Towards a Taxonomy of AI Risks in the Health Domain

Delaram Golpayegani
ADAPT Centre
Trinity College Dublin
Dublin, Ireland
sgolpays@tcd.ie

Joshua Hovsha
School of Law
Trinity College Dublin
Dublin, Ireland
hovshaj@tcd.ie

Leon W. S. Rossmaier
Section of Philosophy (BMS)
University of Twente
Enschede, Netherlands
l.w.s.rossmaier@utwente.nl

Rana Saniei
Ontology Engineering Group
Universidad Politécnica de Madrid
Madrid, Spain
r.saniei@upm.es

Jana Mišić
Section of Philosophy (BMS)
University of Twente and Rathenau Institute
The Hague, Netherlands
j.misic@rathenau.nl

***Abstract*—The adoption of AI in the health sector has its share of benefits and harms to various stakeholder groups and entities. There are critical risks involved in using AI systems in the health domain; risks that can have severe, irreversible, and life-changing impacts on people's lives. With the development of innovative AI-based applications in the medical and healthcare sectors, new types of risks emerge. To benefit from novel AI applications in this domain, the risks need to be managed in order to protect the fundamental interests and rights of those affected. This will increase the level to which these systems become ethically acceptable, legally permissible, and socially sustainable. In this paper, we first discuss the necessity of AI risk management in the health domain from the ethical, legal, and societal perspectives. We then present a taxonomy of risks associated with the use of AI systems in the health domain called HART, accessible online at `https://w3id.org/hart`. HART mirrors the risks of a variety of different real-world incidents caused by use of AI in the health sector. Lastly, we discuss the implications of the taxonomy for different stakeholder groups and further research.**

***Index Terms*—risk, AI systems, health, AI regulation, ethics of AI, AI public policy, taxonomy**

I. Introduction

Application of AI in the health domain has great potential for promoting public health, improving patient care, reducing treatment costs, assisting medics in reaching a diagnosis, and discovering new treatment methods and drugs. However, there are significant risks involved in the use of AI systems such as risk of errors which can lead to injury to patients or risk of disclosing patients' sensitive data [1]. With the huge amount of AI investment in the medical and healthcare sectors for drugs, cancer, molecular, and drug discovery [2], the uncertainties around the newly developed AI systems in these sectors are increased.

This circumstance has led to a lively discussion within the fields of ethics, social sciences, and legal scholarship making it all the more necessary for conceptualising and identifying the exact risks to different stakeholder groups and to use this insight for the different contexts of risk management and impact assessment. Significantly the need for risk-based assessment lies at the heart of four relevant legislative instruments which apply to AI in the health sector within the European Union. However, in the current state of debate in the ethical, legal, and social science literature, it is often not clear how the addressed risks are conceptualised. By establishing a clear understanding of AI risk, we aim to contribute to a better understanding of ethical and legal risk assessment that allows comparison between different methodologies and facilitates communication in interdisciplinary groups conducting AI risk and impact assessments. To this end, we provide a formal taxonomy of risks in the health domain by bringing ethical, legal, and knowledge engineering perspectives together.

In this paper, we take the first step toward creating a shared and formal representation of AI risks in the health domain by proposing a taxonomy called HART (Health AI Risk Taxonomy). HART provides a catalogue of known AI risks, risk sources, risk consequences, and risk impacts identified from real-world incidents indexed in the AI, algorithmic, and automation incidents and controversies (AIAAIC) repository[1]-an open-source corpus of more than 850 incidents caused by AI all around the world.

The remainder of the paper is organised as follows: Section II provides discussions on risk of AI in the health domain from the ethical, societal, and legal perspectives to illustrate how AI risks are conceptualised. Section III describes the proposed taxonomy and Section IV discusses potential applications and impacts of the taxonomy. Finally, Section V concludes the paper by discussing future work.

This project is the result of interdisciplinary research within the PROTECT (Protecting Personal Data Amidst Big Data Innovation) project and has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813497.

[1] https://www.aiaaic.org/aiaaic-repository

II. Discussions on AI risks in the health domain

A. The Ethical Discussion on AI

We situate our risk-based approach to the assessment of AI systems in the medical domain in the interdisciplinary context of ethics, public policy, and legal scholarship to highlight our interpretation of risk as the potential negative impact on individuals, groups, or society at large.

The literature on the ethics of technology currently discusses the application of artificial intelligence systems in various health related fields, such as public health [3], healthcare, or mental healthcare [4], as well as clinical practice [5] and health research [6]. At the centre of the ethical discourse are worries and concerns linked to the development, design, and deployment of artificial intelligence systems, which we refer to as ethical risks. This focus has been, however, challenged because it neglects the circumstance that AI systems not only pose challenges to moral goods like values, rights, or duties but also mediate how we conceptualise such moral goods [7].

