

Document Version

Final published version

Licence

CC BY

Citation (APA)

van Genderen, M. E., van de Sande, D., Hooft, L., Reis, A. A., Cornet, A. D., Oosterhoff, J. H. F., van der Ster, B. J. P., Gommers, D., van den Hoven, J., & More Authors (2024). Charting a new course in healthcare: early-stage AI algorithm registration to enhance trust and transparency. *npj Digital Medicine*, 7(1), Article 119. <https://doi.org/10.1038/s41746-024-01104-w>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Charting a new course in healthcare: early-stage AI algorithm registration to enhance trust and transparency

Michel E. van Genderen, Davy van de Sande, Lotty Hooft, Andreas Alois Reis, Alexander D. Cornet, Jacobien H. F. Oosterhoff, Björn J. P. van der Ster, Joost Huiskens, Reggie Townsend, Jasper van Bommel, Diederik Gommers & Jeroen van den Hoven



AI holds the potential to transform healthcare, promising improvements in patient care. Yet, realizing this potential is hampered by over-reliance on limited datasets and a lack of transparency in validation processes. To overcome these obstacles, we advocate the creation of a detailed registry for AI algorithms. This registry would document the development, training, and validation of AI models, ensuring scientific integrity and transparency. Additionally, it would serve as a platform for peer review and ethical oversight. By bridging the gap between scientific validation and regulatory approval, such as by the FDA, we aim to enhance the integrity and trustworthiness of AI applications in healthcare.

Fueled by the potential to improve patient outcomes and clinical decision-making, artificial intelligence (AI) is poised to broadly reshape medicine resulting in an exponentially growing number of studies, for example in the field of intensive care medicine¹. This trend is exemplified by the rapid growth of AI-based trials registered at clinicaltrials.gov. since 2009 — 76 trials from 2009 to 2019, with an additional 294 trials in just the next three years². In 2019, the US Food and Drug Administration (FDA) has launched a digital health branch, approving 692 AI-based models to date³. Most of these FDA-approved models, however, are based on evidence from retrospective, single-institution data, often unpublished, rather than robust evidence from clinical trials, the cornerstone of medicine^{4,5}.

Our obligation to ensure responsible AI

AI algorithms are increasingly utilized to assist healthcare providers in clinical decision-making. These AI clinical decision support algorithms derives inputs from various clinical sources, aiding in tasks ranging from classification and computer-aided diagnosis in radiology to clinical prediction models for prognostic or quality purposes⁶. The trustworthiness of such AI algorithms is crucial for their successful integration into clinical practice. In 2020, several authors led an initiative to create an open access database exclusively for FDA approved clinical based AI algorithms⁷. Nonetheless, more detailed reporting is necessary to enhance the

understanding and interpretation of AI outputs, thereby fostering user trust and facilitating the integration of AI into a learning healthcare system⁸. The World Health Organization (WHO) has recently published a set of key principles to augment trust in and adoption of AI in health care, including the imperative to improve transparency by detailing the source code, database, data inputs, and analytical approaches used in AI algorithms⁹. While guidelines like SPIRIT-AI¹⁰, CONSORT-AI¹¹, and DECIDE-AI¹² promote algorithmic information reporting in scientific publications for transparency, they lack specific requirements to translate principles into practice¹³.

To ensure the responsible use of AI algorithms, establishing a supportive infrastructure that builds trust in these systems and mitigates biases during early research, clinical evaluation, and development phases is essential. This concept underpins the European Union's AI Act, aiming to regulate AI use by addressing potential risks to human life¹⁴. Thus we advocate for the mandatory registration of early-stage AI algorithms, drawing parallels to the registration of clinical trials.

