Investigating AI in Medical Devices: The Need for Better Establishment of Risk-Assessment and Regulatory Foundations

Sandra Baum[1] and Konstantinos Manikas[1,2(✉)]

[1] Computer Science Department, IT -University of Copenhagen, Rued Langgaards Vej 7, 2300 Copenhagen, Denmark
{sanb,koma}@itu.dk
[2] Accenture Consulting, Bohrsgade 35, 1799 Copenhagen V, Denmark

Abstract. Artificial intelligence (AI) has the potential to revolutionize healthcare in the EU by addressing challenges, such as shortages of healthcare personnel and more effective diagnosis and care. However, the safety concerns surrounding AI-based medical devices have been a major roadblock to the technology's wider adoption. This study aims to further investigate these concerns in the European context by analysing the AI-enabled Medical devices currently available in the European Union market along with their potential safety risks. We do this by applying a combination of three research methods: (1) a survey of the safety risks of AI-enabled Medical Devices published between 2012 and 2023, (2) an analysis of AI-based medical devices in the EUDAMED database, and (3) a survey on the perceptions of the EU Medical AI ecosystem stakeholders. Our study analyzed the state-of-the-art with a literature body of 29 papers and summarized a number of risks related to the use of AI in medical devices along with the reported mitigation strategies. Furthermore, we analyzed the approved medical devices (71 devices) that use AI in the EUDAMED database and found that there is a lack of transparency in whether the devices use AI along with the lack of crucial information necessary to asses the devices' safety risks, such information on training data. Finally, when we survey a number of medical device stakeholders (7 out of 130 respondents) we find that there is a disconnect between the industry and regulators: the medical device representatives emphasize the need for better guidance on post-market surveillance while the regulation representatives feel that they lack expertise in AI.

Keywords: Artificial intelligence · Medical device regulation · Literature survey · Medical device survey

1 Introduction

Artificial intelligence (AI) solutions, like ChatGPT are increasingly entering various aspects of our lives. This tendency is arguably also occurring in the medical

D. Salvi et al. (Eds.): PH 2023, LNICST 572, pp. 51–69, 2024.
https://doi.org/10.1007/978-3-031-59717-6_4

device area [32]. AI-based medical device software holds great promise in addressing the challenges faced by healthcare systems in the European Union, such as the aging population, inefficient medical systems, and lack of healthcare workers. However, these AI-enabled solutions also come with risks, from inaccurate predictions to incorporating various biases. These issues raise concerns about safety risks, which can consequently lead to a lack of trust and pose a barrier to the wide-scale adoption of AI into clinical practices. Lack of information about these devices and mitigation of various risks further decreases trust. In the EU context various aspects of AI- enabled medical devices, such as their characteristics, are unexplored. This paper aims to provide an overview of risks associated with AI-based medical devices and describing the AI-based medical software devices currently on the EU market, with focus on factors affecting their safety. The core questions explored in this paper are:

1. What are the safety risks of AI-enabled medical Devices, and what strategies exist to mitigate them?
   Extensive focus has been put into creating frameworks for evaluating AI-based Medical Devices [5]. However, to the best of the author's knowledge, no survey of the safety risk of such devices has so far been conducted.
2. What kind of AI-based Medical Devices can be currently found on the EU market?
   There has been a lot of discussions and work put into regulating AI in the European Union. However, in comparison to the USA regulatory body (FDA), EU is lagging behind in terms of providing information about AI-enabled Medical Devices. Indeed, until the launch of EUDAMED there lacked a central database of Medical Devices on the EU market.
3. How do the stakeholders of the AI-enabled Medical Device ecosystem perceive the use, risks and regulation of AI-enabled Medical Devices?
   Analysis on stakeholders' perception of Medical AI is so far largely focused on healthcare specialists and their views on AI [40,43]. However, little is known how companies, researchers and regulators perceive the current use of AI various safety risks of AI-enabled Medical Devices and the regulation on Medical-AI in EU.

The rest of the paper is structured as follows: in Sect. 2, we analyze the background and related works; in Sect. 3, we provide a brief overview of the regulation of medical devices in EU; in Sect. 4, we explain the methodologies used in various research steps; in Sect. 5, we present the findings of the research and in Sect. 6, we provide an analysis of the research findings. Following that, we provide a discussion section, where we delve into the implications of our findings.

This paper aims to contribute to the current discussions of legislative and regulatory reforms intended to regulate AI/ML-based medical devices.

2 Background and Related Work

Most studies on AI-based medical devices focus on the US market and on devices approved by the FDA. For example, Wu et al. [41] published a comprehensive

overview of medical AI devices approved by the US Food and Drug Administration, that indicated that evaluation process can mask vulnerabilities of devices when they are deployed on patients. Muehlematter et al. [28] report on a comparative study of Medical Devices approved by FDA and CE-marked in EU between years 2015–2020.

While many papers have investigated the safety risks of AI-based medical devices [15,29,36], we were not able to find a dedicated literature survey of the safety risks of AI-based medical devices.

3 Methodological Approach

In this study we apply a combination of quantitative and qualitative approaches to present multiple findings about AI-based Medical Devices. This mixed approach is chosen to enable triangulation in order to examine the current use and potential safety risks of AI-enabled Software as a Medical Device and from research literature, devices on the EU market and practitioner's viewpoint. The approach applied is: (a) we review the literature of safety risks associated with AI-enabled medical devices; (b) we analyze the current AI-enabled Medical Devices on the EU market achieved by the collection and manual labelling of data from the European Database of Medical Devices EUDAMED; and (c) we survey the stakeholders of the ecosystem of the European medical devices.

3.1 Literature Survey of Risks of AI-Enabled SaMD

We conduct a literature survey on the risk of AI-enabled software as a medical device (SaMD)[1]. We define a protocol based on the PRISMA methodology [30] and by leveraging our previous experience on literature surveys and systematic literature reviews [22–24,39]. Our survey protocol includes:

Sources. The defined literature sources are: (i) Google Scholar, (ii) PubMed, and (iii) Scopus.

