Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

© 2024 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI240472

Proposing an AI Passport as a Mitigating Action of Risk Associated to Artificial Intelligence in Healthcare

Juan M. GARCÍA-GÓMEZ [a,1] , Vicent BLANES-SELVA[a]
and Ascensión DOÑATE-MARTÍNEZ[a]
[a]*BDSLab, Instituto Universitario de Tecnologías de la Información y Comunicaciones, Universitat Politècnica de Valencia, 46022 Valencia, Spain*
ORCiD ID: Juan M. García-Gómez https://orcid.org/0000-0002-3851-1557

Abstract. The integration of Artificial Intelligence (AI) in healthcare signifies a substantial shift, offering benefits to patients and healthcare systems while also introducing new risks. The emphasis on patient safety and performance standards is pivotal, especially with the European Union's strides towards regulating AI through the AI Act. This act focuses on classifying AI systems based on risk levels, mandating stringent requirements for high-risk AI, enhancing transparency, and ensuring ethics in AI applications. The concept of an "AI passport" is introduced as a living document detailing the AI system's purpose, ethical declarations, training, evaluation, and potential biases. This passport aims to enhance transparency and safety in medical AI applications, serving as a comprehensive record for patients, clinicians, and stakeholders. The AI passport, structured in JSON format, encapsulates key information about the AI system as a mechanism for continuous performance evaluation and transparency. This initiative may represent a significant step towards mitigating the risks associated with AI in healthcare, emphasizing the importance of accountability, transparency, and patient safety in the development and application of AI technologies.

Keywords. EU AI act, Patient safety, AI software as Medical Device, AI regulation, AI passport

1. Introduction

With the massive introduction of AI-based tools in professional services, we can expect a large transformation of the conception and industry of healthcare. Added to the great benefits for patients and healthcare professionals, AI tools may also bring risks not seen yet. Developers, manufacturers and healthcare providers would consider their responsibility with the patients as a main line of design. Given their complexity, this would require specific functionalities to ensure the continuous operation of the AI solutions at the highest standards of performance and safety for the patient [1-4].

The recent updates on the EU's Artificial Intelligence (AI) Act [5] mark significant progress towards establishing comprehensive regulations for AI technologies within the European Union. Key aspects of the AI Act include the classification of AI systems based

[1] Corresponding author: Juan M. García-Gómez; E-mail: juanmig@upv.es.

on their risk to fundamental rights, with "high-risk" AI systems subject to stringent requirements such as risk-mitigation systems, high-quality datasets, improved documentation, and human oversight.

The Act introduces binding rules on transparency and ethics, requiring tech companies to notify people when interacting with AI in specific contexts, label AI-generated content, and conduct impact assessments for essential services like healthcare. Despite these measures, there remains flexibility for AI companies, particularly around foundational models, which are powerful AI models used for various purposes. The Act mandates better documentation and compliance for these models.

Aligned with this principle of software design, Garcia-Gómez et al. proposed in [6] fourteen functional requirements to mitigate the Risk of Harm to Patients from Artificial Intelligence in Healthcare. For that, they reviewed the risk analysis performed by the Directorate General for Parliamentary Research Services (EPRS) of the European Parliament [5] where seven main risks of AI in medicine and healthcare where pointed out: 1) patient harm due to AI errors, 2) the misuse of medical AI tools, 3) bias in AI and the perpetuation of existing inequities, 4) lack of transparency, 5) privacy and security issues, 6) gaps in accountability, and 7) obstacles in implementation.

The first and main functional requirements detected in [5] was the issue of an AI passport associated to each AI entity operating in healthcare. The AI passport, implemented as a digital living document with declarations and relevant information, may serve patients, clinicians, and other stakeholders for the safety use of an AI software for medical purposes.

In this study, we delve deeper into the concept of AI passport by analyzing the sources of uncertainty of the risks to the patients associated to the use of AI and the proposed mitigation actions proposed by the EPRS. As a result, we propose a conceptual structure of the AI passport and its implementation in the JSON format. We also provide an example of an AI passport for a Clinical Decision Support System (CDSS) on palliative care interventions.

2. Designing the AI passport based on the mitigation actions to reduce risk for patients

The design of the AI passport is directly focused on reducing the risk of patient harm by providing maximum transparency of the AI tool to the final user. Hence, the AI passport brings direct access to the clinicians and rest of stakeholders of the healthcare system to key elements of the AI-based software used for the medical purpose, the implementation of the internal AI models, the quality [8] and equality of the used datasets, and how they were trained and evaluated.

The design of the AI passport was carried out by mapping the mitigation actions to the risk for patients with the elements involved in the design and deployment of an AI-based software as a Medical Device. As a result, Table 1 shows the mapping of each mitigation action for the main entities we will include in the AI passport.

