

Article

Not peer-reviewed version

An 'Algorithmic Ethics' Effectiveness Impact Assessment Framework' for developers of Artificial Intelligence (AI) systems in healthcare

[Elsa Papadopoulou](#)*, [Themis Exarchos](#)*, [Demetris Gerogiannis](#), Joana Namorado

Posted Date: 4 June 2024

doi: 10.20944/preprints202405.2154.v1

Keywords: Artificial Intelligence; Ethics; Applied Ethics; Bioethics; Computational Ethics; Trustworthy AI; Professional Certification; Medical Devices; Safety; Efficacy; Effectiveness; 'Ethics' Due Diligence'



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

An 'Algorithmic Ethics' Effectiveness Impact Assessment Framework' for developers of Artificial Intelligence (AI) systems in healthcare

Elsa Papadopoulou ^{1,*}, Demetris Gerogiannis ², Joana Namorado ³ and Themis Exarchos ²

¹ Bioinformatics and Human Electrophysiology Lab (BiHELab), Dept. of Informatics, Ionian University, Corfu, Greece; elsapapa30@yahoo.fr

² Computer Vision and Image Processing, Dept. of Computer Science & Engineering, University of Ioannina, Greece; dgerog@gmail.com

³ Citolab Lda; joananamorado1@gmail.com

⁴ Data Modeling and Decision Support Systems, Department of Informatics, Ionian University, Corfu, Greece; themis.exarchos@gmail.com

* Correspondence: elsapapa30@yahoo.fr or c19papa2@ionio.gr; Tel.: +32 (0) 471653122

Abstract: Algorithmic systems used in healthcare contexts are primarily developed for the benefit of the public. It is therefore essential that these systems are trusted by the individuals for whose benefit they are deployed. Drawing inspiration from the principles embedded in the testing of the safety, efficacy and effectiveness of new medicinal products, concurrent design engineering and professional certification requirements, the authors propose, for the first time, a preliminary competency-based 'Algorithmic Ethics' Effectiveness Impact Assessment' framework for developers of AI systems used in healthcare contexts. They concluded that this set of principles should encompass the algorithmic systems 'production lifecycle', to guarantee the optimized use of the AI technologies, avoiding biases and discrimination while ensuring the best possible outcomes, simultaneously increasing decision-making capacity and the accuracy of the results. As AI is as good as those who program it and the system in which it operates [1], the robustness and trustworthiness of their 'creators' and 'deployers', should be fostered by a certification system guaranteeing the latter's knowledge and understanding of ethical aspects as well as their competencies in integrating these aspects in trustworthy AI systems when used in healthcare contexts.

Keywords: Artificial Intelligence; Ethics; Applied Ethics; Bioethics; Computational Ethics; Trustworthy AI; Professional Certification; Medical Devices; Safety; Efficacy; Effectiveness; 'Ethics' Due Diligence'

1. Introduction

Interest in medical devices incorporating AI/ML functionality has increased in recent years and even more so in recent months due to the development of large language models (LLMs). LLMs are AI models that are trained on very large datasets, enabling them to recognize, summarize, translate, predict, and generate content (for example: ChatGPT, Llama, Claude, PaLM, etc.).

The use of any type of AI system, in health, patient care and public health requires particular attention to ensure that the public, both healthy individuals and patients, trusts that the system is properly scrutinized and evaluated, that it is beneficial, fair and conforms to strict standards of quality and ethics. A holistic, transdisciplinary 'Ethics Due Diligence' approach can greatly contribute to an effectively positive ethics' impact assessment enhancing the public's trust towards not only the use of AI technologies but equally important to those who design, develop, deploy and use such technologies in healthcare contexts.

For the purposes of this paper, algorithmic systems are considered either as a medical device themselves or are embedded in medical devices or software systems that enable e.g. diagnosis of

health conditions, drug discovery, treatment recommendation, etc. [2] These systems are used in a wide array of applications in healthcare contexts, for example: detecting clinical conditions in medical imaging and diagnostic services, providing virtual patient care using AI-powered tools, managing electronic health records, augmenting patient engagement and compliance with the treatment plan, reducing the administrative workload of healthcare professionals (HCPs), discovering new drugs and vaccines, spotting medical prescription errors, extensive data storage and analysis, technology-assisted rehabilitation, etc. Nevertheless, this science pitch meets several technical, ethical, and social challenges, such as privacy, safety, choosing who is in a most urgent need for a transplant, the costs of using AI systems in the healthcare provision and reimbursement of such costs ensuring access to the benefits offered by the use of such systems to all, information and consent, efficacy and accuracy of the analysis produced by the AI systems, etc.

Despite existing regulatory processes and provisions that guarantee the safety, efficacy and effectiveness of new medicinal products or medical devices, similar ones for algorithmic systems used in healthcare are very scarce due to the still relatively nascent phase of the use of AI technologies in healthcare contexts.

The governance of AI applications is crucial for patient safety and accountability and for raising the HCPs' belief in enhancing acceptance and boosting significant health consequences. Effective governance of AI systems is a prerequisite to precisely address regulatory, ethical, and trust issues while advancing the acceptance and implementation of AI [3]. Trustworthy AI should not serve as a mere axiom to enable a higher return on investment for AI systems' developers but it should be translated in practical and measurable actions to enable a proper 'due diligence' of the systems' trustworthiness.

In February 2023, the OECD published a report presenting research and findings on accountability and risk in AI systems by providing an overview of how risk-management frameworks and the AI system lifecycle can be integrated to promote trustworthy AI. One of the ten principles put forward in this report refers to the accountability that AI actors should bear for the proper functioning of the AI systems they develop and use. This means that AI actors must take measures to ensure their AI systems are trustworthy – i.e. that they benefit people; respect human rights and fairness; are transparent and explainable; and are robust, secure and safe. To achieve this, actors need to govern and manage risks throughout their AI systems' lifecycle – from planning and design, to data collection and processing, to model building and validation, to deployment, operation and monitoring. The report also identifies four important steps, which can help manage AI risks throughout the system's lifecycle: (1) definition of scope, context, actors and criteria; (2) assessment of the risks at individual, aggregate, and societal levels; (3) treatment of the risks in ways commensurate to cease, prevent or mitigate adverse impacts; and (4) producing a governance framework for the risk management process. Risk management should be an iterative process whereby the findings and outputs of one step continuously inform the others. [4] The risk management governance process should be designed and rolled out as a continuous, dynamic, transdisciplinary feedback loop allowing for a holistic risk 'due diligence'.