However, scholars have pointed out that the deployment of AI systems in the field of medicine holds severe implications for the autonomy of patients rooted in the paternalistic epistemic authority of AI systems [8]. While the potential restrictions of the patient's autonomy are heavily debated [5], [9], the discourse on autonomy holds implications for what is often promised by the implementation of AI systems; to increase the patient's empowerment. The concept of empowerment, as widely understood in terms of increased knowledge, independence or autonomy in medical decision-making, has been challenged due to the advent of AI in medicine and termed more in the mode of digital companionship supporting the patient throughout the care process [10] since AI systems may also offer ways in which the user's health and well-being might benefit from limitations to their autonomy.

The discussion on autonomy is closely connected to worries about the privacy of data subjects or users of AI systems since scholars have argued that respect for privacy is constituent of autonomy [11]. The discussion on privacy mainly focuses on the challenges to the adequate protection of user or patient information collected by AI systems or data that is used as training data [5], [12].

Moreover, trustworthiness of AI systems is a reoccurring theme in the literature [13]. The trustworthiness of AI systems is often challenged since it is dependent on performance. This is especially relevant if the system is deployed in the medical context, for instance as part of diagnosis procedures [14]. Furthermore, the trustworthiness of AI systems has implications for practitioners applying them in everyday practice. Relying too much on the information provided by the AI system cause problems for meaningfully assigning responsibility and accountability in case of over-diagnosis or mistreatment [15].

AI systems deployed in everyday practice pose several difficulties for medical professionals. For instance, it is expected that the use of AI systems results in decreases in empathy from side of the physician including concerns about how physicians perceive their responsibility to care for their patients [16]. Furthermore, scholars have pointed out the risk of deskilling of medical personnel [5], [17], as well as an alternating effects on how physicians perceive the nature of their work [18].

Lastly, AI systems applied in the health domain pose several challenges for social justice. For one, there is the risk of increasing already existing inequalities in health and public health since access and opportunities to gain the benefits resulting from this technology are not distributed equally across society and are largely dependent on the user's digital competency as well as their access to the internet [3]. This circumstance may result in biased training data for AI systems [5], [19], which is particularly problematic because populations that are most in need and likely to benefit most from the systems are often excluded from contributing to the data sets [20]. Such data biases might persist in the developing processes of drugs and treatment methods.

B. Public Policy and Society Context

As hinted earlier, risks of and around AI in public health can be normative or epistemic, or they can relate to individuals, groups, relationships, institutions, and society in general [14]. In the broader societal discussions on public health, AI risks are often consequential to epistemic shortcomings (such as intransparency or lack of digital competency) on the side of users and policymakers alike. One of the central, ever-present risks is the choice of policymakers in determining the problem that AI is aimed at solving. The deployment of AI-based tools in the clinical space, for example, can be conflicted when algorithmic decisions aim to minimise malpractice on the one hand, while simultaneously not allowing the physicians to be fully aware of the decision-making process of the automated system. The intransparency can then contribute to liability issues between the system and the institution or a physician, prompting policymakers to red-tape some models in specific settings [21].

Another set of epistemic risks arises from the uncertainties and the hype surrounding AI systems. The public sector, with public health pioneering the trend, has been increasingly deploying algorithmic systems as policymakers rush to get on the private sector AI-wagon [22]. Intransparency can also be the result of particular ownership models where public health institutions collaborate with private entities providing AI-based technologies. Within these public-private arrangements, the highly proprietary nature of AI systems can obstruct third-party inspection of the technology [23]. The reluctance to transparently inform details of an AI model is worrying as it raises the potential for harm when a model is transferred from a trial research context into a real-world clinical practice [24]. It can also create a risk for determining accountability for AI errors or with privacy breaches within data sharing partnerships between the public health institutions and the private entity [25].

In terms of normative considerations, the broader societal agreement on what is risky, what is inclusive and what is acceptable, poses a challenge for public health AI. For one, and connected to the issue of social justice mentioned earlier, aggregation of big data may not always be representative. For example, patients suffering from cognitive disability or poverty may have less access to smartphones and digital technology [26] and therefore be omitted from training data sets. Risk of segregation and polarisation of individuals and groups can also occur when AI-based platforms or applications are the first-line of inquiry for patients. Patients who partake in online health communities may be skipping the formal diagnosis altogether, relying on self–treatment online [27].

When policymakers discuss normative risks, the challenge of group versus individual characteristics also appears [28]. What is a priority for one individual, may not be for others within the same group as well. This further poses the question of democratic representation and stakeholder participation in determining the goals and acceptability of risk with AI-based systems. Di Nucci (2019) proposes a multi-stakeholder engagement process to understand which tasks are socially acceptable to be delegated to AI-health before making it official policy [14]. While benefits may arise for the efficiency of the public health provision, risks of rising unemployment due to automation of services are present as well. The 2021 World Health Organisation Guidance on Ethics and Governance of Artificial Intelligence for Health [29] echoed scholars' concerns about power and ownership in drawing attention to the 'uberization' of health care [30] as medical professionals' jobs become less stable and secure.