Search string. *((safety) OR (risks)) AND (healthcare) AND (((machine learning)) OR (deep learning)) OR (artificial intelligence)*[2].

Inclusion/exclusion criteria. In order for the paper to be included in the Literature Review the following criteria has to be met: 1) Paper discusses the safety risks of AI enabled devices in medicine; 2) The paper is from the time period 2012–2023; 3) The paper is in English;

[1] SaMD is defined by the International Medical Device Regulators Forum (IMDRF) as "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.".

[2] Further variation of search keywords were tested that included, among other, "AI ML & Safety & Medical Device & Medicine"; "ML & Safety Risks & Medical Device & Healthcare".

The papers are screened by title, abstract, and full text against the inclusion and exclusions criteria defined. After the papers were selected for inclusion, backwards and forward snowballing is used to find further relevant papers and gather a comprehensive and diverse set of studies that are relevant to the research question being addressed.

For all of the papers reviewed, the following information are extracted: (1)Type of article: journal, conference article or book; (2) Bibliographic data such as publication year; (3) Safety risks listed/discussed; (4) Ways of mitigating the safety risks if they were listed/discussed; (5) Reviewer notes, comments, and recommendations from surveying the article.

Although the current study focuses on medical devices limited in the EU region, geographical limitations were intentionally excluded from the literature survey to ensure an adequate literature body and variability in results.

3.2 EU AI-Based Medical Device Survey Protocol

We survey the approved software medical devices in EU and identify the devices with AI-supported functions. To do so, we survey the European database on medical devices (EUDAMED). We extract[3] all software medical devices and collect (a) Trade name, (b) Manufacturer and (c) Classification (risk class).

Having collected the initial device body, validated and cleaned the data, we process it as following. In this study we focus on medium to high risk devices, thus we exclude the low risk devices (class I and Class A) from the dataset. This group is chosen for exclusion, since the devices are subjected to a different, less rigorous approval process.

Furthermore, EUDAMED does not currently provide information on the description of the device, including whether a device is using AI or not. Thus, the resulting data are manually annotated as either AI and non-AI for filtering out non-AI devices. To validate the device data, we follow a three-source approach: FDA list of AI/ML devices, AI for Radiology database, and device publicly available data. As the first step of annotation the devices are cross-referenced with our first two sources: the FDA list of Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices and the AI for Radiology database. AI for Radiology database is a database of CE-marked AI software products for clinical radiology based on vendor-supplied product specifications [18]. The FDA list of AI-enabled Medical Devices is a non-exhaustive list periodically updated by FDA based on publicly available information [6]. The rest of the devices are then manually labeled based on the publicly available data on Google. In this step, we use the device manufactures websites and press releases as primary data sources.

The resulting dataset is categorized by medical specialties using a modified version of the European Union of Medical Specialists' list, created in collaboration with two medical experts. The modified list aims to include all relevant

[3] We apply the Python library BeautifulSoup with extraction date 2022.09.29.

specialties while avoiding excessive granularity for the paper's purpose. Furthermore, the devices are categorized based on the risk categorization principles of the International Medical Device Regulators Forum (IMDRF) set out in Possible Framework for Risk Categorization and Corresponding Considerations [11].

Therefore the devices are further manually labeled across following dimension: the point of healthcare situation or condition the software is intended to be used in; the body part targeted and the medical speciality the device belongs to.

Furthermore, from the point of healthcare situation or condition the software is intended to be used in, the devices were classified in the following categories: critical situation or condition; serious situation or condition; Non-serious situation or condition [11]. Classifying the devices from the point of healthcare situation or condition is done by two labellers - one working in the healthcare sector and another in IT.

3.3 Medical Device Stakeholder Survey

In order to get a more complete view on the area, we conduct a survey of the EU stakeholders in AI-based medical device area. The main focus areas are the current use of AI-based medical devices in EU, potential risk factors and EU legislation regarding Medical Devices. In this survey we intend to validate the findings from the literature survey and rate the risks of using AI-based medical devices. The survey is conducted online with the survey link being sent out to the potential participants. Before launching, the survey is piloted it using a pre-screening [17]. The survey is pre-screened by four researchers with knowledge of both medical and IT field. All lists of options for multiple-choice questions, with the exception of the question about participants role, are randomized to decrease potential measurement errors [17].

The survey has four sections: *Background information* aimed at defining the role of the participant (Expert, Working in SAMD company, Regulator or Other); *AI in the EU market* aimed at defining areas with AI-enabled devices and possible overuse of them; *Safety risks of AI-enabled medical devices* focusing on rating and prioritizing the risks that are noted in the literature; and *Regulation of AI-enabled devices in EU* focusing on the largest technical challenge for ensuring the safety of AI/ML based Medical Devices in EU.

Furthermore the survey collects input in whether the participants felt that the current medical device regulation in EU is sufficient in terms of ensuring the safety of AI/ML based Medical Devices and what changes they would like to see in the medical device regulation in the EU. In this analysis we categorized the participant roles as following: *regulator*, a person working in a AI- enabled SaMD company or an expert in the area. *Expert*, a person who had published research in AI enabled Medical Devices in EU or was or had been part of an expert group or think tank such as the EU expert group on AI. *Regulator*, included in Notified Bodies in the European Union found by looking through the NANDO database for Notified Bodies. SAMD using AI companies were companies which in publicly declared using AI in their devices. Such companies were found firstly by the list

of companies found on the EUDAMED database and manually labeled as AI and secondly using various online databases for companies, such as Danish startup ecosystem database or EIT health database. Experts were found by looking at speakers at relevant conferences, authors of relevant papers or looking at interest groups representing Medical-Device manufacturers. The survey was sent to 130 individuals, including 31 experts, 50 regulators, and 49 individuals from SaMD companies.