Over the past decade, the US Food & Drug Administration (FDA) reviewed and authorized a growing number of devices (marketed via 510(k) clearance, granted De Novo request, or premarket approval) with AI/ML across many different fields of medicine—and expects this trend to continue. As of October 19, 2023, no device was authorized by the US FDA that uses generative AI or artificial general intelligence (AGI) or is powered by LLMs [5]. Furthermore, in 2014, the International Medical Device Regulators Forum (IMDRF) Software as a Medical Device Working Group (WG) published a possible risk categorization framework for Software as a Medical Device. The recommendations provided in this document allow manufactures and regulators to more clearly identify risk categories of Software as a Medical Device based on how the output of a Software as a Medical Device is used for healthcare decisions in different healthcare situations or conditions [6]. In January 2021, the U.S. Food and Drug Administration (FDA) issued the "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan" from the Center for Devices and Radiological Health's Digital Health Center of Excellence (CDRH). The Action Plan was a direct

response to stakeholder feedback to an April 2019 discussion paper, "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning-Based Software as a Medical Device" and outlined five actions the FDA intended to take. The FDA's CDRH considered a total product lifecycle-based regulatory framework for these technologies that would allow for modifications to be made from real-world learning and adaptation, while ensuring that the safety and effectiveness of the software as a medical device are maintained [7]. As per the findings of the consultation that led to the production of the action plan, stakeholders called, among others, for a patient-centered approach incorporating transparency to users. To enhance such patient-centered approach, the development and utilization of AI/ML-based devices need to take into consideration issues such as trust, equity and accountability [8]. In March 2024, the FDA published the "Artificial Intelligence and Medical Products: How CBER, CDER, CDRH, and OCP are Working Together," which represents the FDA's coordinated approach to AI. This paper is intended to complement the "AI/ML SaMD Action Plan" and represents a commitment between the FDA's Center for Biologics Evaluation and Research (CBER), the Center for Drug Evaluation and Research (CDER), and the Center for Devices and Radiological Health (CDRH), and the Office of Combination Products (OCP), to drive alignment and share learnings applicable to AI in medical products more broadly [9].

In September 2022, the European Commission proposed a Directive on adapting non contractual civil liability rules to artificial intelligence (the 'Liability Rules for AI'). The objective of this proposal was to promote the rollout of trustworthy AI. It foresees that victims of damage caused by an AI system could obtain equivalent protection to victims of damage caused by products in general. It also reduces legal uncertainty of businesses developing or using AI regarding their possible exposure to liability and prevents the emergence of fragmented AI-specific adaptations of national civil liability rules [10].

In December 2023, the European Parliament and the Council of the EU reached a political agreement on the EU AI Act, which is expected to enter into force in June 2024. Most of its provisions will become applicable two years after its entry into force. The EU AI Act is the first-ever comprehensive legal framework on AI worldwide. The aim of the new rules is to foster trustworthy AI in Europe and beyond, by ensuring that AI systems respect fundamental rights, safety, and ethical principles and by addressing risks of very powerful and impactful AI models. It foresees a Regulatory Framework that defines 4 levels of risk for AI systems, unacceptable risk, high risk, limited risk, and minimal risk. Among the systems identified as 'High Risk' include AI technology used in safety components of products as for example in the case of an AI application integrated into a medical device [11]. In conjunction with the EU AI Act, the EU Medical Devices regulation [12] governs the safety, performance, and quality of medical devices placed on the market and their use within the European Union. It aims to protect public health and ensure high standards of quality and safety for medical devices. Whereas, the EU AI Act covers a broad range of AI systems, encompassing both general-purpose and specialized AI applications, the EU MDR specifically focuses on medical devices, which may or may not use AI technology, with AI being thus a secondary consideration. Furthermore, the EU AI Act primarily aims to protect the safety and fundamental rights of individuals interacting with AI systems. The EU MDR, on the other hand, is concerned with ensuring the safety, performance, and quality of medical devices, regardless of whether they use AI or not. Due to the broad definition in the EU MDR, many AI systems used in health could be classified as a medical device. Both the EU AI Act and the MDR adopt a risk-based approach. The EU AI Act categorizes AI systems into four risk levels, while the EU MDR classifies medical devices into different risk classes (I, IIa, IIb, and III) based on their potential impact on patient safety and public health. Also, the EU AI Act emphasizes transparency and traceability of AI systems, with requirements for clear and concise information on the system's operation, purpose, and limitations and the EU MDR also mandates transparency and traceability for medical devices, including labeling and documentation requirements. Additionally, both the EU AI Act and the EU MDR stress the importance of human oversight and control over the respective technologies. For AI systems, this may include human oversight in high-risk situations, while for medical devices, it may involve post-market surveillance and monitoring.

In December 2023, the European Medicines Agency (EMA) and the Heads of Medicines Agencies (HMAs) published an AI workplan to 2028, setting out a collaborative and coordinated strategy to maximise the benefits of AI to stakeholders while managing the risks. The workplan will help the European medicines regulatory network (EMRN) to embrace the opportunities of AI for personal productivity, automating processes and systems, increasing insights into data and supporting more robust decision-making to benefit public and animal health [13].

In April 29, 2024, the NIST released a draft publication based on the AI Risk Management Framework (AI RMF) to help manage the risk of Generative AI. The draft AI RMF Generative AI Profile aims to help organizations identify unique risks posed by generative AI and proposes actions for generative AI risk management that best aligns with their goals and priorities. The NIST AI Risk Management Framework (AI RMF) is intended for voluntary use and to improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems. Among others, the AI RMF states that *“for AI systems to be trustworthy, they often need to be responsive to a multiplicity of criteria that are of value to interested parties. Approaches which enhance AI trustworthiness can reduce negative AI risks. This Framework articulates the following characteristics of trustworthy AI and offers guidance for addressing them. Characteristics of trustworthy AI systems include: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed. Creating trustworthy AI requires balancing each of these characteristics based on the AI system’s context of use. While all characteristics are socio-technical system attributes, accountability and transparency also relate to the processes and activities internal to an AI system and its external setting. Neglecting these characteristics can increase the probability and magnitude of negative consequence”* [14]. In the draft AI RMF Generative AI Profile, states among others that *“The integration of GAI systems can involve varying risks of misconfigurations and poor interactions. Human experts may be biased against or “averse” to AI-generated outputs, such as in their perceptions of the quality of generated content. In contrast, due to the complexity and increasing reliability of GAI technology, other human experts may become conditioned to and overly rely upon GAI systems. This phenomenon is known as “automation bias,” which refers to excessive deference to AI systems. Accidental misalignment or mis-specification of system goals or rewards by developers or users can cause a model not to operate as intended. One AI model persistently shared deceptive outputs after a group of researchers taught it to do so, despite applying standards safety techniques to correct its behavior. While deceptive capabilities is an emergent field of risks, adversaries could prompt deceptive behaviors which could lead to other risks”* [15].