From the Table 1 we can interpret most mitigation actions of the risks of harm to patients require tracking the key decisions made during the design and development of the AI-based solution (training and evaluation). Moreover, an incorrect use of the AI tools may result in potential harm to patients. Hence, data used during those processes may reveal potential biases on the AI behavior and subsequent interpretation.

Table 1. Mapping between mitigation actions for risks of harm to patients and AI passport elements including key attributes in parentheses. Acronyms: AI SaMD (AI-based software as Medical Device).

Risks of harm to patients	Mitigation actions	AI passport elements (attributes)
Patient harm due to AI errors	Comprehensive multi-centre evaluation studies to identify instabilities	Data (clinician), Evaluation Strategy (clinical validation, AI performance, continuous evaluation)
	Assistive AI solutions that maintain the clinician as part of the workflow	Data (clinician)
	Unexpected variations in clinical contexts and environments	Data (demographics, clinician, Data Quality assessment), Evaluation Strategy (continuous evaluation)
Misuse of medical AI tools	User-centred design and extensive usability test for the AI algorithm	Evaluation Strategy (perceived utility, perceived usability)
	Increase medical AI knowledge	AI entity, Training Algorithm, Evaluation Strategy
	Better regulation and information on emerging AI technologies	AI entity (AI foundational model, AI model, AI SaMD)
Bias in AI and the perpetuation of existing inequities	According to sex and gender; age differences; ethnic groups; geographic locations; socioeconomic factors	Data (demographics), Training Algorithm (mechanism to ensure equality)
Lack of transparency	Create an 'AI passport' for documenting the model	AI entity & its relationships
	Include traceability and explainability as prerequisites for certification	XAI mechanism
Privacy and security issues	Increased awareness of data privacy, consent, and cybersecurity	AI entity (encryption mechanism, field-tested libraries)
	Regulations to address accountability and protect citizens	AI entity (ID, manufacturer, date of release, role & responsibilities, regulation check, standards, certifications), Data (clinician), Evaluation Strategy (date roles & responsibilities, clinical validation, AI performance)
Gaps in accountability	Process should be implemented to identify the roles of AI developers and clinical users when AI-assisted medical decisions harm individuals	AI entity (ID, manufacturer, date of release, role & responsibilities, regulation check, standards, certifications), Data (clinician), Evaluation Strategy

3. Definition of the AI passport

AI passport was defined in [5] as the "complete statement detailing the AI system purpose, ethical declarations, context of use, training, and evaluation details, including potential biases due to the training datasets." From that definition and the mapping presented in section 2, we have defined the entity-relationship model of the AI passport by identifying eight entities and their six relationships as shown in Figure 1.

