PERSPECTIVES

# Managing Risk and Quality of AI in Healthcare: Are Hospitals Ready for Implementation?

Arian Ranjbar [ORCID][1], Eilin Wermundsen Mork[1], Jesper Ravn[1], Helga Brøgger[2], Per Myrseth[2], Hans Peter Østrem[3], Harry Hallock[2]

[1]Medical Technology and E-Health, Akershus University Hospital, Lørenskog, Norway; [2]Group Research and Development, DNV AS, Høvik, Norway; [3]Business Assurance, DNV AS, Høvik, Norway

Correspondence: Arian Ranjbar, Akershus University Hospital, Sykehusveien 25, 1478, Lørenskog, Norway, Tel +46700436768, Email arian.ranjbar@medisin.uio.no

**Abstract:** Artificial intelligence (AI) provides a unique opportunity to help meet the demands of the future healthcare system. However, hospitals may not be well equipped to handle safe and effective development and/or procurement of AI systems. Furthermore, upcoming regulations such as the EU AI Act may enforce the need to establish new management systems, quality assurance and control mechanisms, novel to healthcare organizations. This paper discusses challenges in AI implementation, particularly potential gaps in current management systems (MS), by reviewing the harmonized standard for AI MS, ISO 42001, as part of a gap analysis of a tertiary acute hospital with ongoing AI activities. Examination of the industry agnostic ISO 42001 reveals a technical debt within healthcare, aligning with previous research on digitalization and AI implementation. To successfully implement AI with quality assurance in mind, emphasis should be put on the foundation and structure of the healthcare organizations, including both workforce and data infrastructure.

**Keywords:** artificial intelligence, management systems, quality assurance, risk management, implementation

## Introduction

The rapid advancements in Artificial Intelligence (AI) have significantly influenced several domains. As the technology continues to mature, attention has been drawn to implementing AI into healthcare, with the potential of improving effectiveness, personalizing treatment and diagnostics, improving patient safety, meeting the rising healthcare demands caused by an ageing population, and improving homecare to relieve the burden on often understaffed hospitals.[1] However, AI differs significantly from conventional software, thus, to mitigate the new associated risks, AI requires new methods for development, procurement, implementation, and management.

Given the risk that AI poses in healthcare and other sectors, entities such as the WHO have proposed regulatory considerations and frameworks for control and regulation.[2] In the US this is being achieved partly by the "Blueprint for AI Bill", but as it is non-binding, it requires implementation by all the federal agencies.[3] Conversely, in the EU this is being addressed by the centralized EU AI Act.[4,5] Globally, standards on AI are being developed by multiple organizations such as ISO/IEC JTC 1/SC 42, one being ISO/IEC 42001 Information technology – Artificial intelligence – Management system, and this has also been proposed as one of numerous future harmonized standards for the EU AI Act.[6] Specifically, ISO 42001 provides guidance for establishing, implementing, maintaining and continually improving an AI management system within the context of an organization.

The introduction of ISO 42001 provides a standardized framework for management of AI and the corresponding risks. This paper adds to the discussion of potential challenges of AI implementation and operation management within healthcare, by reviewing the new standard as part of a gap analysis conducted at Akershus University Hospital.

## Risk Factors and Quality Assurance

AI as a field has a long history, and with its development the scope and definition of terminology, including subfields such as machine learning, have been dynamic.[7] For the purpose of this paper we consider AI-systems in alignment with the ISO-definition: "the engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives".[8]

Following the rapid development during the last decade, implementation of AI has become an increasingly investigated field, with an extensive literature on the corresponding multifaceted risks.[9,10] Among risk factors are technical, e.g., accuracy, reliability (data quality) and data security;[11–13] ethical, e.g., privacy, equity, accessibility and informed consent, but also fairness and bias;[14–16] and organizational, e.g., workforce displacement, acceptance, liability, and trust. Currently, a large proportion of AI systems are targeting decision support, running the risk of overreliance and automation-bias, as well as ignorance due to information or alarm fatigue.[17] In addition to erroneous algorithmic results, these scenarios could potentially lead to unsound medical practice or liability concerns.[18,19]

In response to these risk factors, quality assurance frameworks have been suggested for AI.[20] However, so far, most frameworks have been designed for single projects or model development in isolation.[21,22] This is especially prevalent within research where AI has a longer history, and the primary focus on research quality metrics, e.g., reproducibility;[23] but also domain-specific validation requirements.[24] Similar frameworks have also been suggested for implementation projects, including validation and postimplementation evaluation.[22,25] In this context, recommendations on management around AI implementations have also been suggested, with emphasis on building cross-disciplinary teams, clear problem formulations and set goals, data quality assurance, focus on implementation from the start including engaging end-users etc.[26]

Recently more thorough guidelines have been suggested for a complete life cycle, i.e., from concept to implementation to decommissioning. In a scoping review, suggestions based on 72 relevant guidance documents were made for healthcare.[21] The process is divided into six stages, where recommendations are provided for preparation, development of AI, validation, development of software or application, impact assessment and implementation.