The objective of the above and other pertinent non-legislative, legislative and regulatory texts is to enhance the public’s trust in the use of AI systems in, among others, healthcare contexts as well. As aforementioned, guaranteeing the safety, efficacy and effectiveness of medical devices integrating AI systems, is of primordial importance in order to ensure a wider uptake of such powerful technologies and harness their potential to the fullest. Failure to provide this guarantee can compromise entire areas of Health and Clinical Studies and Public Health. It has proven to be detrimental to the acceptance or development of entire areas of research and can induce increasing resistance and public distrust. In the present challenging economic environment, the resources mobilized during the phases of product research and development are scrutinised as much as the actual objectives, and results.

The ultimate goal and purpose of use of AI systems in healthcare contexts is the preservation or improvement of healthy individuals or patients’ conditions. With the understanding that the underlying purpose of the use being to ‘Do Good’ rather than simply ‘Do No Harm’. Therefore, one needs to be able to assess and measure, not only an algorithm’s efficiency but also its compliance with the set of principles that can yield a ‘Do Good’ result.

Recent regulatory and legislative texts targeting AI systems in general and more specifically those used in healthcare contexts, suggest that the type, the transparency and particularly, the quality of data selected to train the AI models that are used as a departure point in these contexts is decisive to the resulting quality and relevance of the model. These data sets are also used for the analyses and further along the line, for the mitigation of any bias. This requires that adequate efforts are put into

eliminating, or at least into mitigation of “Publication Bias” and to ensure that negative results as well as positive ones, be part of the datasets analysed.

The current paper shall introduce two novel notions: that of an ‘Ethics Due Diligence’ as well as that of a competency-based certification framework for professionals that design and deploy AI systems in healthcare contexts. Such a framework will serve as an ‘Algorithmic Ethics Effectiveness’ Impact Assessment’ measurement in real-world settings and should be tailored to the ever-evolving landscape of AI technologies. It is designed to be dynamic and adaptable and to identify critical skill categories for AI professionals, including regulatory compliance, ethical use and bias removal, validation and testing, continuous monitoring and feedback, deployment and scalability, risk assessment and mitigation, and security and privacy. The framework will enable the AI professionals and users exercise their creativity, duty of care and service provision to harness their potential to the maximum while ensuring that the ultimate goal for which the AI systems are used in healthcare is served: the preservation and protection of human life by abiding to a ‘Do Good’ principle.

Throughout the paper the term ‘AI uses in healthcare’ reflects the entire cycle of processes encompassing the maintenance and/or improvement of an individuals health via the use of AI systems for the prevention, diagnosis, treatment and monitoring of an individual’s physical and mental well-being.

2. Materials and Methods

The authors combined their own knowledge and expertise in legal, medical and computer science with the review of literature and practices pertinent to **applied ethics, bioethics, medical ethics, due diligence, design thinking, value sensitive design, concurrent engineering principles, algorithmic systems design and development lifecycle**.

Based on the insights provided by a transdisciplinary approach of a legal professional, a medical doctor and a computer scientist, this paper discusses how principles emanating from the aforementioned methodologies, approaches and practices combined with notions from professional certification requirements and provisions, can ensure the safety, efficacy and effectiveness of responsible AI in healthcare contents through the use of an ethics compliance framework for those who design and deploy such systems.

3. Results

From theory to practice: Preliminary set of principles for an ‘Algorithmic Ethics’ Impact Assessment, Competency-based Certification Framework’ for AI systems used in healthcare contexts

To foster a trustworthy AI system while helping the developers to conceive, design, deploy, test and maintain the AI systems’ ethics efficacy and robustness, and help protect the AI systems’ developers from any liability risks emanating from the relevant legislative and regulatory texts allowing them at the same time to innovate and provide to citizens high-fidelity AI systems beneficial for their health and well-being, the authors propose the below set of questions-principles as a basis for a future certification system for the developers themselves.

The classification of the questions-principles follows the development and deployment stages of an AI system:

- Development:
 - Design,
 - Database creation,
 - Learning
 - Training
 - Validation
 - Testing
 - Deployment
 - Calibration
 - Use

- Maintenance
- Improvement

Design

1. Am I aware of how 'Human-Centered Designed' principles can be embedded in the design of algorithms and models?
2. Have I embedded 'Human-Centered Designed' principles during the conception and development phase of my algorithmic systems?
3. Am I aware of the notion of 'ethically designed' algorithms?
4. Have I taken into consideration ethics aspects when designing my algorithmic systems?
5. Do I understand what a high-risk AI system is?
6. Have I integrated risk mitigation aspects in the design process of my algorithmic systems?

Database Creation

1. Are my algorithmic systems providing an accurate explanation, which correctly reflects the reason for generating the output and/or accurately reflects the system's process?
2. Can I explain your algorithmic system's 'knowledge limits', i.e. which data was used, which data was not used and why?
3. Have I embedded a review process of the dataset's update and bias detection? Have I created a log of adversarial findings? Have I introduced corrective measures to eliminate inaccuracies and bias? Have I registered the positive results of the process?

Learning

1. Have I created a log and a reporting template as well as a governance framework concerning these processes of the information on my AI systems behaviour?

Calibration

1. Am I aware of the regulatory and legislative texts that apply to my activities?
2. Have I reviewed whether my planned or implemented AI use cases fall under one of the regulated categories under the relevant regulatory and legislative texts?
3. Have I embedded the obligations stemming from the relevant regulatory and legislative provisions in my algorithmic systems?
4. Have I ensured that my algorithmic systems do not pose an unacceptable risk to fundamental rights?
5. Does my algorithmic system deliver or contain accompanying evidence or reason(s) for outputs and/or processes ensuring compliance with ethics and non-infringement of fundamental rights provisions?

Use

1. Have I embedded principles to ensure that the outcomes are understandable to the intended consumer(s)?
2. Am I aware of the legal consequences of the potential harms that my algorithmic systems will bear for me and anyone involved and have I applied due care in order to mitigate the potential harms and the legal consequences?

Maintenance

1. Have I ensured that all relevant stakeholders have been involved in the design, testing and deployment phase of my algorithmic system?
2. Have I described in the design, testing and deployment protocol the stakeholders involved, their contributions, the adverse effects' log and the solutions?