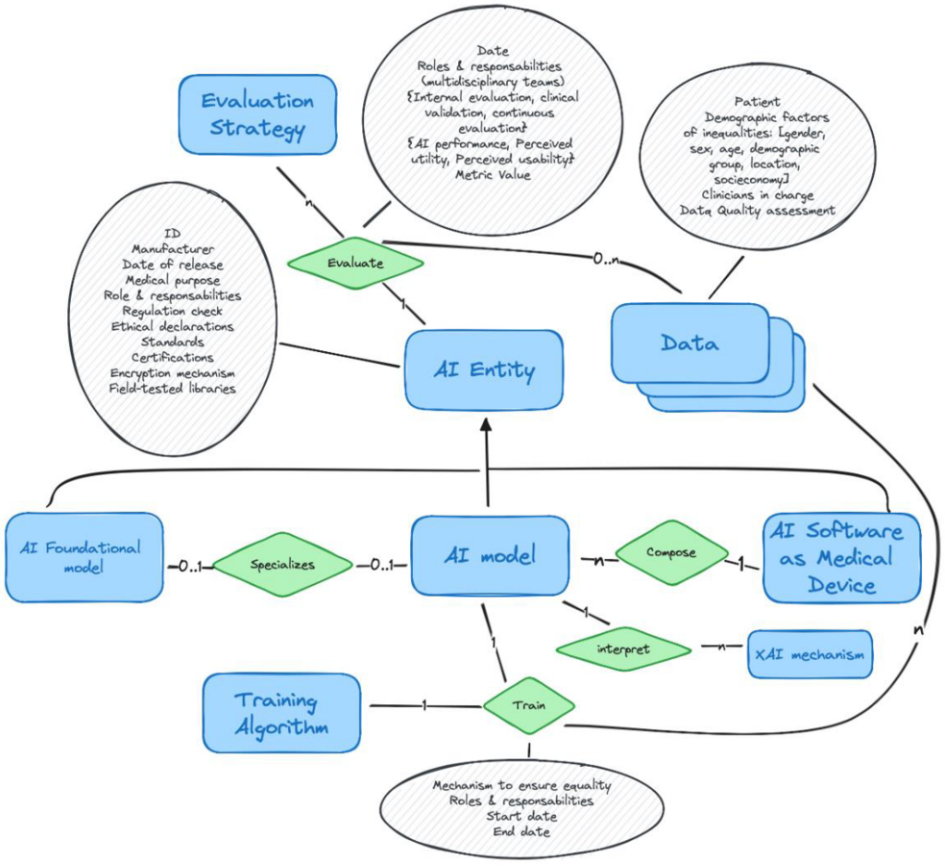

Figure 1. Entity-relationship model of AI passport associated to an AI software for medical purpose, where only main attributes are shown (access the GitHub of the AI passport for details: https://github.com/bdslab-upv/ai_passport.git).

It is worth noting that although the AI software is the medical product directly interacting with the user, it can be composed of several AI models that are trained and evaluated with complementary datasets. Moreover, current (and future) AI models derive from Foundational models [9], which may belong to different manufacturers and follow alternative standards.

In the schema, we have also specified the most relevant aspects for a safety use of the AI software and models, including basic information such as its clinical purpose and manufacturer along with its complete information about training, evaluation and use and mechanism to ensure equality, interpretability, and highest standards of design.

We conceive four methods associated to AI passport to deploy its functionality:

- Declaration: AI passport may serve as the primary document of the medical purpose, limitations, accountabilities, and liabilities of the software [10].
- Validation: automatic validation services may check the compliance of the AI passport with respect to the structure, model registers and data repositories.

- Update: although a first version of the AI passport would be created from the manufacturer statements, it should be updated with the results of the continuous performance evaluation and the usability tests.
- Disclaimer: maximum transparency of how the AI software is designed allows clinicians and stakeholders to use it with the highest standards of security, making them also responsible for their acts by interacting with the system.

A repository with the complete implementation in JSON of the AI Passport and its validator can be downloaded for academic purposes at https://github.com/bdslab-upv/ai_passport.git. Moreover, this repository includes a complete example applied to the Aleph CDSS for Palliative Care, including the AI passports of the CDSS and its three internal AI models.

4. Conclusions

In this work we have presented a complete definition and structure of the AI passport associated to an AI software for a medical purpose. This passport details its medical purpose, ethical declarations, context of use, training, and evaluation details, including potential biases due to the training datasets. This initiative represents a significant step towards mitigating the risks associated with AI in healthcare, emphasizing the importance of accountability, transparency, and patient safety in the development and application of AI technologies. The JSON implementation is free for academic purposes.

References

[1] Amann J, Vayena E, Ormond KE, Frey D, Madai VI, Blasimme A. Expectations and attitudes towards medical AI: A qualitative study in the field of stroke. PLoS One. 2023;18(1).
[2] Tang L, Li J, Fantus S. Medical artificial intelligence ethics: A systematic review of empirical studies. Digit Health. 2023;9. https://doi.org/10.1177/20552076231186.
[3] Oala L, Fehr J, Gilli L, Balachandran P, Leite AW, Calderon-Ramirez S, et al. Ml4h auditing: From paper to practice. In: Machine learning for health. PMLR; 2020. p. 280-317.
[4] Fehr J, Jaramillo-Gutierrez G, Oala L, Gröschel MI, Bierwirth M, Balachandran P, et al. Piloting A Survey-Based Assessment of Transparency and Trustworthiness with Three Medical AI Tools. Healthcare (Basel). 2022 Sep;10(10):1923.
[5] The European Commission. Proposal for a regulation of the European parliament and the council laying down harmonized rules artificial intelligence (artificial intelligence act) and amending certain union legislative. 2021.
[6] García-Gómez JM, Blanes-Selva V, Cenzano JCDB, Cebolla-Cornejo J, Doñate-Martínez A. Functional requirements to mitigate the Risk of Harm to Patients from Artificial Intelligence in Healthcare. arXiv preprint arXiv:2309.10424.
[7] European Parliamentary Research Service. Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts. 2022.
[8] Sáez C, Martínez-Miranda J, Robles M, García-Gómez JM. Organizing Data Quality Assessment of Shifting Biomedical Data, Quality of Life through Quality of Information. Stud Health Technol Inform. 2012;180:721-5.
[9] Bommasani R, Hudson D, Adeli E, Altman R, Arora S, von Arx S, et al. On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258. 2021.
[10] The European Commission. Proposal for a directive of the European parliament and the council on liability for defective products. 2022.