ISO 42001 is a standard for AI management, that adds an additional layer of considerations, by placing the individual projects in the context of an organization.[6] This requires further understanding of the relation, needs and expectations of the AI system in relation to the organisation; appointing competent leadership, with a clear policy, resources, and commitment; and enabling a workforce with competence and awareness, with sufficient communication and documentation. Moreover, the lifecycle may vary widely among different solutions depending on implementation via in-house development or procurement of commercially available products, imposing further demands on operation, validation and improvement. Given management and governance of healthcare data is almost exclusively done by the hospitals themselves, with limited access to data by third parties, the rate of involvement, either through in-house development or collaboration, may significantly increase in the foreseeable future.

Although AI implementation within healthcare is still in its early stages, ISO 42001 gives a standardized framework with a broad scope. Reviewing ISO 42001 may thereby provide a basis for a discussion ground on challenges of AI implementation, and requirements for both safe adoption and compliance with upcoming regulation.[20]

## Review and Gap Analysis

The aim of this paper was to add to the discussion of potential challenges of AI implementation in healthcare, particularly by reviewing the ISO 42001 standard from a healthcare perspective. To further expand on the perspective, a gap analysis of ISO 42001 was conducted at Akershus University Hospital, a large tertiary acute care hospital in Norway. The purpose of the gap analysis was to investigate the industry agnostic standard applied to a healthcare organization. The investigation is thereby limited, not adhering to requirements for statistical conclusions on the elements of the standard themselves, while still presenting discussion grounds as a case on AI implementation in healthcare.

The gap analysis was carried out in a process of understanding the requirements of ISO 42001, reviewing current practices, and identifying discrepancies. It was carried out by an inter-disciplinary working group with experts in AI, IT, health, information security, legal, business management, research and development, and management systems.

## Identified Challenges

ISO 42001 is structured similarly to other management standards such as ISO 9001 and ISO 14001, as well as information technology standards such as ISO 27001 and ISO 27701, with a set of 10 clauses common to all ISO management systems. In Norway many hospitals have already implemented ISO 9001 and ISO 14001, whilst some are additionally working towards ISO 27001; numerous hospitals throughout the US and the UK are ISO 9001 certified. The implementation of the requirements listed in ISO 42001 may therefore often be built upon existing practices. Naturally, overlapping areas are easier to comply with, since most hospitals will already have good practices in place, for example risk management processes e.g., risk assessment and treatment, procedures for operational control, incident reporting, and performance evaluation e.g., internal control and management review; all which can be adapted with adjustments for AI-specific requirements.

On the other hand, the examination identified that hospitals may lack the organizational maturity and technical resources regarding: (1) AI system development life cycle, for example specifications, design, verification and validation, deployment, operation and monitoring and technical documentation for AI systems; (2) data needs for AI systems, for example data governance policies and sound methods for data acquisition, data quality and data preparation. If these two topics are not sufficiently addressed, they pose a significant risk as they can have a negative impact on the safety and performance of the AI system.

## AI System Development Life Cycle

The major difference in AI compared with conventional software and medical equipment is the life cycle. Traditional software operates based on predefined instructions and rules, following a set of programmed commands to perform a specific task. Due to the static nature, it is easily verifiable and unwanted behaviour can be removed through code changes. AI solutions, on the other hand, are adaptable without explicit programming through the use of large datasets. However, since the data are part of forming the model, in a non-linear instruction set, the end result may be more stochastic; requiring comprehensive validation. Improving the models often involves further training, and cannot be done through simple code changes. These considerable differences reflect necessary changes to traditional hospital management. The following have been identified as key areas to consider:

Validation (and development) - Due to the data dependency, local validation is required independent of internal development or procurement, to an extent not experienced with conventional software. Thereby, the organization needs a support function more akin to development teams with expertise on data selection and assessment, testing requirements and verification metrics. However, this assumes models are robust against domain shifts; in many scenarios domain adaptation or further improvements will be necessary, requiring deeper involvement of the implementing organization. For internal development, the support function evidently needs further knowledge on machine learning modelling and software engineering. Since hospitals are the biggest health data administrators, they must acknowledge they are "big data" custodians and thus build the necessary management, competence, and infrastructure to use these data to improve healthcare.