Improvement

1. Can I provide explanations on the outcomes produced by my algorithmic system?
2. What is the testing and validation governance framework of the AI system that I have designed?

4. Discussion

Applied Ethics

Applied ethics is a branch of philosophical inquiry that involves the application of moral principles and values to real-world situations and problems. Unlike theoretical ethics, which focuses on the nature and foundations of morality, applied ethics deals with practical issues and dilemmas that arise in everyday life, such as in professional settings, public policy, and personal relationships. The goal of applied ethics is to provide a structured, principled approach to addressing complex moral issues in real-world settings.

In the case of medical ethics, applied ethics would involve the practical application of moral principles to issues like patient confidentiality, informed consent, end-of-life care, and the allocation of scarce medical resources. By using a framework of moral principles, professionals in applied ethics aim to make principled decisions in complex situations, balancing possibly priorities and values, for example identifying which patient would more urgently need a kidney transplant.

The study of applied ethics requires not only an understanding of philosophical concepts, but also an awareness of the practical realities and contexts in which decisions are made. It involves not just abstract theorizing, but also careful deliberation and reasoning, often within a transdisciplinary framework.

Combining 'applied ethics' and 'trustworthy AI' involves integrating moral principles and values into the design, development, and deployment of AI systems. This is a multifaceted process that involves various stakeholders, including AI developers, policymakers, users, and ethicists.

To do so an 'AI Ethics Risk Due Diligence' framework should be developed and used, serving as a systematic approach to identify, assess, and mitigate potential risks and harms associated with AI systems used in healthcare contexts from an ethical perspective. Such a framework should combine the following steps:

1. Identification and definition of the scope and objectives of the AI system, including the intended purpose of use, the target users or population, and any potential secondary applications combined with the moral values and principles to guide the development and use of trustworthy AI within the specific healthcare context. These values typically include transparency, fairness, privacy, accountability, and human oversight.

2. Embedding the moral values and principles in the conception and establishment of guidelines and frameworks for the design, development, and deployment of AI systems. These guidelines should be agile and ensure clarity and transparency. AI developers should incorporate these values and principles into the design, development, and deployment of AI systems. This can be achieved through the use of techniques like Value-Sensitive Design (VSD), where AI systems are designed to adhere to specific moral values. Design Thinking (DT) is based, among others, on the principle that empathy is embedded in the design phase with the users for whom the innovation is developed to understand their pains and problems fully. The latter, in turn, is converted to the Human-Centred Design (HCD) that focuses on understanding the perception, the needs, and expectations of the person who are looking for a solution to a specific problem and whether the proposed solution has been designed in a way to and will effectively and efficiently resolve the problem for which it was designed. HCD can be further enriched by Value Sensitive Design (VSD) principles, which is a method that embeds values into a technical design.

As proposed in an analysis published at the beginning of 2021 by Steven Umbrello & Ibo van de Poel (2021), VSD could be integrated into AI systems design to address the challenges posed by the need for transparency, explicability, and accountability of AI systems as well as those posed by Machine Learning (ML) which may lead to AI systems adapting in ways that "disembody" the values embedded in them.

As an example, a study discussed the moral precepts and how could these VSD principles be operationalized in the design of the Quality of Life (QoL), QoL-ME, which is an eHealth and mHealth

application that is expected to address important human values in the tool's design, using VSD principles for integrating important human values during the development of the tool. (Maathuis, Niezen, Buitengeweg, Bongers, & Nieuwenhuizen, 2020). [16].

4. Ensure transparency and explainability meaning that users should be able to understand how AI systems make decisions and the reasoning behind these decisions. This can be achieved through techniques like interpretable machine learning and explainable AI.

5. Foster accountability and responsibility by a well-thought and well-structured risk assessment and evaluation plan to mitigate potential risks and harms. This can include the use of techniques like interpretable machine learning, explainable AI, and audit trails.

6. Embed continuous monitoring and evaluation pathways through the use of techniques like impact assessments, which can be used to identify potential risks and harms associated with AI systems. This is why an 'Ethics' Effectiveness Impact Assessment Framework' is of primordial importance in healthcare contexts where the ultimate purpose of use is worthwhile – the preservation of human health and well-being.

7. Finally, trustworthy AI requires a transdisciplinary approach with the involvement and collaboration of various stakeholders who may be affected by the AI system, including AI developers, policymakers, users, healthy individuals and patients as well as ethicists and legal professionals. This can be achieved through the establishment of multi-stakeholder forums and the ongoing engagement of stakeholders in the development and deployment of AI systems.

By following these steps, the principles of applied ethics can be combined with the development of trustworthy AI, thus ensuring that AI systems used in healthcare contexts are designed and deployed in a manner that is consistent with moral values and principles.

Bioethics

Since the dawn of medicine, its practice and that of biomedical research was governed by a set of human-centered, ethics principles.

One of the first practitioners who explicitly enshrined the notion of medical ethics in a set of four principles of medical ethics was Hippocrates around 400 BC. He was a physician and teacher of the ancient Greek classical period and is known as the 'Father of Medicine' and he probably reflected the practices and principles customary in the practice of medicine. The principles are the following [17]:

- Autonomy – respect for the patient's right to self-determination
- Beneficence – the duty to 'do good'
- Non-Maleficence – the duty to 'not do Harm'
- Justice – to treat all people equally and equitably

The respect or non-respect of these principles during medical practice entails consequences, which are subject to legal provisions under the Law e.g. even into criminal law in case of severe infringements. However, *"ethics drives our behaviour, not the law; in contrast, hopefully, the law largely reflects ethics"* [18].

With the rise of scientific medicine and research in the medical field, and largely as a result of appalling scandals originated by the publicity resulting from a series of severe infringements of the medical deontological practices, the notion of bioethics was introduced in the 1970s and the field of bioethics became prominent in many discussion and decision-making, leading to a series of principles and declarations at EU and global level.

In Article 4, the UNESCO's "Universal Declaration on Bioethics and Human Rights" foresees that "In applying and advancing scientific knowledge, medical practice and associated technologies, direct and indirect benefits to patients, research participants and other affected individuals should be maximized and any possible harm to such individuals should be minimized" [19].