4 Findings

4.1 Literature Survey

When applying the literature survey protocol we retrieve a total of 27.073 results. 24 of the originally resulted papers are included in the final literature body. An additional five papers are added by snowballing. The literature survey process can be summarized in Fig. 1. The complete reference list can be found in the Appendix 7. Details of the included papers are also summarised in Fig. 1. The papers covered the time period 2016–2022; they came from the European Union (EU), United Kingdom (UK), United States of America (USA), and Canada; and they were primarily journal articles. Journals published in 2022 were most common.

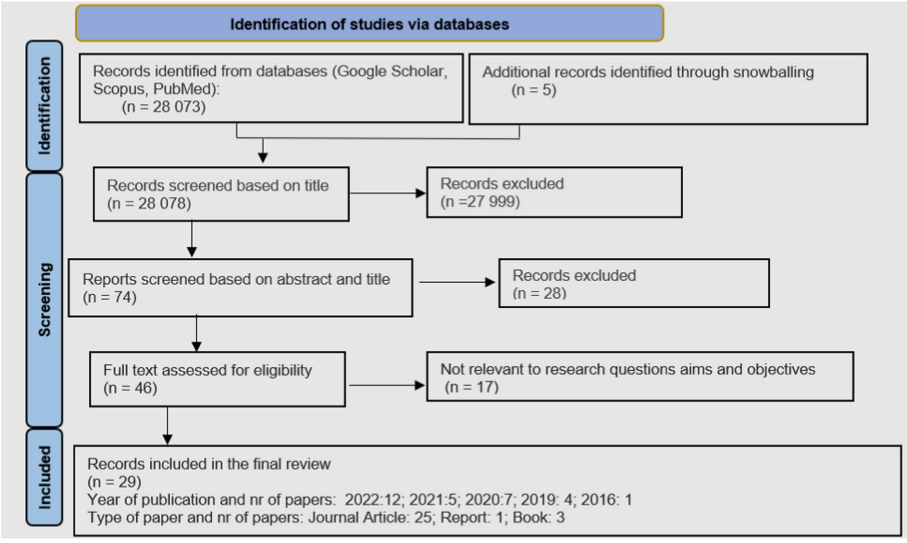

Fig. 1. Process for defining the literature body.

The risks identified in the literature survey can be classified into four categories: data-centric risks, transparency-related risks, cyber-security risks, and

risks stemming from user-machine interaction. Summary of various risks and their mitigation strategies can be found in Table 1 at the end of the chapter.

Results of the survey show that data-related risks are most prevalent in the literature. These include aspects such as data drift, distributional shift or calibration drift.

Data-Centric Risks [1,3,12,13,19,21,25,34,38]. Many risks mentioned in the literature stem from the data used for the Artificial Intelligence.

Bias is a prominent topic in literature and mainly stems from disparities between training and operational data. Bias errors occur, since machine learning models do not generalize well beyond the data they were trained on [21]. Bias has different forms such as distributional shift, which arises when the underlying data distribution used for model development differs from the data where the model is deployed [34]. For instance, a skin cancer prediction model may perform poorly on dark-skinned patients [3], revealing a distributional shift due to selection bias, which can occur when marginalized populations are not adequately represented in the training data.

If not effectively implemented, evaluated, and regulated, AI solutions in the future may perpetuate and amplify systemic disparities and human biases, contributing to healthcare inequities [19].

Distributional shift can also occur due to minor difference in the radiology equipment in different hospitals [21] resulting in medical images, such as X-rays with slightly different characteristics. Another sub-type of bias is calibration drift, which can occur due to unanticipated changes in clinical practices or patient behaviour [38].

However, bias can also stem from factors beyond the shortcomings of inadequate training data. Such as measurement bias - omitting critical data-fields during model training. For example an algorithm predicting survival of post-menopausal women, that did not perform well, partly because it lacked relevant blood test results [38]. Another source for bias can be incorrect data. This is especially true in cases where data from consumer-facing health apps are merged with clinical data to create predictions. An instance is the Fitbit PurePulse Trackers' unreliable heart rate measurements [13].

Bias is further worsened by the characteristics of data-sets available. Healthcare data is often sparse and imbalanced, for example contain more samples of patients with a mild condition, due to naturally occurring distribution. This is especially prominent in fields such as pathology and mental health [1].

Further data related risks include noise and artefacts in model inputs and hidden confounders. *Noise and artefacts in model inputs* - Noise in data refers to meaningless or irrelevant data, that the model can pick up on [12]. Noisy data is often caused by the differences in or issues with medical equipment used. For example, scanning errors or differences in hospital imagining protocols [25]. Noise can also come from Imaging artifacts and poor imaging quality [25].

Hidden Confounders. are factors unmeasured in the observational data affecting both treatment and outcome. An example of hidden confounders in a clinical

setting are physicians prescribing medication based on indicators not present in the health record [16]. They can reduce both model generalizability and interpretation [42].

Transparency-Related Risks [14,19–21,34,37,42]. Risks related to the explainablility or interpretability also referred to as transparency of AI based devices were also a common topic in the literature.

Lack of Transparency: i.e. the 'black box' nature of many AI systems, like deep learning models, makes their decision-making process unclear [20,21,42]. This lack of transparency can make it difficult to determine the accuracy or reliability of the AI's output, may erode trust of patients and healthcare specialists and may make it more difficult to identify and correct errors [14,34,37]. A well-known example of a difficultly identifiable error is a case of AI models predicting pneumonia mortality risk mistakenly labeled asthma patients as lower risk of mortality, since they were treated more aggressively and quickly according to hospital protocol, which reduced their risk of death [42]. Transparency can be split into [19]: *Traceability* (clarity of AI development and usage) and *Explainability* (clarity of AI decisions).