C. EU Legal and Regulatory Context

Two relevant Regulations impacting medical devices across the EU are relevant for the work of HART: the Medical Devices Regulation (MDR) [31] which came into effect on 26 May 2021, and the In-Vitro Diagnostic Devices Regulation (IVDR) [32] which has been forced since 26 May 2022.

Significantly, a common clause found in both of these Regulations (see MDR Recital 19, 2017; IVDR Recital 17, 2017) states that software is to be understood as a medical device/in-vitro diagnostic device. This is true regardless of whether the software is a stand-alone offering or is simply a component in a wider medical device. Significantly, risk assessment lies at the core of the requirements for both regulations.

Beyond these two legal instruments, the processing of personal data and use will fall under the scope of the General Data Protection Regulation (GDPR) [33] which is currently in force. Additionally, the Proposed AI Act [34] would be significant for the governance of AI within the Union in the near term.

1) Personal Data and the General Data Protection Regulation: The right to Privacy and the right to Data Protection are two distinct rights enshrined in EU law in Article 7 and Article 8 of the Charter of Fundamental Rights of the European Union [35]. These rights are given effect through legislation such as GDPR.

Article 35 of the GDPR mandates that a Data Protection Impact Assessment (DPIA) be undertaken where processing of data is likely to result in a "high risk to the rights and freedoms" of natural persons. From the onset, it is clear that the concern to be resolved by a DPIA is not limited to privacy rights or data protection. Instead, the requirement for risk assessment is triggered whenever data processing involves any form of high risk to any "right and freedoms".

Going further, Article 35 (3) of the GDPR outlines a non-exhaustive list of processing activities which require a Data Protection Impact Assessment. A data protection impact assessment referred to in paragraph 1 shall in particular be required in the case of:

(a) a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;
(b) processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; or
(c) a systematic monitoring of a publicly accessible area on a large scale. (Emphasis added)

As such, Working Party 29 [36] outlines nine criteria which signal high risks and thus require a full assessment. These are derived from Article 35 of GDPR along with recitals 71, 75 and 91, and a review of the phrasing 'likely to result in a high risk' throughout the Regulation. These are:

1) Evaluation and scoring
2) Automated decision making
3) Systemic monitoring
4) Processing of sensitive data or data of a highly personal nature
5) Data processed at a large scale
6) Matching and combining datasets
7) Data concerning vulnerable data subjects
8) Innovative use of technology or applying new technological or organisational solutions
9) When the processing in its own right "prevents data subjects from exercising a right or using a service or a contract" as laid out in Article 22 and Recital 91 of GDPR (Working Party 29, 2018, pp. 9–10).

Many of these criteria would apply within the context of AI use for healthcare. What is more the Data Protection Regulators within EU member states have built upon this guidance requiring the completion of Article 35 Impact Assessments when processing meets certain requirements.

Thus, for instance, the French Commission Nationale de l'Informatique et des Libertés [37] has laid its own categories for processing of data requiring a DPIA. Of these the following criteria would be applicable in all cases of medical use of AI:

1) Health data processing implemented by health establishments or medical-social establishments for the care of persons;
2) Processing dealing with genetic data of so-called "vulnerable" people (patients, employees, children, etc.);

3) Processing whose purpose is the management of alerts and reports in social and health matters;
4) Processing whose purpose is the management of alerts and reports in professional matters;
5) Processing of health data necessary for the establishment of a data warehouse or register;
6) Processing involving the profiling of persons that may lead to their exclusion from the benefit of a contract or the suspension or even the breaking of the contract;
7) Processing for the purpose of providing social or medico-social support to persons.

From the weight of this overlapping guidance, we conclude that the GDPR will require a full impact assessment as a default for all AI systems in the healthcare sector.

2) AI Act Requirements: As with the GDPR, the Draft AI Act [34] would have territorial application beyond EU member states when companies utilise AI on individuals or groups within the EU. The European Commission describes the AI Act as a risk-based legal instrument. Four levels of Risk associated AI have been identified. (I) At the highest level are 'unacceptable risks' which are prohibited. (II) Below this level are AI uses involving high risks, these are permitted when certain compliance obligations and assessments have been carried out. (III) Thereafter, certain AI uses with mandatory transparency obligations are outlined. Importantly level two and three are not exclusive as such certain activities may fall within the remit of both levels' obligations. Finally, (IV) at level four are AI activities with minimal to no risk, which are permitted with no restrictions.

Annex III of the Draft AI Act delineates the concept of 'High Risk Usage' as applying to products or safety components of products already covered by EU health and safety harmonisation legislation (such as toys, machinery, lifts, or medical devices). As has been noted above, both the Medical Devices Regulation as well as In-Vitro Diagnostic Devices Regulation accept software programs as part of their definitions. As such the category of 'High-Risk' AI system will apply in the case of medical AI. It should be assumed that the requirements of 'High Risk' AI systems will be applicable in these cases.