Furthermore, the "World Medical Association's Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects", under "Risks, Burdens and Benefits" foresee that "17. All medical research involving human subjects must be preceded by careful assessment of predictable risks and burdens to the individuals and groups involved in the research in comparison with foreseeable benefits to them and to other individuals or groups affected by the condition under investigation. Measures to minimise the risks

must be implemented. The risks must be continuously monitored, assessed and documented by the researcher. 18. Physicians may not be involved in a research study involving human subjects unless they are confident that the risks have been adequately assessed and can be satisfactorily managed. When the risks are found to outweigh the potential benefits or when there is conclusive proof of definitive outcomes, physicians must assess whether to continue, modify or immediately stop the study” [20].

Key ethical concerns in bioethics often involve big questions such as:

- What should I do? How should I act?
- How should I treat others? What are my obligations or responsibilities toward others?
- What type of person should I be? What does it mean to be a good doctor or a good nurse or a good bench scientist?
- Big moral considerations in bioethics often revolve around questions about:
- Whether one ought to act to maximize the best outcomes or ought to act to uphold important moral rules and duties? Or how to do both?
- Are we required only not to harm others or must we also act in ways that benefit them or make their lives better?
- What should be done when we think policies or law are unethical because they don't treat people fairly or equally? What does it mean to treat people fairly?
- How could we design access to a scarce resource such that all people have a fair or maybe an equal opportunity to obtain that scarce resource, e.g., organ allocation policies?
- How and when should we share information about a medical treatment to best permit others make informed and voluntary decisions about what is done or not done to their bodies? What resources are needed to support people in making these decisions?
- When can minors make their own health care decisions? Who should decide if a minor child's opinions about a medical treatment for them differs from that of his/her parent(s) [21]?

Concurrent Design

The methodology of concurrent engineering or concurrent design encompasses all the processes where specialists from different disciplines work together in parallel and concurrently, towards an initially identified outcome, instead of working consecutively.

In an article that provides practical guidelines on how to merge different research methodologies using both quantitative and qualitative inquiries in biology education research, the author states that, as part of the concurrent design methodology the data for the qualitative and quantitative enquiries are collected in a single phase. Because the general aim of the concurrent design approach is to better understand or obtain more developed understanding of the phenomenon under study, the data can be collected from the same participants or similar target populations. The goal being to obtain different but complementary data that validate the overall results [22].

The principles of concurrent design can be applied during the conception, design and deployment of algorithmic systems for uses in healthcare contexts where developers, medical practitioners and users would in parallel contribute both qualitative and quantitative data to ensure robust, trustworthy and ethically designed systems. Such a working methodology would result in important economies of scale as systems would be tested and improved very rapidly due to a holistic approach as well as reduction of time-to-market as the trust of end users would be a feature embedded already in the design phase.

Algorithmic Ethics Effectiveness, Efficacy and Safety Assessment in healthcare contexts

As aforementioned, one of the scientific fields the current analysis draws inspiration from is also the process leading to the development of a new pharmaceutical product's measurement of its, efficacy, effectiveness and safety.

A drug (or any medical treatment) should be used only when it will benefit a patient. Benefit takes into account both the drug's ability to produce the desired result (efficacy) and the type and likelihood of adverse effects (safety). Cost is commonly also balanced with benefit. Efficacy is the capacity to produce an effect (eg, lower blood pressure) and it can be assessed accurately only in ideal

conditions (ie, when patients are selected by proper criteria and strictly adhere to the dosing schedule). Thus, efficacy is measured under expert supervision in a group of patients most likely to have a response to a drug, such as in a controlled clinical trial. Effectiveness differs from efficacy in that it takes into account how well a drug works in real-world use. Often, a drug that is efficacious in clinical trials is not very effective in actual use. For example, a drug may have high efficacy in lowering blood pressure but may have low effectiveness because it causes so many adverse effects that patients stop taking it. Effectiveness also may be lower than efficacy if clinicians inadvertently prescribe the drug inappropriately (eg, giving a fibrinolytic drug to a patient thought to have an ischemic stroke, but who had an unrecognized cerebral hemorrhage on CT scan). Thus, effectiveness tends to be lower than efficacy [23].

The need to ensure and monitor the safety of pharmaceutical products led to the introduction of the notion of pharmacovigilance. Pharmacovigilance is derived from the combination of the Greek word 'Φάρμακο' which means medicina product and the Latin word 'Vigilia' which means 'to keep watch'. It is aimed at monitoring the risk/benefit ratio of medicinal products, to preserve a patient's safety and the quality of life through the science of detection, assessment, understanding, and prevention of adverse effects of drugs or other related problems. The importance of pharmacovigilance was first highlighted in 1848, when a girl named Hannah Greener from England passed away after being administered chloroform for anesthesia to remove an infected toenail. Due to concerns around the safety of using anesthetics, the Lancet set up a commission to tackle this issue, encouraging doctors to report deaths caused by anesthesia. The need for safety monitoring has evolved around unfortunate incidents in history, with deaths caused by anesthesia and congenital malformations from thalidomide use. Reports from adverse drug reactions (ADRs) are stored in a global database and can be used to evaluate the associations between various medications and associated ADRs. Clinicians play an important role in the recognition and reporting of ADRs to national pharmacovigilance centers (NPCs). The purpose of NPCs is to make the clinicians understand their functions, including the monitoring, investigation, and assessment of ADR reports, along with periodical benefit-risk assessments of medications via multiple sources [24].

When an algorithmic system is used in a healthcare context to assist a clinician with e.g. the diagnosis of a health condition or with identifying the most suitable treatment for a specific individual's health treatment or a researcher with discovering a novel very beneficial medicinal product, this system should also be subject to the processes ensuring its safety, efficacy and effectiveness.

An algorithm is usually tested and measured against its efficiency. Algorithmic efficiency relates to how many resources a computer needs to expend to process an algorithm. It's a measure of how well an algorithm performs in terms of time and space, which are the two main measures of efficiency. Time complexity refers to the computational complexity that describes the amount of time an algorithm takes to run as a function of the size of the input to the program. Space complexity, on the other hand, refers to the amount of memory an algorithm uses to process the input. Efficiency is crucial because it directly impacts the performance of the system running the algorithm. The efficiency of an algorithm needs to be determined to ensure it can perform without the risk of crashes or severe delays. If an algorithm is not efficient, it is unlikely to be fit for its purpose. An algorithm's efficiency is measured by how many resources are used to process it. An efficient algorithm uses minimal resources to perform its functions. An inefficient algorithm can lead to longer execution times, higher costs, and potentially frustrated users if the algorithm is part of a user-facing application. Algorithmic efficiency can be measured using techniques like Big O notation, which provides an upper bound on the time complexity in the worst-case scenario. This notation helps to compare different algorithms based on their maximum running time. However, it's important to note that the efficiency of an algorithm can also depend on factors such as the specific data it's processing. For example, some sorting algorithms perform poorly on data that is already sorted or sorted in reverse order. In practice, the choice of the most efficient algorithm often depends on the specific requirements of the task at hand, including factors like the available computational resources, the size and nature of the input data, and the required accuracy or reliability of the results [22,25].