Why AI algorithms should be registered

The integrity of clinical trials rests in large part on medical practitioners' ethical obligation to ensure patient health and well-being, including those involved in research. As the research landscape rapidly evolves, the Declaration of Helsinki is subject to changes to safeguard and maintain trust in research¹⁵. The Council of Europe's Helsinki 2019 update conference underscored the need for algorithmic transparency and effective supervisory mechanisms in AI's design, development, and deployment phases¹⁶. These measures are necessary to fulfilling ethical obligations, mitigating algorithmic bias, and fostering trust, thereby maximizing benefits and minimizing risks to human rights. AI trials, like human participant studies, must uphold ethical standards, considering the emerging risks of human-AI interactions, interpretability challenges, and data constraints¹⁷. To ensure a safe translation of AI algorithms into medical practice, it is crucial to understand the design, development, and clinical validation process to infer potential risks of bias and avoiding harm to patients, which would be unethical and could expedite serious negative consequences¹¹. Transparency is needed to assess the quality of AI algorithms for stakeholders and to enable medical end-users and patients.

On the other hand, AI algorithm producers (vendors or industry) may be unwilling to provide training datasets or summary information due to intellectual property (IP) and trade secrecy. The intent here is, however, to strike a balance between disclosing algorithm information and protecting IP to promote greater transparency while allowing entities to safeguard their innovations. For instance, enhancing model transparency by disclosing information on model development, training, and validation datasets, and

Table 1 | Proposed registration information for early stage clinical AI algorithms in healthcare

Item ^a	Description	Generative AI ^b	Non-Generative AI ^b
Name, Version, and AI Model Type	Name of the system, its version, and the type of AI model used (e.g., deep learning, decision tree, etc.)	✓	✓
Training and validation population	Demographics (e.g., age, gender, ethnicity) of the patient population on which the algorithm was trained	✓	✓
Clinical context	Model application (e.g. used for administrative purpose or for instance to predict a specific illness)	✓	✓
Performance Metrics	Performance of the AI system in preclinical development/validation and prior clinical studies (e.g., model discrimination and calibration)	✓	✓
Input Data Types	Types of data used as inputs by the AI system (e.g., images, clinical notes, lab results, etc.)	✓	✓
Data Acquisition and Processing	Process of data acquisition, the steps required for input data entry, the pre-processing procedures applied, and the methodologies employed for handling missing or low-quality data	✓	✓
Output Types and Presentation	Types of outputs generated by the AI system (e.g., predictions, recommendations, etc.) and how these outputs are presented to the users	✓	✓
Registrant Information	Name, affiliation, and contact information of the person or organization that registered the AI system	✓	✓
Foundation model-specific information	Type and version of the foundation model used (e.g., LLM, version: GPT-4 or PaLM 2)	✓	
	Manufacturer or company that developed the foundation model (e.g., OpenAI, Microsoft, Google etc.)	✓	
	Fine-tuning or grounding process used on the foundation model	✓	

AI Artificial Intelligence, LLM Large Language Model, GPT-4 Generative Pre-trained Transformer 4, PaLM 2 Pathways Language Model.

^aItems to be registered have been adapted from DECIDE-AI and CONSORT-AI guidelines.

^bThe terminology “Generative AI” and “Non-Generative AI” specifies if certain data or criteria are relevant to generative AI models, non-generative AI models, or both. The presence of a checkmark (✓) in a column signals that the mentioned data or criteria pertain to that AI category. Generative AI models are capable of producing new content, such as the way Large Language Models (LLMs) can craft text that mimics human writing. On the other hand, Non-Generative AI models are designed to interpret and learn from pre-existing data to make forecasts or decisions. For instance, these models analyze and learn from existing data to make predictions or decisions.

clinical performance is a critical step toward trustworthy AI. This transparency is essential to address AI algorithms’ core components and mitigate potential biases and safety issues. AI algorithm registration should support a dynamic learning healthcare system, allowing for modifications to AI systems post-approval. This iterative design promotes trust and ensures AI algorithm registration aligns with stakeholders’ moral obligation to avoid harm^{18,19}.