Operation requires continuous monitoring of performance, unintended use, information security threats, adverse events etc. Although most hospitals already have monitoring systems in place for both clinical and technical incidents, it is typically done by incidence reporting in silos, i.e., through separate systems. AI systems carry potential combined technical and clinical adversities, requiring aggregated reporting.

Improvement includes greater software control and facilitating updates. However, this can be challenging since hospitals lack the necessary software and data pipeline for continuous integration. If systems are procured from external vendors, they may need internal data in order to improve and update the systems. Furthermore, if these challenges are resolved, the EU Medical Device Regulation limits continuous improvement for medical device software, as updates may be deemed a "significant change", thus potentially requiring re-certification.

If AI is to be implemented, which as discussed may have large benefits to healthcare organizations and society, a greater focus needs to be put onto enablers of the technology. This includes upskilling and recruiting a more multidisciplinary workforce that in addition to healthcare professionals, possesses competence across data science, software engineering, software implementation and operation, data management, information security, law and policy, and quality assurance to perform these tasks. Additionally, parts of the organizations may require transition towards management that reflects a more agile and software development-like approach.

## Data and Infrastructure

Organizations are potentially falling for the hype of AI before getting their house, or more so data, in order. An ongoing challenge is insufficient data governance and management, and more generally a misconception on data quality. Historically, hospitals have acquired data to provide healthcare services, allocate resources, manage finances, collect governance-related data, and track the quality of provided treatments. The format and structure of such data may be sufficient for the intended purpose, but can provide challenges when queried for general AI solutions. A typical example is the use of ICD-codes to annotate disease, traditionally used for administrative and financial purposes.[27] In theory, diagnosis codes may seem like a potential candidate as ground truth data for AI training, validation or monitoring. However, the validity of ICD-codes is known to be limited,[28] not adverse for administrative purposes, but detrimental as labelling for AI. Moving forward, not only is a thorough understanding of the data being used in development necessary, but potentially new medical data standards for the purpose of AI.

The challenge with building data pipelines in general is well known, particularly with regards to quality.[12] For sufficient development, including validation, both availability and reliability is relevant. The former is currently a challenge within many healthcare organizations due to reliance on a mix of legacy and contemporary systems, often with ad hoc solutions; and lack of standards among data formats, including vendor locking. Reliability, as seen in the examples discussed above, is a challenge since data have not been collected with AI development in mind, but rather to be quickly accessible and read by humans. Although AI is reaching human capabilities on unstructured data, performance is still better on structured data while providing a simpler environment for validation. To sufficiently address the data challenges, enabling technology is necessary, often available in other industries but currently lacking within many healthcare organizations. This includes adequate data governance and management, beyond administrative use cases, with accessible standardized data formats and databases.

## Conclusion

The findings from reviewing ISO 42001 and conducting a gap analysis align with previous research discussing the challenges of AI implementation, especially in regards to the AI life cycle.[9,20,21] Additionally, with the industry agnostic nature of ISO 42001, the review reveals a potential large technical debt within healthcare, agreeing with previous observations in the digitalization literature.[29] AI systems require deeper involvement by the healthcare organization, even when procured, compared with previous digital tools and software solutions. This is particularly evident in the requirements of the standard since the underlying data are administrated by the organizations themselves. For other industries, where the digitalization process has reached further,[30] and where software is a larger part of the everyday business, adapting AI with the right quality assurance is a smaller step.[31] Within many regards, hospitals are already data-driven entities, but lack modern software engineering and data management practices, since the degree of involvement in development and operation of software have previously been limited.[10] At the current state, to appropriately implement and reach the full potential of AI, hospitals need to put more emphasis on the foundation, including both the workforce and infrastructure. The digital transformation of the organization, reaching sufficient management in relation to AI, is not only necessary to reduce risk and ensure the quality of care, but will also be a requirement to comply with upcoming regulation.

## Funding

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019;6(2):94–98. doi:10.7861/futurehosp.6-2-94

2. World Health Organization. Regulatory considerations on artificial intelligence for health. World Health Organization; 2023. Available from: https://iris.who.int/handle/10665/373421. Accessed February 29, 2024.

3. Hine E, Floridi L. The blueprint for an ai bill of rights: in search of enaction, at risk of inaction. *Minds Mach*. 2023;33(2):285–292. doi:10.1007/s11023-023-09625-1

4. Veale M, Zuiderveen Borgesius F. Demystifying the Draft EU Artificial Intelligence Act — analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Rev Int*. 2021;22(4):97–112. doi:10.9785/cri-2021-220402

5. Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts; 2021. Available from: Accessed February 27, 2024. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.