As a human being is partly characterized by their DNA and biomarkers, similarly an algorithm is defined by its design and its data input. For the purposes of this analysis, an algorithm used within a health & care context is considered as a 'living' entity that is defined initially by its core design and is subsequently affected by its 'environment'. In both the core design phase and the 'surrounding' environments the algorithm is going to 'operate', a set of principles, translated eventually into norms, must encompass its 'existence' in order to guarantee an optimised algorithmic ethics efficacy, effectiveness and safety avoiding biases, discrimination and ensuring the best outcome, while simultaneously augmenting the medical personnel's decision-making capacity and increasing the accuracy of results (e.g. diagnosis, treatment, monitoring, etc). The ultimate goal and purpose of its use being the preservation or improvement of healthy individuals or patients' conditions.

Due to the particular features of healthcare contexts, an algorithmic systems efficiency should be enhanced and measured against also its ethics effectiveness and safety, similar to medicinal products effectiveness and safety, to ensure the highest degree of accuracy and reliability in real world settings and eliminate the possibility of bias which can lead to unfair or inaccurate results.

To measure the 'algorithm's ethics effectiveness', analogously to the effectiveness of a medicinal product, and different from an 'algorithm's efficiency', this paper presents an 'Algorithmic Ethics' Effectiveness Assessment Framework' methodology with a focus on the competencies of the individual that develops the algorithmic system. The proposed framework comprises a set of principles that can be eventually translated to a competencies' checklist in order to evaluate an algorithm's ethics' performance within a given time and space when used in a health & care context.

Approaches towards Mitigating Errors and Biases

As all human beings, no-one is infallible. However, the introduction of so powerful technologies is done with the purpose of reducing and almost eliminating errors e.g. diagnosis, treatment, cure, etc. So, the question arises :- what happens and who's to blame when algorithms go wrong? A more detailed analysis of the purely legal challenges will be presented in a subsequent paper.

What could be the best approach for mitigating the risks of bias and legal repercussions stemming from the use of AI technologies in health & care settings?

"Before computer scientists can even start theorizing about how to build such "novelty-adaptive" agents, they need a rigorous method for evaluating them. Traditionally, most AI systems are tested by the same people who build them. Competitions are more impartial, but to date, no competition has evaluated AI systems in situations so unexpected that not even the system designers could have foreseen them. Such an evaluation is the gold standard for testing AI on novelty, similar to randomized controlled trials for evaluating drugs [26]".

A few points could be drawn from the above extract that are important when AI systems are designed for and deployed in a health and care context: the capacity of AI systems to adapt to unexpected and novel circumstances; the design of rigorous, robust yet adaptable evaluation processes prior to building such systems.

A recent study examined how well a machine learning model performed across several independent clinical trials of antipsychotic medication for schizophrenia. Models predicted patient outcomes with high accuracy within the trial in which the model was developed but performed no better than chance when applied out-of-sample. Pooling data across trials to predict outcomes did not improve predictions. These results suggest that models predicting treatment outcomes in schizophrenia are highly context-dependent and may have limited generalizability [27].

As the degree of extraordinary and novel circumstances during experimentation, diagnosis to discovery and treatment is very high in the health & care context, expanding on these aspects the analysis will explore what could be a rigorous, robust yet adaptable evaluation process or methodology for AI systems from an ethics point of view to enable and support efficient and effective performance of such systems under ordinary but also extraordinary and novel circumstances, in other words, the algorithm's ethics effectiveness embedded from its conception and design phase and in principles that its designer and developer will need to adhere to.

The public's trust depends in large part on fitting the expected results to the resources put into developing and running the products. In other words, the result must be commensurable with the resources and sacrifices invested in its development and implementation. The public is generally favourable to innovation, while it trusts the institutions that develop, products and regulate them. But this is based on trust, and trust only. Scientists are among the most trusted professionals in the world and preserving this trust rests on applying the existing legislation and regulation in a fair and equitable manner rather than on drafting new rules that will only burden the administration while having no effect on the actual climate of trust.

To do so, this trust must be enhanced. And can and should be extended to systematically impact/affect? AI use, particularly as health is an area that has a rich tradition of ethics and understanding of the ethics permeating the development process of health innovation.

A tool that has proved its worth in several areas of medical innovation is a-priori ethics review, and a pre-emptive response to ethics concerns. This form of ethics review has been applied in European funded research and has proven to be satisfactory to the researchers and to the public. It has a proven track record in medical research, development of medical devices and acceptance of the introduction of IT into the Health policies. The confidence of the public is a non-negligible part of the acceptance of a technology or product, and this, in turn, makes for a significant economy of resources and effort. The example of Clinical Trials, for example, must be analysed carefully, and proper extrapolations apart, lessons learnt weighed in. Furthermore, similarly to drug development, but to a much larger scale, in the case of AI applications, the very testing of the code is so resource consuming, that it is desirable to carry out an ethics evaluation as early as possible in the process, as the actual Ethical issues are, in fact, the issues of Quality and Standards. Making sure that ethics review is carried out, implemented and later, documented/certified is a sure way to provide proof that the AI application has been screened, assessed and monitored for its fairness, equity and transparency.

Good software, like all other tools, must conform to standards, not only for quality (to be evaluated for trustworthiness and reproducibility, accuracy of its data and transparency) but also the source of any code components. This is not just an issue of "attribution" or "ownership", but particularly, it is a record of its original aim and how it was modified along the way.

This facilitates understanding, corrections, and improvements. Careful version control is an essential part of high-quality programming, and this fact underlines the need for transparency, veracity, and accountability. This naturally, underlines the connection of quality, standards, and ethics.

Predict, Prevent and Eliminate Bias

The very nature of machine learning algorithms makes it plain that, however unintentionally, one could develop a biased system or accept biased results. Though intent matters in ethics – and many actions are assessed based on intent – and that failure to take note, correct or mitigate a problem can also be cause for blame, and failure to do so is a negligence.

The same way a surgeon won't deliberately operate on the wrong limb or remove the healthy kidney, it is sensible to scrutinize the datasets used for training machine learning algorithms for embedded bias or lead to biased interpretations, even if doing so is a difficult and labour-intensive approach to take.