Currently, the majority of the 14 available CE-certified AI-based radiology products in Europe lack information on training data collection and population characteristics, and none report potential performance limitations related to bias mitigation characteristics, such as ethnicity and age²⁰. Both are an obstacle to assess the risk of algorithmic bias. An example is the sepsis prediction algorithm developed by Epic (Epic Systems, WI, USA), which, despite its deployment in several U.S. hospitals, faced poor performance during external validation in 27,697 patients due to a lack of transparent information on performance metrics and the dataset²¹. Early registration could mitigate potential harm by mandating the disclosure of key AI algorithm aspects prior to clinical implementation, encouraging the publication of negative results, and preventing publication bias or overly optimistic interpretations of results. This is exemplified by studies that demonstrated that AI was found to reinforce systematic health disparities^{22,23}. Although transparency alone does not ensure bias-free algorithms, it is crucial for identifying and eliminating bias, thereby facilitating continuous improvement and accountability²⁴.

Welcoming AI registration in medicine

The practice of registering clinical trials was initiated decades ago, with the WHO establishing the International Clinical Trials Registry Platform (ICTRP) in 2005 and the World Medical Association’s Declaration of Helsinki mandating prospective registration of all clinical trials since

2008^{5,25}. Clinical trial registration has been effective in logging and providing comprehensive information about experimental clinical interventions, significantly enhancing transparency and reducing reporting bias. Similarly, the recent WHO guidance on large multi-modal models encourages the early-stage registration of AI algorithms to improve “explainability,” for instance, by disclosing performance in internal testing²⁶. However, current databases like EUDAMED, the FDA database, as well as clinical trial registries, lack fields for early stage algorithm or training data information^{20,27}. Given AI algorithms’ potential impact on patient care, traceability and comprehensive documentation of the development process and pre-clinical evaluations are essential. Our proposed set of minimum criteria for an AI algorithm registry aims to fill this gap, requiring registration to encompass the entire model, including data acquisition process details, training data characteristics, model specifications, and information presentation to end-users. (Table 1). This registry does not aim to share code, safeguarding IP, but to ensure that general algorithm information is disclosed, facilitating a safe, transparent, and responsible integration of AI in healthcare. Importantly, the AI system content should not be a concern in terms of patent infringement as only general algorithm information are required. AI algorithms should be registered prior to its deployment in clinical practice and before submitting a trial protocol for ethics approval in preparation for clinical assessment, once the registry is open for enrollment. The registry is designed to capture the lifecycle of AI algorithms in healthcare, recognizing that these models evolve through active learning or subsequent updates with new data. While the focus is initially on the ‘base’ algorithm, the system is intentionally designed to accommodate modifications. The registry should differentiate between minor adjustments, unlikely to impact the AI’s fundamental decision-making process and substantial changes that might affect the model’s performance. Such modifications, including retraining on new data or alterations in algorithmic processing,

Table 2 | Minimal functional requirements for the registry

Requirement	Description
Data quality	- Verification mechanism for registration data accuracy - Assignment of unique identifiers to distinct algorithms - Prevention mechanism for duplicate algorithm registrations - Accessible audit trail and version management
Accessibility	- Accessible to all potential registrants - Freely available to the public - User-friendly for contributions - Electronically searchable - Offered in relevant languages - Open 24/7 for submissions and searches
Integration with other sources	- Enable linking with clinical trial identifiers - Enable linking with published study DOIs - Enable linking with FDA or European Commission device identifiers - Automate data transfer between third parties, reducing redundancy
Technical functionality	- Routine maintenance and updates - Ensure permanence of entries - Efficient data storage and management - Robust data protection against loss and corruption
Governance	- Central global hub for data entry, reducing redundancy - Administered by a non-profit organization like NIH's clinicaltrials.gov

AI Artificial Intelligence, NIH National Institutes of Health, FDA Food and Drug Administration, DOI Digital Object Identifier.