Of relevance to the taxonomy presented in this work, Article 9 of the Draft AI Act mandates a 'continuous iterative process' of risk identification and management to be run through the 'entire lifecycle' of High-Risk AI systems.

As such, the need for risk-based assessment lies at the heart of all four relevant legislative instruments which apply to AI in the Health Sector at a Union-wide level: The MDR, IVDR, GDPR, and Draft AI Act.

III. HART: Health AI Risk Taxonomy

A. Methodology

To develop HART, we follow Noy and McGuinness' guidelines [38] alongside the steps specified in the Linked Open Terms (LOT) methodology [39]. The methodology is summarised as follows:

- Requirements specification: The requirements are specified in the form of competency questions- questions that the taxonomy is designed to provide answers to.
- Use-case identification and analysis: Incidents where use of AI in the health domain involved critical risks are selected from the AIAAIC repository and then analysed to extract information regarding the AI systems and their risks.
- Implementation: Adopting a bottom-up approach, the identified instances are classified into more general categories. In this process, we consider reusing AI risk concepts from existing risk ontologies such as AIRO (AI Risk Ontology) [40] and RiskOnto [41] and existing classifications including WHO classification of AI for health [29]. The machine-readable format of the taxonomy is generated in OWL (Web Ontology Language) using Protégé[2]- an open-source tool for ontology development. HART documentation is made available online at `https://w3id.org/hart` under the CC BY 4.0 license.
- Evaluation: HART is evaluated against the competency questions. To ensure the quality of the taxonomy, we follow W3C Best Practice Recipes for Publishing RDF Vocabularies[3] and FAIR best practices [42].
- Maintenance: HART will be periodically updated in light of new emerging AI risks.

B. HART Requirements

The competency questions used in the development and evaluation of HART are derived from the discussions presented in Section II, and are as follows:

1) What are the purposes of using AI?
2) What types of AI applications are available?
3) Which AI techniques are utilised in AI systems?
4) Which entities develop AI systems?
5) Who uses AI systems?
6) What are the types of risks associated with use of AI ?
7) What are the categories of risk sources?
8) What are the categories of consequences of AI risks?
9) What are the types of impacts of AI?
10) Who can be impacted by AI systems?
11) What rights can be negatively impacted by AI?

C. Use-case Identification & Analysis

By filtering use-cases in AIAAIC according to the sector, we identified 52 cases where incidents occurred in the 'Health' or 'Govt - Health' sectors. We annotated each of the use-cases to extract the following information about the AI system: the intended *Purpose* of the system, the *AI Technique(s)* used in the system, the system's *Application*, *Operator(s)*, *Developer(s)*, and *User(s)* of the system. Regarding the incident described in the use-cases the following information is extracted: *Risk* imposed by the AI system, its *Risk Source(s)*,

[2] https://protege.stanford.edu/

[3] https://www.w3.org/TR/swbp-vocab-pub/

Consequence(s), *Impact(s)*, *Subjects*, and *Area(s) of Impact*. Examples of annotated use-cases are shown in Table I.

TABLE I
EXAMPLES OF USE-CASE ANNOTATION

AIAAIC ID	AIAAIC0758	AIAAIC0633
AI System	NarxCare	DermAssist
Purpose	Assessing and predicting drug abuse	Diagnosis of skin, nail, and hair issues
AI Technique	Machine learning techniques	Deep learning & computer vision techniques
AI Application	N/A	Image analysis
AI Developer	Bamboo Health	Google
AI Provider	California Department of Justice (DOJ)	Google
AI User	Pharmacists, clinicians	Patients
Risk Source	Training data	Training data
Risk	Inaccuracy	a) Inaccuracy, b) Unlawful use of sensitive data
Consequence	Inaccurate outcome	a) Complex conditions being missed, b) Privacy violation
Impact	Discrimination (enforcement of racial and gender bias)	a) Physical injury, b) Fundamental rights infringement
Area Of Impact	Justice	a) Health b) Right to data protection
AI Subject	Patients, medical staff	Patients

D. HART Overview

Concerning AI systems used in the health domain, we have classified types of techniques, applications, as well as purposes, users, developers and providers of these systems. Regarding risks associated with AI systems applied in the health care settings, categories of risk sources, risks, consequences, and impacts are identified. In addition, a classification of the areas and the stakeholders that could be negatively impacted by AI systems is provided. Figure 1 illustrates an overview of HART. In the following, we provide definitions of the main concepts in HART.

