (series of processing activities, from collection, structuring, storage, interconnection and erasure, e.g.) and data controller (legal person who defines the means and purposes of the processing, e.g. manufactures, deployers, sponsors of a clinical study), under article 4 (1) (2) (7) and (8) of the GDPR.[9]

Once the key concepts have been established, the next step is to safeguard the interest in setting up datasets to train AIaMD overall. What is the most appropriate legal basis of the GDPR, and how does it make his interest legitimate, without jeopardizing biomedical innovation and patients' rights?

Putting the issue in perspective

The application of AIaMD has raised concerns about its reliability and safety. A 2019 study compared the performance of AI with that of human doctors in diagnosing diseases, revealing that although AI needs more data and refinement, it shows promising results, especially in fields like image recognition. However, challenges persist. In a study carried out in Uganda in 2023, an AI algorithm showed poor performance in identifying dermatological diseases in dark skin compared to light skin, e.g.[3] Diversifying training data throughout the device's lifecycle remains critical for continuous safety and performance monitoring across a wide

1. Centre for Justice and Governance. University of Minho Law School. Braga. Portugal.
2. Medicine Centre Association P5 (ACMP5). University of Minho Medical School. Braga. Portugal.
3. ICVS/3B's. Escola de Medicina. Universidade do Minho. Braga. Portugal.
4. Agency for Clinical Research and Biomedical Innovation (AICIB). Lisbon. Portugal.
✉ **Autor correspondente:** Yuri Ladeia. rodrigueslaw@outlook.pt
Recebido/*Received*: 28/10/2024 - **Aceite/*Accepted*:** 27/01/2025 - **Publicado/*Published*:** 02/05/2025

spectrum of scenarios of AI application in healthcare.[10]

The digitalization of healthcare is generating a massive amount of data beyond human processing capacity. By increasing cognitive capacities, AI can support doctors in better-informed and precise decision-making. However, the success of AIaMD depends on the representativeness of the training data,[11] especially as the reliability of AI capabilities is fundamental to patient care, and for the trust and accountability of the doctor as the human operator and overseer of the AI system, as established in article 14(2), and recitals 66 and 73 of the AIA.

The World Health Organization 2021 guidelines report that "generating evidence for AI-Based MD" "emphasizes using high-quality and representative data". The recommendations underline the need for a robust database to ensure the safe performance of AIaMD in medical practice.[12]

From the data protection angle, processing training data for AIaMD requires a consistent legal basis. The GDPR provides several without hierarchy, including the subject's consent and the data controller's legitimate interest – manufacturer or deployer, for example.

The development and deployment data processing phases of artificial intelligence involve separate purposes (thus constituting separate processing activities). In life sciences, consent is often applied as a legal basis, as it provides data subjects with greater transparency and control. However, consent may not be the optimal data protection approach, particularly in large-scale cases where it is unfeasible to ensure individual consent and its conditions.[13]

Hence, if the conditions for consent are not met under GDPR article 7 at the data collection development phase, then in subsequent deployments, the lawfulness will be affected, thus contradicting the continuous monitoring and improvement requirements.[14] For instance, when data processing in the deployment phase is based on controller or third-party legitimate interest, in the event that the original data collection was unlawful (for example, consent conditions not appropriate due not informing data subjects about secondary data processing purposes), when assessing the legitimate interest adequacy as legal basis (data subjects may not expect such further processing, for example), it could be poisoned and compromised, according to "Opinion 28/2024" of the European Data Protection Board (EDPB) "on AI models for the processing of personal data".[15]

On the other hand, legitimate interest may provide a more appropriate legal basis in both phases, as long as it can be demonstrated that data processing is indispensable for safety and performance intended purposes of the AIaMDs. This quality and safety based legitimate interest must be carefully balanced so the rights of the data subjects are not jeopardized.

Strategic perspective

The GDPR may allow the processing of sensitive health-related personal data, as long as there are valid legal bases. The training data processing must ensure safety and performance, balancing the privacy of the data subjects. The decision between consent *versus* legitimate interest depends on the specific scenario, and a meticulous analysis is essential to determine the most appropriate legal basis on a case-by-case basis.