necessitate updates to the registration. Moreover, for models engaged in active learning or subject to frequent updates, we advocate for a mechanism within the registry that allows for the periodic reporting of updated performance metrics, ensuring the registry accurately reflects each algorithm's current capabilities and performance in practical applications. The registry's functional requirements should at least allow for data quality, accessibility, source integration, technical functionality, and governance requirements (Table 2). Specifically because foundation models, i.e. generative AI, such as ChatGPT, released by OpenAI in 2022, differ from well-known general AI models that have the ability to perform specific clinical tasks, such as predicting sepsis²⁸. These generative models, characterized by their training on extensive datasets and the utilization of billions of parameters, demand specific hardware and exhibit a dynamic nature. Despite these variances, it's imperative to trace and log key characteristics to ensure responsible use of AI in clinical decision support²⁹. This is much needed because current uses of generative AI within healthcare are limited by their lack of generalizability and limitations of model details, such as model weights, published due to data privacy concerns²⁷. Our proposed registry, therefore, distinguishes between generative and general AI in terms of required documentation (Table 1), encompassing training data knowledge corpus of the foundation model (such as time period of training, geographical regions, and languages), implemented policies to prevent the dissemination of sensitive input data into foundation models, details about the manufacturer, and software version.

Institutional review boards should consider algorithm registration a prerequisite for approval, and scientific journals could make registration a condition for publication, continuing a tradition of rigorous scientific accountability, as has been done in the past⁵. Healthcare institutions should consider the prerequisite of early registration, fostering a culture of transparency, even in situations not subject to regulatory or other

oversight. Such proactive measures act as a safeguard against the deployment of unverified algorithms that might endanger patient safety. Integrating algorithm registration into the current practice could ensure the safe, transparent, and responsible integration of AI in healthcare. While early registration will foster transparency, accountability, and eventually ensure patient safety, it is imperative to strike a balance between capturing knowledge at an early stage and minimizing registration burden. We therefore advocate for an iterative and flexible registration process that can adapt to the evolving landscape of AI in healthcare. AI registration represents a crucial important advancement to improve safety and responsible use of AI in healthcare. It responds to the growing ask and need for regulatory frameworks, regulatory oversight and robust solutions^{27,30}. We encourage governmental agencies, national and international organizations, AI experts, and the private sector (including tech companies) to bundle forces and knowledge to facilitate and regulate such a registry.

Michel E. van Genderen¹✉, **Davy van de Sande**¹, **Lotty Hooff**², **Andreas Alois Reis**³, **Alexander D. Cornet**^{4,5}, **Jacobien H. F. Oosterhoff**⁶, **Björn J. P. van der Ster**¹, **Joost Huiskens**^{7,10}, **Reggie Townsend**^{8,9,10}, **Jasper van Bommel**¹, **Diederik Gommers**¹ & **Jeroen van den Hoven**⁶

¹Erasmus MC University Medical Center, Department of Adult Intensive Care, Rotterdam, The Netherlands. ²Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands. ³Department of Research for Health, Division of the Chief Scientist, World Health Organization, Geneva, Switzerland.

⁴Section editor Intensive Care at Nederlands Tijdschrift voor Geneeskunde (Dutch Journal of Medicine), Amsterdam, The Netherlands. ⁵Department of Intensive Care, Medisch Spectrum Twente, Enschede, The Netherlands.

⁶Delft University of Technology, Faculty of Technology, Policy and Management, Delft, The Netherlands. ⁷Microsoft, Healthcare, Amsterdam, The Netherlands. ⁸Vice President Data Ethics Practice, SAS Worldwide Headquarters, Cary, N.C., USA. ⁹National Artificial Intelligence Advisory Committee, Executive Office of the President of the United States, Washington, D.C., USA. ¹⁰Views expressed are authors own and do not represent employer or other related affiliations

✉ e-mail: m.vangenderen@erasmusmc.nl

Received: 16 June 2023; Accepted: 15 April 2024;