AI Technique refers to techniques and approaches utilised in the development of an AI system [34]. In the analysed incidents, *Machine Learning* techniques were the most commonly used techniques. ***AI Application*** indicates the particular use of an AI system, such as *Facial Recognition*, *Image Analysis*, *Voice Recognition*, and *Text Analysis*. ***AI Developer*** refers to an entity who has developed the AI system and ***AI Provider*** is the entity that provides the AI system to users. ***AI Users*** in the health care settings are quite diverse; AI systems can be used by *Individuals* including patients, *Groups* e.g. employees, and *Health Care Providers* such as hospitals, pharmacies, clinics, and health care professionals for different ***Purposes***, e.g. *Diagnosis*, *Health Research*, and *Disease Prevention*.

Risk in HART represents harmful risk of AI systems to reflect the AI Act's interpretation of risk. ***Risk Source*** indicates an event that has the potential to give rise to risks [43]. We classified risk sources into three top-level categories according to the origin of the source. Our analysis shows that approximately half of the reviewed incidents were caused by *Data-Related Risk Sources* such as use of unrepresentative training data and improper anonymisation of patient data. ***Consequence*** indicates the outcome of an event affecting objectives [43]. ***Impact*** represents adverse outcomes of a consequence on individuals, groups, and society [40]. We classified the identified impacts into three categories *Well-being Impact*, *Fairness Impact*, and *Fundamental Rights Infringement*. ***AI Subjects*** are the stakeholders who are impacted by AI systems, including those who are not even direct users of the system. In our study, we discovered that in the majority of incidents, *Individuals*, in particular patients, were negatively affected by AI systems. ***Area of Impact*** refers to the areas that can be negatively affected by AI systems. Analysis of incidents suggests that impacts related to *Well-being* occurred more commonly.

IV. Anticipated Application and Impacts of HART

In this section, we discuss the impact of the proposed taxonomy on various stakeholders explaining how it can be used to increase the feasibility of interdisciplinary work on the risk assessment of AI systems utilised in the health context.

The benefit of adopting a taxonomy approach to AI for healthcare is that the categories of risk sources, risks, impacts, and 'things to check' can be defined for a class of use-cases, e.g. diagnosis, drug development, or research, but specific issues that might arise due to the particular contexts of use, e.g. choice of technology, can be addressed as a subclass of that category if a different action, risk treatment, or control requirement arises. This approach also allows for the identification of common issues, risks, and approaches to risk treatment.

A. The Implications for Different Stakeholder Groups

The ambition of our taxonomy is to provide guidance and support for multiple contexts concerning AI systems in the health domain. Interdisciplinary researchers working on risk of development and use of AI systems in the health domain can use HART as a starting point by either altering it or adding to it. Thereby, our proposed taxonomy contributes to the EU's initiative of Responsible Research and Innovation that requests a common language or set of concepts to enable comprehensive interdisciplinary work [44]. Moreover, this taxonomy presents a starting point for the assessment of ethical and legal risks within legally required risk assessments including data protection impact assessments as required by GDPR and conformity assessments as laid out in the EU's proposed AI Act.

HART assists organisations in conducting AI risk management and impact assessment by providing an open-access taxonomy of known risks. In addition, they can benefit from HART in addressing issues related to ethical and trustworthy AI. Since HART provides insights into the individual's interests at stake, organisations following a value-sensitive-design approach can more easily identify potential risks, further evaluate them, and decide on values guiding the entire development and design process.