User and System Interaction Related Risks [7,19–21,26,31,38]. The least discussed aspect influencing the safety of the devices was User and System Interaction related risks. Examples of such risks are input errors, automation complacency, and cognitive bias of the user. Input errors can occur as a result of the user misspelling, confusing clinical terms, users employing local definitions or misrepresenting findings. This issue is further exacerbated by quickly changing medical definitions [7]. Automation complacency refers to when specialists rely too heavily on the models predictions. Research has shown that specialists tend to over-rely and delegate full responsibility to systems and lose vigilance or become deskilled [21,31]. Evidence suggests that when a clinician is uncertain, they may defer to models predictions [26]. The black box nature of many modern AI-systems will likely contribute to the worsening of this phenomenon [21,38]. Over time, automation complacency might lead to misdiagnoses and inappropriate interventions, as algorithms may lean towards overdiagnosis by detecting subclinical findings [38].

Cognitive bias of the user includes errors that are closely related to automation complacency of the users. Cognitive biases have many forms and can include to misunderstandings of statistics and mathematical rationality or be one of many forms of human cognitive biases, such as Search satisfying: Ceasing to look for further information or alternative answers when the first plausible solution is found [7].

Various User and System Interaction risks are exacerbated by healthcare specialists limited knowledge of AI. Varied studies have showed that, that healthcare specialists have received little education regarding AI and do not rate their knowledge of AI highly hidden [19].

Cybersecurity Risks [2,38]. AI-enabled medical devices largely share common cybersecurity risks with non-AI healthcare systems, but the use of AI in healthcare increases exposure to data privacy and integrity risks, due to creating an increased need between the interconnectivity between systems and dataset [2]. Resulting attacks can compromise model accuracy, lead to harmful predictions, re-identify de-identified data, or result in data loss. Various cybersecurity risks are discussed below.

AI increases *reidentification* opportunities in anonymized patient datasets, exemplified by Liangyuan's research [2], which demonstrated that over 90% of adults' physical activity data could be reidentified using ML models.

Adversarial Attacks on AI, categorized into white-box attacks that employ subversion, such as gradient-based techniques, and black-box attacks that poison datasets [2] leading to harmful or incorrect predictions or undetectable software corruption, are not easily detectable [38].

The risks discussed can also interplay and mutually amplify each other, such as the interaction between sampling and diagnostic bias or automation complacency and lack of transparency, with the latter making it more difficult to identify bias in the training data.

Mitigation Strategies. Most papers included in the survey presented potential mitigation strategies for the risks. In this chapter mitigation strategies from the summary table warranting additional clarification are described.

Transparency. *Visualization tools for increased transparency.*

Visualization tools for ML predictions, like Local Interpretable Algorithm-Agnostic Explanations (LIME) and Shapley Values (SHAP), help visualize the key features influencing the algorithm's predictions However, a challenge remains in clinicians understanding the language of these explanations. To address this, a platform connecting medical experts with ML researchers could help establish standardized representations of explanations [35].

Cybersecurity. *Encryption* various encryption measures are usually employed for data in transfer [2].

Adversarial Training - machine learning technique, which improves models robustness and generalization ability by training it to learn data samples that are designed to be have small and often human-imperceptible differences from the original data, but, which a model misclassifies [9]. For example images, with added pixels. This technique helps the model becomes more resistant to errors and to better handle real-world inputs that may be similarly ambiguous [8].

Masking Measures. Masking techniques, such as adding random statistical noise, collapsing variables, creating synthetic data, or using ML models to generate statistically similar datasets, are commonly employed when sharing data with external stakeholders to protect sensitive information [2].

GAN-Generative Adversarial Network (GAN) is a type of machine learning model, that can be used to generate adversarial data for a model to classify to bolster a model's robustness against attacks [33].

Statistical Approaches - using statistical tests, adversarial inputs from the operational data can be detected. Statistical tests rely on the fact that adversarial examples are statistically different from other inputs [10].

Federated Learning. This technique allows the training of an algorithm on sensitive data, present at multiple decentralized sites, without the exchange of data. For example, a number of hospitals can contribute toward the training of a model, without the data itself ever leaving each hospital's data center [42].

4.2 EU AI-Based Medical Device Survey

At the time of the gathering the data of this paper[4], the EUDAMED database lists 955 medical software items, which are reduced to 765 unique devices after eliminating duplicates. Excluding lower risk devices left 327, with 5 listed in the AI Radiology database, 13 in the FDAs database for AI based medical devices The other devices are manually labeled following the protocol outlined in Chap. 3. This results in 71 AI devices. These are labeled as serious (10), non-serious (20), or critical (41). More AI devices are classified as class IIa (low to moderate risk) than non-AI devices (72% versus 68%). The most common target body parts are the heart (13 devices) and lungs (10 devices), with 13 devices targeting multiple parts. Some devices belong to two specialities. The most common speciality the device are aimed at was radiology with 28 devices, followed by cardiology with 14 devices.

4.3 Survey of Stakeholder Perceptions

The survey is send out to 130 potential respondents. Seven provided a valid response. Four of the respondents are experts and three are working in a SaMD company. The responses do not include any regulators. However, one of the regulators reports that they feel that they do not have sufficient information to fill out the survey.

The respondents report that *pathology* and *emergency* medicine are areas that AI can be used while *radiology* and *nuclear medicine* are areas where AI is underused.

Participants are requested to evaluate various aspects of AI-enabled medical devices for their potential impact on device safety. The selected characteristics are based on prominent elements from the literature survey and guidelines from the International Medical Device Regulators Forum [11].

In this question, the participant assess that whether the devices are: (a) informing of options for treatment/diagnosing, or (b) for aiding in treatment or in diagnoses. (b) had the most influence. This element received a score of 4,6

[4] Extraction date: 2022.09.29.