The data controller's legitimate interest relies on three cumulative conditions. As a balancing test, the interest should be lawful – not contrary to the law; clearly and precisely articulated – not general; real and present – not speculative, as per the "Guidelines 1/2024 on the processing of personal data based on Article 6(1) (f) GDPR" by EDPB.[15]

In the AIaMD development and deployment, under Quality Management Systems, those requirements are in place when the common purpose is related to ensuring high standards of quality, safety and privacy, as per article 10 AIA and the complementary legal basis under Article 9 (2) subparagraph (i) GDPR, allowing health data processing. It is also a condition for placing and maintaining the product on the EU market, under Article 5 MDR.

Therefore, it is suggested that collecting and (re)using health data (development and deployment) throughout the AIaMD lifecycle could be grounded in the legal basis of legitimate interest; provided this interest is not limited to the economic sense, and also provided there is recognition of this interest by the law, as is the case with security and quality purposes. Although controversies in case law that are not covered in this paper due to space constraints, in general, if the interest is purely material, this legal basis may have no effect.[10]

CONCLUSION

If the AI model is non-anonymous (the probabilistic chances of extracting personal data for development and from queries are not insignificant) as per EDPB "Opinion 28/2024", we suggest that legitimate interest based on quality and safety reasons could reveal a balanced legal basis for health data processing on development and deployment of AIaMDs, as the device's performance and patients' safety are commonly paramount to EU law, suppliers, deployers, doctors, and patients.

AUTHOR CONTRIBUTIONS

The authors contributed equally to this manuscript and approved the final version to be published.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

FUNDING SOURCES

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

1. O'Leary K. AlphaFold gets an upgrade (and a Nobel). Nat Med. 2024;30:3393.
2. European Parliament and Council. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024. Strasbourg: EPC; 2024.
3. Kamulegeya L, Bwanika J, Okello M, Rusoke D, Nassiwa F, Lubega W, et al. Using artificial intelligence on dermatology conditions in Uganda: a case for diversity in training data sets for machine learning. Afr Health Sci. 2023;23:753-63.
4. Stamhuis E. European Data Protection Board: "It depends". How AI can be legitimate under the GDPR. Blogpost for the AI-MAPS project. [cited 2025 Feb 25]. Available from: https://www.eur.nl/en/news/european-data-protection-board-it-depends-how-ai-can-be-legitimate-under-gdpr.
5. European Parliament and Council. 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. Strasbourg: EPC; 2016.
6. Gonçalves AS de S, Costa TB da. Os dados pessoais dos utilizadores das plataformas colaborativas à luz do RGPD. In: Economia Colaborativa. Braga: UMinho Editora/JusGov; 2023. p.365-96.
7. European Parliament and Council. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending directive 2001/83/EC, regulation (EC) no 178/2002 and regulation (EC) No 1223/2009 and repealing Council directives 90/385/EEC and 93/42/EEC. Strasbourg: EPC; 2017.
8. EUR-Lex. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 april 2017 on in vitro diagnostic medical devices and repealing directive 98/79/EC and commission decision 2010/227/EU. 2017. [cited 2024 Aug 23]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0746.
9. Shen J, Zhang CJ, Jiang B, Chen J, Song J, Liu Z, et al. Artificial intelligence versus clinicians in disease diagnosis: systematic review. JMIR Med Inform. 2019;7:e10010.
10. Khalifa M, Albadawy M. Artificial intelligence for clinical prediction: exploring key domains and essential functions. Comput. Methods Programs Biomed. Update. 2024;5:1-11.
11. World Health Organization. Generating evidence for artificial intelligence based medical devices: a framework for training validation and evaluation. 2021. [cited 2024 Aug 23]. Available from: https://www.who.int/publications/i/item/9789240038462.
12. Schermer BW, Custers B, van der Hof S. The crisis of consent: how stronger legal protection may lead to weaker consent in data protection. Ethics Inf Technol. 2014;16:171-82.
13. Kramcsák PT. Can legitimate interest be an appropriate lawful basis for processing artificial intelligence training datasets? Comput Law Secur Rev. 2023;48:1-11.
14. European Data Protection Board. Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models. Brussels: EDPB; 2024.
15. European Data Protection Board. Guidelines 1/2024 on processing of personal data based on article 6(1)(f) GDPR. Brussels: EDPB; 2024.