Published online: 08 May 2024

References

- van de Sande, D., van Genderen, M. E., Huiskens, J., Gommers, D. & van Bommel, J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. *Intensive Care Med.* **47**, 750–760 (2021).
- Medicine, U.N.L.o. National Library of Medicine(U.S.). Clinicaltrials.gov. (2024).
- FDA. *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices* (FDA, 2024).
- Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
- De Angelis, C. et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N. Engl. J. Med.* **351**, 1250–1251 (2004).
- Bajgain, B., Lorenzetti, D., Lee, J. & Sauro, K. Determinants of implementing artificial intelligence-based clinical decision support tools in healthcare: a scoping review protocol. *BMJ Open* **13**, e068373 (2023).
- Benjamens, S., Dhunnoo, P. & Mesko, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit. Med.* **3**, 118 (2020).
- Badal, K., Lee, C. M. & Esserman, L. J. Guiding principles for the responsible development of artificial intelligence tools for healthcare. *Commun. Med.* **3**, 47 (2023).
- World Health Organization. *Ethics and governance of artificial intelligence for health* (WHO, 2021).
- Cruz Rivera, S. et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digital Health* **2**, e549–e560 (2020).
- Liu, X. et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat. Med.* **26**, 1364–1374 (2020).

12. Vasey, B. et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat. Med.* **28**, 924–933 (2022).
13. Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **1**, 501–507 (2019).
14. European Commission. *Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts* (European Commission, 2021).
15. Wilson, C. B. An updated Declaration of Helsinki will provide more protection. *Nat. Med.* **19**, 664 (2013).
16. Council of Europe. *Artificial Intelligence: Helsinki conference conclusions*. 2023 (Council of Europe, 2019).
17. Perni, S., Lehmann, L. S. & Bitterman, D. S. Patients should be informed when AI systems are used in clinical trials. *Nat. Med.* **29**, 1890–1891 (2023).
18. London, A. J. Artificial intelligence in medicine: Overcoming or recapitulating structural challenges to improving patient care? *Cell Rep. Med.* **3**, 100622 (2022).
19. Hightower, M., Kohane, I.S. & Gotbaum, R. Is Medicine Ready for AI? *N. Engl. J. Med.* **388**, e49 (2023).
20. Fehr, J., Citro, B., Malpani, R., Lippert, C. & Madai, V. I. A trustworthy AI reality-check: the lack of transparency of artificial intelligence products in healthcare. *Front. Digital Health* **6**, 1267290 (2024).
21. Wong, A. et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern. Med.* **181**, 1065–1070 (2021).
22. Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y. & Ghassemi, M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in underserved patient populations. *Nat. Med.* **27**, 2176–2182 (2021).
23. Omiye, J. A., Lester, J. C., Spichak, S., Rotemberg, V. & Daneshjou, R. Large language models propagate race-based medicine. *npj Digital Med.* **6**, 195 (2023).
24. Lambert, S. I. et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *npj Digital Med.* **6**, 111 (2023).
25. Ghersi, D. & Pang, T. En route to international clinical trial transparency. *Lancet* **372**, 1531–1532 (2008).
26. WHO. *Ethics and governance of artificial intelligence for health - Guidance on large multi-modal models* (WHO, 2024).
27. Meskó, B. & Topol, E. J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digital Med.* **6**, 120 (2023).
28. Komorowski, M. Clinical management of sepsis can be improved by artificial intelligence: yes. *Intensive Care Med.* **46**, 375–377 (2020).
29. Li, H. et al. Ethics of large language models in medicine and medical research. *Lancet Digit Health* **5**, e333–e335 (2023).
30. Raza, M. M., Venkatesh, K. P. & Kvedar, J. C. Generative AI and large language models in health care: pathways to implementation. *npj Digital Med.* **7**, 62 (2024).

Author contributions

MvG and DvdS conceptualized and wrote the manuscript. The manuscript was edited and critically reviewed by JvdH, LH, AR, AC, JH, RT, JvB, BvdS, and JO. DG directed overall research and edited the paper. All authors read and approved the final manuscript and had final responsibility for the decision to submit for publication.

Competing interests

D.G. has received speakers fees and travel expenses from Dräger, GE Healthcare (medical advisory board 2009–12), Maquet, and Novalung (medical advisory board 2015–18). All other authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Michel E. van Genderen.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024