Table 1. Risks and mitigation strategies identified in the literature

Risks and Mitigation Strategies	
Risk	Mitigation Strategy
Bias [3,21]	1. Pooling of data from various countries and organisations to create large and diverse data sets,across various areas, such as race and ethnicity [3,42] 2. Verify AI technology product claims on local data set [3] 3. Comprehensive multi-location evaluation studies to identify instabilities [19] 4. Reporting performance of models across relevant subgroups [26]
Hidden Confounders	No Mitigation Strategy Suggested
Noise and artefacts in model inputs [12]	Polishing, such as relabeling of data, or filtering out the noise [12]
Adversarial Attacks	1. Adversarial training 2. Generative Adversarial Network (GAN) 3. Statistical approaches [2]
Data Privacy Attacks [2]	1. De-identification algorithms [2] 2. Federated approaches for decentralised AI [19] 3. Full disk encryption [2] 4. Masking measures [2]
Lack of Tranparency (Blackbox nature of AI) [20,21]	1. An 'AI passport' for standardised description and traceability of medical AI tools [19] 2. Auditing [21] 3. For some models, visualization software, i.e. as SHAP and LIME [27,38]
Input errors [7]	Providing the user with background information and a glossary of clinical terms used in the model [7]
Cognitive biases	Training healthcare specialists to not lose vigilance [21]
Automation Complacency [21]	1. Improving the interpretability of AI systems [21] 2. Curriculum combining medicine and engineering to allow for better understanding of the workings of models [4] 3. Training healthcare specialists to not lose vigilance [21]

(in a scale from 1-5). The aspect receiving the lowest score in terms of influence was *the medical specialisation the device is deployed in* with an average of 3. Whether it is used for critical, serious or non critical, illness/condition received an average rating of 4.1 The remaining scores were: *Testing the AI algorithm on data from the hospital where it is deployed in the launch phase* - 3.28. *How much data has been used to train the device* - 4.14. *Interaction between the user and the device* - 4.14. *The users understanding of AI/ML* - 3.28.

When examining the current regulation of AI enabled devices three participants did not feel that the current system in EU was sufficient in terms of ensuring the safety of AI/ML based Medical Devices, while two participants were unfamiliar with the system and two felt that the current system was sufficient. Reasons noted for feeling that the system was insufficient were: (i) No guidance had been published by the European authorities on surveillance following implementation; (ii) Notified Body scrutinises for safety, clinical experts review Clinical Evaluation Report; (iii) Does not sufficiently account for potential biases or oversights in the training and validation data; (iv) Re-certification of data-centric AI and learning algorithms are not fully incorporated; (v) Lack of life-cycle understanding;

Lastly, the changes the participants would like to see in the medical device regulation were: (a) more focus on post market surveillance; (b) more streamlined process, that are less dependant on the availability of notified bodies or their specific interpretations; (c) Better guidelines for post-market surveillance; (d) Better guidelines on ensuring sufficiency of data.

5 Analysis

5.1 Literature Survey

It is evident from the literature survey that the main risks stem from the AI-enabled devices reliance and interaction with data - not only during the pre-launch phase, but also during production.

This is due to the fact, that unlike traditional medical devices that function in a rather predictable, deterministic way, AI-enabled devices can evolve and change their behavior based on the data they interact with.

This means that for AI-enabled devices, post-market surveillance and real-world performance monitoring are as, if not more, important. This requires a change in regulatory frameworks, which have traditionally focused heavily on the pre-market phase where devices are tested extensively in controlled lab settings.

Second aspect, unique to AI-enabled medical devices, is risks related to the interpretability of AI-enabled Devices. Dangerous or unhelpful patterns learned by the model can be difficult to detect, as for many AI- models it is difficult and at times impossible to understand why they have reached a certain conclusion. Furthermore, the lack of transparency can exacerbate risks related to user-system interaction. To address this, regulatory frameworks need to include requirements on the level of transparency and interpretability of AI systems. A very promising direction is the use of explainable AI (XAI) techniques, that aim to make the decision-making process of AI systems more interpretable to humans.

In conclusion, the dynamic nature of AI-enabled medical devices, as well as their complexity, calls for a significant shift in thinking when designing regulatory frameworks.

5.2 AI-Based Medical Devices in EU

The current data fields available in EUDAMED point to a possible regulatory issue, as they do not contain a lot of information that would be needed to assess the safety of AI-based devices, such as information about the data - for example potential biases in the training data or amount of data used for training. As the review of the safety risks showed, having clear documentation about the design of the system, including the data used, helps mitigate risks associated with AI-enabled devices, such as the black-box nature of such devices.

Analysis of the distribution of risk classes revealed that the medical AI devices in the EUDAMED dataset had a higher proportion of "class IIa" and "class IIb" risk classes and a lower proportion of "IVD general" risk classes compared to the non-AI devices.

"Class IIa" risk classes are considered to be of low to moderate risk, while "class IIb" risk classes are considered to be of moderate risk. "IVD general" risk classes are not included in class IIa or IIb, and their risk level is not specified.

This suggests that the medical AI devices in EUDAMED are often lower to moderate risk compared to the non-AI devices. It is worth noting that this comparison is based on the proportion of devices in each risk class, and it is not necessarily indicative of the overall risk level of the medical and non-AI devices.

The second finding of the analysis of EUDAEMD is that most AI-enabled devices are dealing with critical or serious illnesses and conditions, such as stroke or cancer. Analysis of the correlation between the severity of the condition or illness targeted and the risk class of the device showed that devices targeting serious illnesses and conditions doe not necessarily get a higher risk class. This is not surprising as, the severity of the targeted condition or illness is just one factor among many considered when assigning a risk class to a medical device, with the intervention of the device carrying the most significant weight.