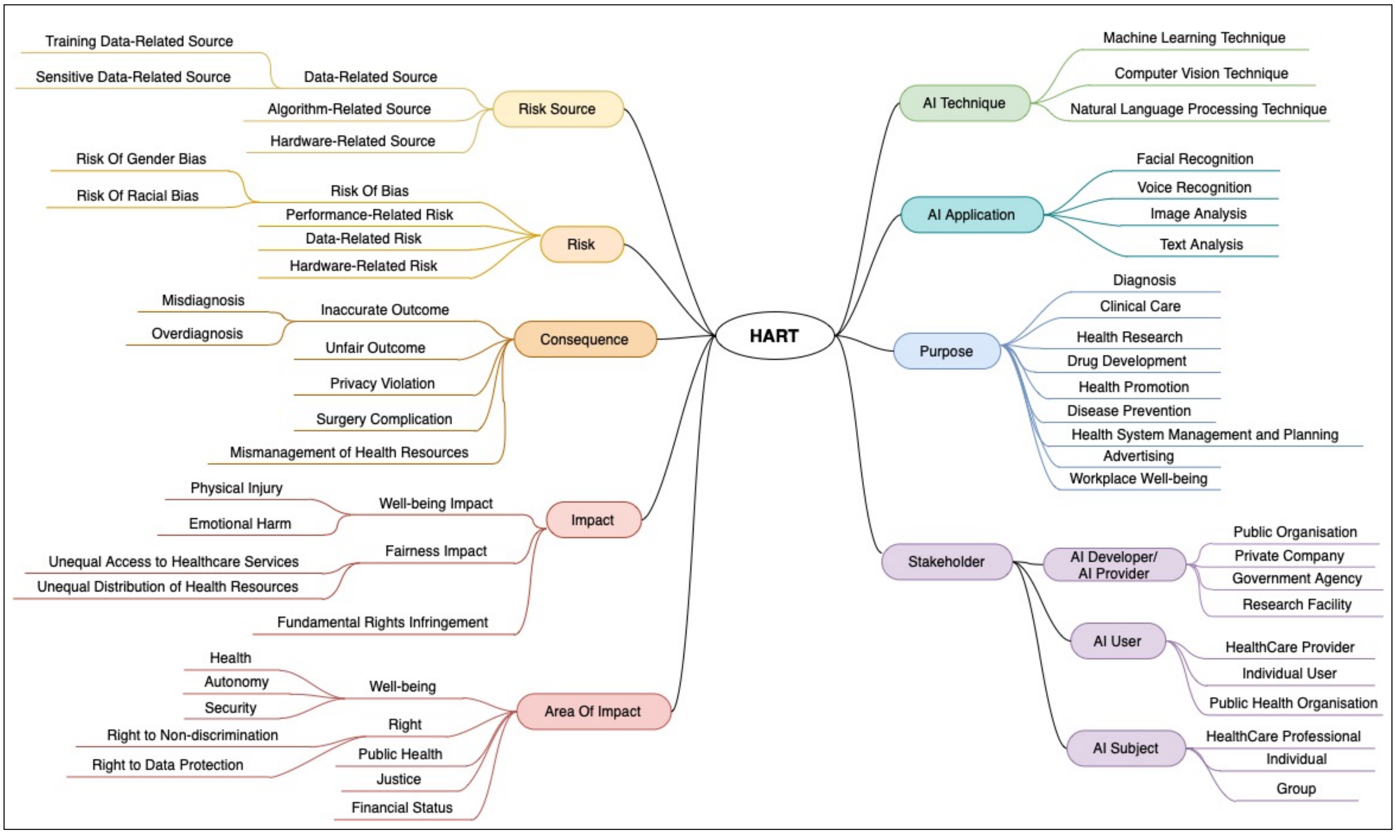

Fig. 1. An overview of HART

AI providers and users can gain a comprehensive overview on the potential risks and impacts, facilitating decisions on whether to deploy or use a given system or meeting legal requirements for risk assessments. This is especially relevant since an ethical and legal risk assessment is required by DPIAs.

Regulators may base the development of guidelines or regulations on a comprehensive understanding of the potential risks, their sources, and impacts. Particularly since many risks pose challenges for the non-violation of fundamental rights. Furthermore, HART might inform the creation or alteration of standards for the development and deployment of AI systems in the health domain.

B. The Application in Risk Based Assessment

Significantly the need for risk-based assessment lies at the heart of all four relevant legislative instruments which apply to AI in the Health Sector at a Union-wide level. The MDR and IVDR create unique risk profiles with specific assessment procedures. Article 35 of GDPR mandates that a DPIA be undertaken where the processing of data is likely to result in a "high risk to the rights and freedoms" of natural persons as has been demonstrated to be the case with AI use in the health sector. Finally, Article 9 of the Draft AI Act mandates a 'continuous iterative process' of risk identification and management to be run through the 'entire lifecycle' of high-risk systems such as AI in the health sector.

Article 35 (7) of GDPR sets out the minimum requirements for Data Protection Impact Assessments:

- The approach be systematic (i.e., should follow a defined and repeatable process).
- It should describe the proposed processing operations and their purposes.
- It should include an assessment of the risks to the rights and freedoms of individuals.
- It should define measures that are envisaged to address these risks, including safeguards, security measures, and other mechanisms to ensure the protection of personal data and to demonstrate compliance with data protection law, taking into account the rights and legitimate interests of data subjects and others.

Similarly, under Article 9(2) of the Draft AI Act a risk assessment must include

(a) identification and analysis of the known and foreseeable risks associated with each high-risk AI system;
(b) estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose and under conditions of reasonably foreseeable misuse;
(c) evaluation of other possibly arising risks based on the analysis of data gathered from the post-market monitoring system referred to in Article 61;
(d) adoption of suitable risk management measures.

We contend that HART facilitates risk management under all four Regulations by providing a structured framework for identifying use-cases and scenarios for the deployment of AI in the Healthcare sector. Through our taxonomy, it will be possible to (i) identify commonly occurring areas of risk, (ii) standardise templates for the documentation of impact assessments for the deployment of AI in specific circumstances, (iii) recommend common risk mitigation controls and auditing processes, and (iv) apply a 'design pattern' approach to any new use-cases or emerging technologies which researchers, medical practitioners, technology designers and developers may apply.

It is recommended that this taxonomy is developed as a structured framework to support scalable and repeatable processes for assessing and mitigating legal and ethical risks in AI deployment for healthcare.

V. Conclusion and Future Work

In the present work, we first provided discussions on the harmful risks associated with use of AI in the health sector from the ethical, societal, and legal perspectives. Then, we presented HART, a taxonomy for AI risks in the health domain, which is developed based on real-world AI incidents that occurred in the health domain from the AIAAIC repository.