Another point is to proceed with the careful screening of output to identify and filter possible bias, and perhaps to identify ways to modify the algorithm to suppress this bias. It is now even possible to incorporate anti-bias features into the code itself.

This is a promising approach, and implementing an ethics review process it might very well be an effective way to ensure that public health surveillance and prediction, disease diagnosis and treatment, and health policy are not corrupted by bias.

AI for Good and not Evil

Although attitudes may vary overtime, the fact remains that the trends to evaluate good and bad remain. Reducing suffering, eliminating disparities, and improving health are “Good”. Depriving people of rights, using people for political or economic purposes without permission given voluntarily, and harming people for profit are “Evil”. The use of powerful technologies such as the AI technologies should augment human capacity to ‘Do Good’. And the healthcare field is the ideal ground to let AI technologies demonstrate our capacity to ‘Do Good’ - and not simply ‘Do No Harm’ - to the maximum.

But is it possible for an algorithmic system to perform ‘Good’ or ‘Evil’ tasks? Very unlikely so and as mentioned before, algorithmic systems are as good as those who conceive, design, develop, test and deploy them.

In a November 2023 article that appeared in ‘Medium’, the author elaborates on the notions of good, bad and evil engineers. She states, among others that *“A good engineer possesses 3 qualities: exceptional knowledge, commitment to truth, and commitment to result. A bad engineer lacks either exceptional knowledge or commitment to results. However, they do have a medium level of commitment to truth. An evil engineer has no or little commitment to truth. The result is of no importance to them. They care about other aspects (perhaps the appearance of results), or they don’t care about anything at all. It’s rare for an evil engineer to have exceptional knowledge, but if they do, it’s not relevant anyway, as again, they care neither for the truth, nor the result. Some of you may find that there is not a clear distinction between the bad and evil engineers here. Normally, evil often does harm — so you would expect an evil engineer to introduce malicious code with bad intentions, or to cover their past mistakes. I agree with that. Yet, what I’d like to highlight here is where I draw the line between bad and evil: It doesn’t necessarily require a malicious action for the engineer to be evil, once the engineer starts ignoring the truth in front of their eyes (i.e. pretending not to see the problems), they cross into the realm of evil. And the more facts they ignore, the more evil they will become”* [28].

From the above it is thus evident that not only the algorithmic systems themselves needs to be provided guardrails but also their developers so that both can ‘Do Good’.

Provide Robust Evaluation

To what basic queries should the ethics assessment provide answers :

- Are data reliable and who is responsible for ensuring reliability?
- How exactly does the calculation work?
- How should it be determined who should use these calculators, and for what purposes?

These, of course, parallel the “lessons learned” we earlier identified for appropriate use of machine learning or any other medical software.

Ethics, Standards, and Public Policy

Once we understand how to get something right, it would be irrational to insist it should be ignored. The evolution of standards in health care has improved quality, increased safety, and saved resources. This is also true for health informatics. If continuously refined and improved the standards achieve results. If those achievements improve human health and welfare, then there is an ethical imperative to develop and improve them. Providing the public with a visible sign that ethics is embedded in the fabric of the code that provides the health care can only enhance the virtuous cycle of trust. Without trust, the best software will fail.

Concurrent Design of an Algorithmic Ethics Effectiveness Impact Assessment, Competency-Based Framework

Article 4 Article 4 of the Convention on Human Rights and Biomedicine states : *“Any intervention in the health field, including research, must be carried out in accordance with relevant professional obligations and standards”*[29].

In a report issued in June 2022 by the Council of Europe’s Steering Committee for Human Rights in the fields of Biomedicine and Health (CDBIO) on the impact of artificial intelligence on the doctor-patient relationship it states that it remains unclear whether developers, manufacturers, and service

providers for AI systems will be bound by the same professional standards as the ones referred to in Article 4 of the Oviedo Convention care and that careful consideration must be given to the role played by healthcare professions bound by professional standards when incorporating AI systems that interact directly with patients [30].

As in the case of bioethics and medical research, medical practitioners and researchers are being trained for and must uphold certain values, abide to principles and demonstrate due care when interacting with individuals or when performing research to maximise the benefits and weigh the risks and burdens on the individuals, in the same way, the developers of algorithmic systems for uses in healthcare contexts, must also be required to provide proof that they have also been trained for and are in a position to design and deploy systems that are beneficial to the individuals they are destined for.

Although, the existing texts refer to the 'Do No Harm' principle, the use of AI technologies, renders this principle in a way obsolete. The use of the powerful analytic capacity of AI technologies should enable medical practitioners, researchers and other users of such technologies in the biomedical field to augment their skills and quality of outcome towards a 'Do Good' principle since the target should be to exceed the state-of-the-art, this being the 'Do No Harm' principle.

Designers and developers of AI systems used in a healthcare context should target to produce not only extremely efficient systems but also systems that are effective when used at wider scale, outside of a controlled environment, which have a 'Do Good' impact for the individuals (healthy and requiring medical assistance) and which can be easily audited, calibrated, verified and validated. Therefore, designers and developers should be subject to liability rules for the systems they develop in the same way medical professionals are held liable in case of medical error. However, due to the very specialised skillset required to produce beneficial outcomes in a healthcare context, the legal notion of joint and several liability could be foreseen as a fair and equitable solution to distribute the liability between those who design and develop the systems, those who distribute and deploy the systems and those who use the systems.

Drawing from the above, and in order to ensure the safety, efficacy and effectiveness of algorithmic systems used in healthcare contexts, systems' designers and developers, systems vendors, systems users and receivers of the services and benefits offered by these systems, should sit together and discuss on how to best design, develop and deploy such systems. (to be calibrated) How to best do that? Design a competency-based framework that ensures robust AI system from its conception and design stage.

Competency-Based Certification Framework for AI Systems' Developers in Healthcare Contexts

"Accordingly, the development of procedures to assess whether an AI will perform as expected is vital. Since Machine Learning will drive AI for the foreseeable future, humans will remain unaware of what an AI is learning and how it knows what it has learned. While this may be disconcerting, it should: human learning is similarly opaque. [...] To cope with this opacity, societies have developed myriad professional certification programs, regulations, and laws. Similar techniques should be applied for AIs; for example, societies could permit AI to be employed only after its creators demonstrate its reliability through testing processes. Developing professional certification, compliance monitoring and oversight programs for AI – and the auditing expertise their execution will require – will be a crucial societal project"[31].