5.3 Survey of Stakeholder Perceptions

The low number of survey participants means that the results are not suitable for representing the medical device ecosystem as a whole, since such a small sample size can lead to sampling bias. However, the results are still useful for supplementing the literature review and for providing insights into potential safety concerns form the stakeholders perspective. Furthermore, the finding points out potential pain-points in the EU regulation of medical devices from the stakeholders perspective. Additionally, the results could potentially be used to inform future research or to identify areas for improvement in the distribution process. The fact that none of the regulators filled out the survey coupled with the fact that one of the regulators reported that he feels that they do not have sufficient information to fill out the survey, points to possible gap in regulators knowledge of AI- enabled devices. While it must be noted that since we only have one datapoint we currently have weak evidence. None the less, this is an interesting finding that could point to a future research direction. It is possible that the regulator who wrote back indicating that they did not have sufficient knowledge

to fill out the survey represents a broader trend among regulators. This could point to a serious issue in the EU legislation of AI-enabled medical devices.

The concerns and pain-points pointed out by the participants show that many risks in AI -enabled medical devices are data-centric, such better guidelines on sufficient amounts of data. This feedback from actors currently active in the ecosystem indicates that EU regulators have not sufficiently addressed several data-centric risks of AI-enabled medical devices.

Furthermore the survey participants also underlined the need for better post-market guidance. This highlights another unique feature of AI that the EU might not have tackled sufficiently. Namely the changing nature of AI models and algorithms and the large amounts of risks stemming of it that can manifest in the post-market phase. Unlike many traditional medical devices such as contact lenses or pumps, AI performance can vary widely in different locations, for example in different hospitals and can also dangerously degrade when coming in contact with new data while in productions. These unique aspects would need to be clearly addressed by the EU regulators.

6 Discussion

The majority of AI devices are identified in the survey on EUDAMED were within the field of radiology or cardiology and most commonly dealt with critical conditions or illnesses, such as cancer or stroke. The fact that devices on the EU market commonly deal with critical illnesses highlights the potential severe outcomes of various risks not being properly mitigated. It was challenging for this study to identify AI in devices as the current information in EUDAMED is inadequate. EUDAMED does not contain information on whether a device is utilising AI and lacks information that would be needed to assess the safety of AI-based devices, e.g. information on the data used for the device.

In the literature survey of the risks and mitigation strategies of AI-enabled Medical Devices most papers discussed data-centric risks in various detail. Other risk categories identified were cybersecurity risks, transparency related risks and lastly user and system interaction related risks. The amount and nature of risks identified in combination with the domain mission criticality of the devices underline the importance of good praxis in the adaptation of AI and the high risks in improper adaptation.

The analysis of stakeholder perceptions found that several post-market data-related risks presented in the academic literature, such as bias, were also a concern for the stakeholders, who emphasised the need for better testing and regulatory guidance to address such risks. Some stakeholders felt that the current EU regulation on Medical-AI is inadequate, citing a lack of post-implementation guidance and guidelines on data sufficiency as examples. This points to a need for regulatory guidelines that in a larger degree take into account the dynamic and data-centric properties of AI enabled medical devices. However, there was indication that regulators feel a lack of expertise about AI.

The findings of this paper highlight lack of regulation and establishment of common understanding of safety risks of AI-enabled Medical Devices. This

can be attributed in part to the immaturity of the field, however, the potential impact of risks in the medical domain are severe. The dynamic nature of AI models compared to traditional medical devices requires a stronger focus on the post-market phase from the regulators. Therefore, the study underscores the need for a more comprehensive understanding, and clear and robust regulatory guidelines to navigate through these potential hazards.

7 Conclusion

In this paper we investigate the adoption of AI in medical devices. Currently, concerns regarding the safety risks surrounding AI-based medical devices currently stand in the way of their wider adoption. In this study we conduct: (1) a survey of the safety risks of AI-enabled Medical Devices published between 2012 and 2023, (2) an analysis of AI-based medical devices in the EUDAMED database. and (3) a survey on the perceptions of Medical AI ecosystem stakeholders. Our analysis body includes 29 reviewed papers, 71 AI-based medical devices and seven responded questionnaires out of an original 130 participants. Our findings show that the presence of unique risks, such as bias or lack of transparency, in AI-enabled Medical Devices is undeniable. Looking at data available at EUDAMED we can see that it is currently hard to even pinpoint which devices in EU use AI and we have to look at company websites, press statements or published papers to discover that. We also uncovered that many AI enabled devices in EU deal with severe conditions such as arrhythmia or stroke, which further underlines the severity of potential risks manifesting. Experts and companies in the Medical AI ecosystem feel a need for guidance and regulation that covers the whole life-cycle of AI products, with more emphasis on the post-market phase, and incorporates aspects related to data-centric risks of the products. This demonstrates an openness to more structured guidelines from the industry. However, our research suggests that regulators feel that do not have expertise in AI, indicating that a gap exists between the complexities of AI technology and the understanding of those responsible for its oversight. Based on the findings we propose, that more clear and encompassing regulatory guidelines would be needed to mitigate the risks of AI-enabled Medical Devices in EU.

Appendix A - Literature Body of the Literature Survey

[21] Magrabi, F., Ammenwerth, E., McNair, J.B., De Keizer, N.F., Hyppönen, H., Nykänen, P., Rigby, M., Scott, P.J., Vehko, T., Wong, Z.S.Y., et al.: Artificial intelligence in clinical decision support: challenges for evaluating ai and practical implications. Yearbook of medical informatics **28**(01), 128–134 (2019)
[38] Scott, I., Carter, S., Coiera, E.: Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health & Care Informatics **28**(1) (2021)
[2] Bohr, A., Memarzadeh, K.: Artificial intelligence in healthcare. Academic Press (2020)