HART assists organisations developing or using AI in the health domain in conducting AI risk and impact assessments as it introduces categories of risk sources, risks, consequences, and impacts. It also provides insights into the stakeholders who could be negatively affected by AI as well as the areas which might be adversely impacted. Adopting a structured and strategic approach to assessing and planning for the trustworthy and responsible use of AI in Health makes it possible to adapt appropriate technologies to the needs of researchers, policymakers, practitioners, and users. Our contribution can serve as the starting point for this work.

HART does not provide an exhaustive overview of risks in the health domain as it is built upon publicly available resources which usually do not provide detailed information regarding the risks. In our future work, we aim to extend the taxonomy by engaging stakeholders such as health domain experts and analysing resources such as the European Parliament's study on the use of AI in healthcare [45] and OECD's paper on trustworthy AI in health [22]. We also plan to explore the application of the taxonomy in real-world scenarios and evaluate the taxonomy using expert assessment.

Acknowledgment

We thank Víctor Rodríguez-Doncel from Ontology Engineering Group, Harshvardhan J. Pandit and Dave Lewis from ADAPT Centre for the discussions on the subject of this paper.

Citation: D. Golpayegani, J. Hovsha, L. W. S. Rossmaier, R. Saniei and J. Mišić, "Towards a Taxonomy of AI Risks in the Health Domain," 2022 Fourth International Conference on Transdisciplinary AI (TransAI), 2022, pp. 1-8, doi: 10.1109/TransAI54797.2022.00007.

References

[1] S. Sunarti, F. F. Rahman, M. Naufal, M. Risky, K. Febriyanto, and R. Masnina, "Artificial intelligence in healthcare: Opportunities and risk for future," *Gaceta Sanitaria*, vol. 35, S67–S70, 2021.

[2] D. Zhang, S. Mishra, E. Brynjolfsson, *et al.*, "The ai index 2021 annual report," *arXiv preprint arXiv:2103.06312*, 2021.

[3] K. Paldan, H. Sauer, and N.-F. Wagner, "Promoting inequality? Self-monitoring applications and the problem of social justice," *AI and Society*, pp. 1–11, 2018.

[4] C. Burr, J. Morley, M. Taddeo, and L. Floridi, "Digital psychiatry: Ethical risks and opportunities for public health and well-being," *IEEE Transactions on Technology and Society*, vol. 1, no. 1, pp. 21–33, 2020.

[5] R. Sparrow and J. J. Hatherley, "The Promise and Perils of AI in Medicine," *International Journal of Chinese and Comparative Philosophy of Medicine*, vol. 17, no. 2, pp. 79–109, 2019.

[6] T. Sharon, "The Googelization of health research: From disruptive innovaiton to disruptve ethics," *Personalized Medicine*, vol. 13, no. 6, pp. 563–574, 2016.

[7] B. de Boer and O. Kudina, "What is morally at stake when using algorithms to make medical diagnoses? Expanding the discussion beyond risks and harms," *Theoretical Medicine and Bioethics*, vol. 42, no. 5, pp. 245–266, 2021, Publisher: Springer Verlag.

[8] R. J. McDougall, "Computer knows best? The need for value-flexibility in medical AI," *Journal of Medical Ethics*, vol. 45, no. 3, pp. 156–160, 2019.

[9] E. Di Nucci, "Should we be afraid of medical AI?" *Journal of Medical Ethics*, vol. 45, no. 8, pp. 556–558, 2019.

[10] J. Morley and L. Floridi, "The Limits of Empowerment: How to Reframe the Role of mHealth Tools in the Healthcare Ecosystem," *Science and Engineering Ethics*, pp. 1–25, 2019.

[11] B. Rössler, *The value of privacy*. Cambridge: Polity Press, 2005.

[12] B. Murdoch, "Privacy and artificial intelligence: Challenges for protecting health information in a new era," *BMC Medical Ethics*, vol. 22, no. 1, pp. 1–5, 2021, Publisher: Biomed Central.

[13] T. Araujo, N. Helberger, S. Kruikemeier, and C. H. de Vreese, "In AI we trust? Perceptions about automated decision-making by artificial intelligence," *AI and Society*, vol. 35, no. 3, pp. 611–623, 2020.

[14] J. Morley, L. Floridi, and B. Goldacre, "The poor performance of apps assessing skin cancer risk," *British Medical Journal*, vol. 368, no. 8233, 2020.

[15] T. Grote, "Trustworthy medical AI systems need to know when they don't know," *Journal of Medical Ethics*, vol. 47, no. 5, pp. 337–338, 2021.

[16] C. Montemayor, J. Halpern, and A. Fairweather, "In principle obstacles for empathic AI: Why we can't

replace human empathy in healthcare," *AI and Society*, pp. 1–7,

[17] P. Shipley, "The keyboard blues: Modern technology and the rights and risks of people at work," *AI and Society*, vol. 9, no. 1, pp. 57–79, 1995, Publisher: Springer London.