6. ISO/IEC FDIS 42001. ISO. Available from: https://www.iso.org/standard/81230.html. Accessed September 16, 2023.

7. Wang P. On defining artificial intelligence. *J Artifi Gen Intell*. 2019;10(2):1–37. doi:10.2478/jagi-2019-0002

8. ISO/IEC 22989:2022. ISO. Available from: https://www.iso.org/standard/74296.html. Accessed February 29, 2024.

9. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25 (1):30–36. doi:10.1038/s41591-018-0307-0

10. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *J Big Data*. 2019;6(1):54. doi:10.1186/s40537-019-0217-0

11. Cabitza F, Zeitoun JD. The proof of the pudding: in praise of a culture of real-world validation for medical artificial intelligence. *Ann Translat Med*. 2019;7(8):161. doi:10.21037/atm.2019.04.07

12. Cai L, Zhu Y. The challenges of data quality and data quality assessment in the big data era. *Data Sci J*. 2015;14:2. doi:10.5334/dsj-2015-002

13. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethic*. 2021;22(1):122. doi:10.1186/s12910-021-00687-3

14. Schönberger D. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. *Int J Law Inform Technol*. 2019;27 (2):171–203. doi:10.1093/ijlit/eaz004

15. Murphy K, Di Ruggiero E, Upshur R, et al. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med Ethics*. 2021;22(1):14. doi:10.1186/s12910-021-00577-8

16. Naik N, Hameed BM, Shetty DK, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? *Front Surg*. 2022;9:266. doi:10.3389/fsurg.2022.862322

17. Fernandes CO, Miles S, Lucena CJPD, Cowan D. Artificial intelligence technologies for coping with alarm fatigue in hospital environments because of sensory overload: algorithm development and validation. *J Med Int Res*. 2019;21(11):e15406. doi:10.2196/15406

18. Mathis MR, Kheterpal S, Najarian K. Artificial intelligence for anesthesia: what the practicing clinician needs to know: more than black magic for the art of the dark. *Anesthesiology*. 2018;129(4):619–622. doi:10.1097/ALN.0000000000002384

19. Tang A, Tam R, Cadrin-Chênevert A, et al. Canadian association of radiologists white paper on artificial intelligence in radiology. *Can Assoc Radiol J*. 2018;69(2):120–135. doi:10.1016/j.carj.2018.02.002

20. Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P. Implementation frameworks for artificial intelligence translation into health care practice: scoping review. *J Me Int Res*. 2022;24(1):e32215. doi:10.2196/32215

21. de Hond AAH, Leeuwenberg AM, Hooft L, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *Npj Digit Med*. 2022;5(1):1–13. doi:10.1038/s41746-021-00549-7

22. Reddy S, Rogers W, Makinen VP, et al. Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ Health Care Inform*. 2021;28(1):e100444. doi:10.1136/bmjhci-2021-100444

23. Pineau J, Vincent-Lamarre P, Sinha K, et al. Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). *J Mach Learn Res*. 2021;22(1):7459–7478.

24. Sengupta PP, Shrestha S, Berthon B, et al. Proposed requirements for cardiovascular imaging-related machine learning evaluation (PRIME): a checklist: reviewed by the American College of Cardiology Healthcare Innovation Council. *Cardiovasc Imaging*. 2020;13(9):2017–2035. doi:10.1016/j.jcmg.2020.07.015

25. Antoniou T, Mamdani M. Evaluation of machine learning solutions in medicine. *CMAJ*. 2021;193(36):E1425–E1429. doi:10.1503/cmaj.210036

26. Verma AA, Murray J, Greiner R, et al. Implementing machine learning in medicine. *CMAJ*. 2021;193(34):E1351–E1357. doi:10.1503/cmaj.202434

27. World Health Organization. International statistical classification of diseases and related health problems (11th ed); 2019. Available from: https://icd.who.int/. Accessed March 25, 2024.

28. Quan H, Li B, Duncan Saunders L, et al. Assessing validity of ICD-9-CM and ICD-10 Administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res*. 2008;43(4):1424–1441. doi:10.1111/j.1475-6773.2007.00822.x

29. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: analyzing the current state-of-research. *J Busi Res*. 2021;123:557–567. doi:10.1016/j.jbusres.2020.10.030

30. Kutnjak A, Pihiri I, Furjan MT Digital transformation case studies across industries–literature review. In: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE; 2019:1293–1298. Available from: https://ieeexplore.ieee.org/abstract/document/8756911/?casa_token=OJJqylKJh-oAAAAA:1PtzwUCK3ED1D52_QpeQ9Bz_toi8fs4xm_zpbmpVOitrSe-S8SbnNbqVyXaI7-16Bqosj8yR. Accessed February 29, 2024.

31. Alsheibani S, Cheung Y, Messom C. Artificial intelligence adoption: AI-readiness at firm-level. *PACIS*. 2018;4:231–245.