[35] Rasheed, K., Qayyum, A., Ghaly, M., Al-Fuqaha, A., Razi, A., Qadir, J.: Explainable, trustworthy, and ethical machine learning for healthcare: A survey. Computers in Biology and Medicine p. 106043 (2022)
[9] Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014)
[8] Geiping, J., Fowl, L., Somepalli, G., Goldblum, M., Moeller, M., Goldstein, T.: What doesn't kill you makes you robust (er): Adversarial training against poisons and backdoors. arXiv preprint arXiv:2102.13624 **1**(7) (2021)
[33] Qiu, S., Liu, Q., Zhou, S., Wu, C.: Review of artificial intelligence adversarial attack and defense technologies. Applied Sciences **9**(5), 909 (2019)
[10] Grosse, K., Manoharan, P., Papernot, N., Backes, M., McDaniel, P.: On the (statistical) detection of adversarial examples. arXiv preprint arXiv:1702.06280 (2017)
[42] Xing, L., Giger, M.L., Min, J.K.: Artificial intelligence in medicine: technical basis and clinical applications. Academic Press (2020)
[3] Borycki, E., Kushniruk, A.: Artificial intelligence and safety in healthcare. In: AI and Society, pp. 17–32. Chapman and Hall/CRC (2022)
[19] Lekadir, K., Quaglio, G., Garmendia, A.T., Gallin, C.: Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts. EPRS (European Parliamentary Research Service) (2022)
[26] McCradden, M.D., Joshi, S., Anderson, J.A., Mazwi, M., Goldenberg, A., Zlotnik Shaul, R.: Patient safety and quality improvement: Ethical principles for a regulatory approach to bias in healthcare machine learning. Journal of the American Medical Informatics Association **27**(12), 2024–2027 (2020)
[20] Macrae, C.: Governing the safety of artificial intelligence in healthcare. BMJ quality & safety **28**(6), 495–498 (2019)
[27] Moore, C.M.: The challenges of health inequities and ai. Intelligence-Based Medicine p. 100067 (2022)
[7] Galitsky, B., Goldberg, S.: Artificial Intelligence for Healthcare Applications and Management. Academic Press (2022)
[4] Briganti, G., Le Moine, O.: Artificial intelligence in medicine: today and tomorrow. Frontiers in medicine **7**, 27 (2020)
[31] Paton, C., Kobayashi, S.: An open science approach to artificial intelligence in healthcare. Yearbook of medical informatics **28**(01), 047–051 (2019)
[42] Xing, L., Giger, M.L., Min, J.K.: Artificial intelligence in medicine: technical basis and clinical applications. Academic Press (2020)
[37] Rubinger, L., Gazendam, A., Ekhtiari, S., Bhandari, M.: Machine learning and artificial intelligence in research and healthcare. Injury (2022)
[34] Quinn, T.P., Jacobs, S., Senadeera, M., Le, V., Coghlan, S.: The three ghosts of medical ai: Can the black-box present deliver? Artificial intelligence in medicine **124**, 102158 (2022)
[14] Jia, Y., McDermid, J.A., Lawton, T., Habli, I.: The role of explainability in assuring safety of machine learning in healthcare. IEEE Transactions on Emerging Topics in Computing (2022)

[25] Martin, C., DeStefano, K., Haran, H., Zink, S., Dai, J., Ahmed, D., Razzak, A., Lin, K., Kogler, A., Waller, J., et al.: The ethical considerations including inclusion and biases, data protection, and proper implementation among ai in radiology and potential implications. Intelligence-Based Medicine p. 100073 (2022)
[12] Gupta, S., Gupta, A.: Dealing with noise problem in machine learning datasets: A systematic review. Procedia Computer Science **161**, 466–474 (2019)
[1] Barh, D.: Artificial Intelligence in Precision Health: From Concept to Applications. Academic Press (2020)
[13] Hamid, S.: The opportunities and risks of artificial intelligence in medicine and healthcare. Apollo - University of Cambridge Repository (2016)
[16] Kallus, N., Puli, A.M., Shalit, U.: Removing hidden confounding by experimental grounding. Advances in neural information processing systems **31** (2018)

References

1. Barh, D.: Artificial Intelligence in Precision Health: From Concept to Applications. Academic Press, Cambridge (2020)
2. Bohr, A., Memarzadeh, K.: Artificial Intelligence in Healthcare. Academic Press, Cambridge (2020)
3. Borycki, E., Kushniruk, A.: Artificial intelligence and safety in healthcare. In: AI and Society, pp. 17–32. Chapman and Hall/CRC, Boca Raton (2022)
4. Briganti, G., Le Moine, O.: Artificial intelligence in medicine: today and tomorrow. Front. Med. **7**, 27 (2020)
5. Crossnohere, N.L., Elsaid, M., Paskett, J., Bose-Brill, S., Bridges, J.F.: Guidelines for artificial intelligence in medicine: literature review and content analysis of frameworks. J. Med. Internet Res. **24**(8), e36823 (2022)
6. Center for Devices and Radiological Health: Artificial intelligence and machine learning (AI/ML)-enabled medical d, October 2022. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices
7. Galitsky, B., Goldberg, S.: Artificial Intelligence for Healthcare Applications and Management. Academic Press, Cambridge (2022)
8. Geiping, J., Fowl, L., Somepalli, G., Goldblum, M., Moeller, M., Goldstein, T.: What doesn't kill you makes you robust (ER): adversarial training against poisons and backdoors. arXiv preprint arXiv:2102.13624 **1**(7) (2021)
9. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014)
10. Grosse, K., Manoharan, P., Papernot, N., Backes, M., McDaniel, P.: On the (statistical) detection of adversarial examples. arXiv preprint arXiv:1702.06280 (2017)
11. Group, I.S.W., et al.: "Software as a medical device": possible framework for risk categorization and corresponding considerations. In: International Medical Device Regulators Forum (2014)
12. Gupta, S., Gupta, A.: Dealing with noise problem in machine learning data-sets: a systematic review. Procedia Comput. Sci. **161**, 466–474 (2019)
13. Hamid, S.: The Opportunities and Risks of Artificial Intelligence in Medicine and Healthcare. Apollo - University of Cambridge Repository (2016)