[18] E. V. Vvedenskaya, "Ethical Problems of Digitization and Robotization in Medicine," *Russian Journal of Philosophical Sciences*, vol. 63, no. 2, pp. 104–122, 2020.

[19] J. A. Skorburg and J. Yam, "Is there an app for that?: Ethical issues in the digital mental health response to COVID-19," *American Journal of Bioethics Neuroscience*, pp. 1–14, 2021.

[20] C. Villongco and F. Khan, ""sorry i didn't hear you." the ethics of voice computing and ai in high risk mental health populations," *American Journal of Bioethics Neuroscience*, vol. 11, no. 2, pp. 105–112, 2020.

[21] S. Gerke, T. Minssen, and G. Cohen, "Ethical and legal challenges of artificial intelligence-driven healthcare," in *Artificial intelligence in healthcare*, Elsevier, 2020, pp. 295–336.

[22] *Trustworthy ai in health*, OECD, 2020.

[23] S. Monteith and T. Glenn, "Automated decision-making and big data: Concerns for people with mental illness," *Current Psychiatry Reports*, vol. 18, no. 12, pp. 1–12, 2016.

[24] N. D. Shah, E. W. Steyerberg, and D. M. Kent, "Big data and predictive analytics: Recalibrating expectations," *Jama*, vol. 320, no. 1, pp. 27–28, 2018.

[25] A. Verghese, N. H. Shah, and R. A. Harrington, "What this computer needs is a physician: Humanism and artificial intelligence," *Jama*, vol. 319, no. 1, pp. 19–20, 2018.

[26] C. J. Miller, D. K. McInnes, K. Stolzmann, and M. S. Bauer, "Interest in use of technology for healthcare among veterans receiving treatment for mental health," *Telemedicine and e-Health*, vol. 22, no. 10, pp. 847–854, 2016.

[27] K. Weigmann, "Health research 2.0: The use in research of personal fitness or health data shared on social network raises both scientific and ethical concerns," *EMBO reports*, vol. 15, no. 3, pp. 223–226, 2014.

[28] T. Z. Zarsky, "Understanding discrimination in the scored society," *Wash. L. Rev.*, vol. 89, p. 1375, 2014.

[29] World Health Organization, *Ethics and governance of artificial intelligence for health*, 2021.

[30] F. Khan, "The uberization of healthcare: The forthcoming legal storm over mobile health technology's impact on the medical profession," *Health matrix*, vol. 26, p. 123, 2016.

[31] European Union, *Medical devices regulation 2017/745 oj l 117*, 2017.

[32] ——, *In-vitro diagnostic devices regulation 2017/746 oj l 117*, 2017.

[33] ——, *Regulation (eu) 2016/679, 27 april 2016, general data protection regulation ("gdpr") oj l 119*, 2016.

[34] European Commission, *Artificial intelligence act: Proposal for a regulation of the european parliament and the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts*, 2021.

[35] European Union, *Charter of fundamental rights of the european union*, 2000.

[36] The Working Party on the Protection of Individuals with regard to the Processing of Personal Data, *Guidelines on automated individual decision-making and profiling for the purposes of regulation 2016/679*, 2017.

[37] CNIL, *Délibération n° 2018-326 du 11 octobre 2018 portant adoption de lignes directrices sur les analyses d'impact relatives à la protection des données (aipd) prévues par le règlement général sur la protection des données (rgpd)*, 2018.

[38] N. F. Noy and D. L. McGuinness, *Ontology development 101: A guide to creating your first ontology*, 2001.

[39] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, and R. Garcıa-Castro, "Lot: An industrial oriented ontology engineering framework," *Engineering Applications of Artificial Intelligence*, vol. 111, p. 104 755, 2022.

[40] D. Golpayegani, H. Pandit, and D. Lewis, "Airo: An ontology for representing ai risks based on the proposed eu ai act and iso risk management standards," in *International Conference on Semantic Systems (SEMANTiCS)*, in press, 2022.

[41] H. Pandit, "A semantic specification for data protection impact assessments (dpia)," in *International Conference on Semantic Systems (SEMANTiCS)*, in press, 2022.

[42] M. Poveda-Villalón, P. Espinoza-Arias, D. Garijo, and O. Corcho, "Coming to terms with fair ontologies," in *International Conference on Knowledge Engineering and Knowledge Management*, Springer, 2020, pp. 255–270.

[43] *Iso 31000 risk management — guidelines*, International Standardization Organization, 2018.

[44] R. von Schomberg, "Introduction: Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields," in 2011, pp. 7–15.

[45] K. Lekadir, G. Quaglio, A. Tselioudis Garmendia, and C. Gallin, *Artificial intelligence in healthcare applications, risks, and ethical and societal impacts*, European Parliament, 2022.