14. Jia, Y., McDermid, J.A., Lawton, T., Habli, I.: The role of explainability in assuring safety of machine learning in healthcare. IEEE Trans. Emerg. Top. Comput. (2022)
15. Jiang, L., et al.: Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies. J. Int. Med. Res. **49**(3), 03000605211000157 (2021)
16. Kallus, N., Puli, A.M., Shalit, U.: Removing hidden confounding by experimental grounding. In: Advances in Neural Information Processing Systems, vol. 31 (2018)
17. Lavrakas, P.J.: Encyclopedia of Survey Research Methods. Sage Publications, Thousand Oaks (2008)
18. van Leeuwen, K.G., Schalekamp, S., Rutten, M.J., van Ginneken, B., de Rooij, M.: Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur. Radiol. **31**(6), 3797–3804 (2021)
19. Lekadir, K., Quaglio, G., Garmendia, A.T., Gallin, C.: Artificial intelligence in healthcare: applications, risks, and ethical and societal impacts. EPRS (European Parliamentary Research Service) (2022)
20. Macrae, C.: Governing the safety of artificial intelligence in healthcare. BMJ Qual. Saf. **28**(6), 495–498 (2019)
21. Magrabi, F., et al.: Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearb. Med. Inform. **28**(01), 128–134 (2019)
22. Manikas, K.: Revisiting software ecosystems research: a longitudinal literature study. J. Syst. Softw. **117**, 84–103 (2016). https://doi.org/10.1016/j.jss.2016.02.003, https://www.sciencedirect.com/science/article/pii/S0164121216000406
23. Manikas, K.: Supporting the evolution of research in software ecosystems: reviewing the empirical literature. In: Maglyas, A., Lamprecht, A.-L. (eds.) Software Business. LNBIP, vol. 240, pp. 63–78. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-40515-5_5
24. Manikas, K., Hansen, K.M.: Software ecosystems – a systematic literature review. J. Syst. Softw. **86**(5), 1294–1306 (2013). https://doi.org/10.1016/j.jss.2012.12.026, https://www.sciencedirect.com/science/article/pii/S016412121200338X
25. Martin, C., et al.: The ethical considerations including inclusion and biases, data protection, and proper implementation among AI in radiology and potential implications. Intell.-Based Med. 100073 (2022)
26. McCradden, M.D., Joshi, S., Anderson, J.A., Mazwi, M., Goldenberg, A., Zlotnik Shaul, R.: Patient safety and quality improvement: ethical principles for a regulatory approach to bias in healthcare machine learning. J. Am. Med. Inform. Assoc. **27**(12), 2024–2027 (2020)
27. Moore, C.M.: The challenges of health inequities and AI. Intell.-Based Med. 100067 (2022)
28. Muehlematter, U.J., Daniore, P., Vokinger, K.N.: Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. Lancet Digit. Health **3**(3), e195–e203 (2021)
29. Newaz, A.I., Sikder, A.K., Rahman, M.A., Uluagac, A.S.: A survey on security and privacy issues in modern healthcare systems: attacks and defenses. ACM Trans. Comput. Healthc. **2**(3), 1–44 (2021)
30. Page, M.J., et al.: The Prisma 2020 statement: an updated guideline for reporting systematic reviews. Syst. Control Found. Appl. **10**(1), 1–11 (2021)
31. Paton, C., Kobayashi, S.: An open science approach to artificial intelligence in healthcare. Yearb. Med. Inform. **28**(01), 047–051 (2019)

32. Powell, A.: AI Revolution in Medicine. Harvard Gazette, November 2020. https://news.harvard.edu/gazette/story/2020/11/risks-and-benefits-of-an-ai-revolution-in-medicine/
33. Qiu, S., Liu, Q., Zhou, S., Wu, C.: Review of artificial intelligence adversarial attack and defense technologies. Appl. Sci. **9**(5), 909 (2019)
34. Quinn, T.P., Jacobs, S., Senadeera, M., Le, V., Coghlan, S.: The three ghosts of medical AI: can the black-box present deliver? Artif. Intell. Med. **124**, 102158 (2022)
35. Rasheed, K., Qayyum, A., Ghaly, M., Al-Fuqaha, A., Razi, A., Qadir, J.: Explainable, trustworthy, and ethical machine learning for healthcare: a survey. Comput. Biol. Med. 106043 (2022)
36. Ross, P., Spates, K.: Considering the safety and quality of artificial intelligence in health care. Jt. Comm. J. Qual. Patient Saf. **46**(10), 596 (2020)
37. Rubinger, L., Gazendam, A., Ekhtiari, S., Bhandari, M.: Machine learning and artificial intelligence in research and healthcare. Injury (2022)
38. Scott, I., Carter, S., Coiera, E.: Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health Care Inform. **28**(1) (2021)
39. Seppänen, M., Hyrynsalmi, S., Manikas, K., Suominen, A.: Yet another ecosystem literature review: 10+1 research communities. In: 2017 IEEE European Technology and Engineering Management Summit (E-TEMS), pp. 1–8 (2017). https://doi.org/10.1109/E-TEMS.2017.8244229
40. Sujan, M.A., White, S., Habli, I., Reynolds, N.: Stakeholder perceptions of the safety and assurance of artificial intelligence in healthcare. Saf. Sci. **155**, 105870 (2022)
41. Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D.E., Zou, J.: How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nat. Med. **27**(4), 582–584 (2021)
42. Xing, L., Giger, M.L., Min, J.K.: Artificial Intelligence in Medicine: Technical Basis and Clinical Applications. Academic Press, Cambridge (2020)
43. Yang, L., Ene, I.C., Arabi Belaghi, R., Koff, D., Stein, N., Santaguida, P.L.: Stakeholders' perspectives on the future of artificial intelligence in radiology: a scoping review. Eur. Radiol. **32**(3), 1477–1495 (2022)