The remarkable expansion of AI has generated a pressing need for a standardized certification process that can effectively evaluate the competencies of individuals engaged in this burgeoning field. The applications of AI span diverse sectors such as healthcare, finance, transportation, and manufacturing. As AI continues to advance, the demand for proficient AI professionals is concurrently on the rise.

This paper introduces a comprehensive competency-based certification framework tailored for AI professionals, strategically aligned with prevailing industry standards. The framework is meticulously crafted to be dynamic and adaptable, aligning with the ever-evolving landscape of AI.

Furthermore, it explicitly expresses its openness to collaboration with industry stakeholders, ensuring continuous relevance and currency.

The proposed certification framework delineates three distinct levels based on proven track record:

- Entry-level: This level signifies that an individual possesses a foundational understanding of AI principles and can proficiently apply them to solve uncomplicated problems.
- Intermediate-level: Certification at this tier attests to an individual's advanced grasp of AI principles and the ability to apply them to address more intricate challenges.
- Expert-level: This pinnacle certification validates that an individual has a profound understanding of AI principles and can adeptly apply them to solve the most complex problems.

Skills Required for AI Professionals

Beyond the certification levels, the framework identifies some critical skill categories for AI professionals, including:

1. Regulatory compliance: Mastery of understanding and navigating regulatory frameworks is imperative to ensure adherence to pertinent laws and regulations.

2. Ethical use and bias removal: The capacity to identify and eliminate bias in AI systems is crucial for fostering fair and unbiased assessments.

3. Validation and testing: Proficiency in rigorously testing AI systems is vital to guarantee their intended functionality and error-free operation.

4. Continuous monitoring and feedback: The ability to monitor AI systems' performance and collect feedback for ongoing enhancement is essential for maintaining their effectiveness and relevance.

5. Deployment and scalability: Competence in preparing for AI model deployment, evaluating infrastructure requirements, and seamless integration into existing systems are indispensable for successful implementation.

6. Risk assessment and mitigation: The capability to conduct comprehensive risk assessments is necessary for identifying potential risks associated with AI systems and implementing suitable measures for mitigation.

7. Security and Privacy: Although this section could fall under other items of this list, in the era of AI, the defense of an individual's privacy is an activity one should highly take into consideration. Therefore, skills related to standard process to ensure anonymity of the datasets used in the model training are important. The more complex the collected data become, the more advanced and deep the techniques have to be to ensure that reverse engineering will not permit the identification of individuals. At the same time, these techniques must guarantee that the anonymized data maintain the intrinsic knowledge (i.e. patterns) to enable the models training.

8. End-of-life processes and ethical implications: Understanding the future functionality, sustainability, end-of-life processes, ethical implications, and societal impact of AI systems is crucial for responsible AI development.

9. Trans-disciplinary Team Collaboration: Experience in working with a multi-disciplinary team is necessary for ensuring a comprehensive assessment and mitigation of potential harms associated with healthcare AI systems. This involves collaborating with experts from various fields, such as data scientists, healthcare providers, and legal experts, to ensure that the AI system is designed and implemented effectively and in compliance with the relevant regulations. Therefore, the assessment of the individual's soft skills is important in order to be able to convey correct information to the collaborators.

10. Domain and Sector Knowledge: Knowledge of the specific healthcare domain and sector in which the AI system is being deployed, as well as the unique challenges and requirements of that domain. This includes understanding the healthcare-specific context, identifying potential challenges, and tailoring the AI system to meet the unique needs of the healthcare industry.

5. Conclusions

Taking the example of the profession of a data scientist, data science certifications have become increasingly important for job seekers looking to enter this industry. A data science certification is a credential that validates an individual's proficiency in utilizing data analysis techniques, interpreting data, and deriving insights from it. It demonstrates the expertise required to work with data science tools and technologies [32].

The benefits of a certification can differ depending on the industry and position you're pursuing. Here are some common benefits you might consider when determining if a certification is useful to a person's career:

- Emphasise professionalism: Having a certification in the industry or position someone is pursuing can emphasise a person's professionalism and commitment. This can help to establish credibility within the person's professional network, which can be beneficial for future endeavours.
- Improve a person's resume: When pursuing a job search and applying for positions, having a certification on a person's resume can improve it significantly. It might show a recruiting manager the person's enthusiasm and commitment to fulfilling the responsibilities within the position to a high standard.
- Work more efficiently: The education required when studying to achieve a certification is likely to provide a person with theoretical and practical knowledge to work more efficiently. The course work when completing a certification, can teach a person strategies and tools to produce high-quality products and services within the industry.
- Enhance the person's earning potential: A professional with a certification might enhance their earning potential because they have specialised knowledge and training. This allows a professional with a certification to pursue higher-level positions [33].

A plethora of analysis has been produced up to now targeting the algorithmic system itself. Although it is an absolute necessity so as to be able to evaluate the system's ethics impact effectiveness, it is equally necessary to embed in the process a set of professional certification standards targeting ethics for the designers and developers of such systems both at individual and corporate level. As in the examples of existing certifications and standards in many professions where the individuals and corporates need to certify their technical skills, they should be required to certify that they have adopted an inclusive approach, ethics-informed and human-centered designed, right from the conception phase and that they understand the requirements set by regulatory, legislative and ethics principles and texts and they have ensured that these are incorporated in their methodologies and review processes of the systems they develop.

Conflicts of Interest: The authors declare no conflicts of interest.

Authors' declaration on the use of Generative AI: The text in lines 140-160 and the sub-chapter on 'Applied Ethics' were produced with the assistance of Generative AI. The remaining part of the analysis was drafted by the authors on the basis of own ideas and text.

References

1. Bob Kocher, Zeke Emanuel. *BROOKINGS*. [Online] USC-Brookings Schaeffer Initiative for Health Policy, March 05, 2019. [Cited: May 05, 2024.] <https://www.brookings.edu/articles/will-robots-replace-doctors/>.
2. Artificial intelligence in medical device software and high-risk medical devices - a review of definitions, expert recommendations and regulatory initiatives. Fraser AG, Biasin E, Bijns B, Bruining N, Caiani EG, Cobbaert K, Davies RH, Gilbert SH, Hovestadt L, Kamenjasevic E, Kwade Z, McGauran G, O'Connor G, Vasey B, Rademakers FE. 6, 2023, Expert Review of Medical Devices, Vol. 20, pp. 467-491.
3. *A Review of the Role of Artificial Intelligence in Healthcare*. Al Kuwaiti, Ahmed et al. 6, 2023, Journal of Personalized Medicine, Vol. 13, p. 951.
4. OECD. Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI. Paris : OECD Digital Economy Papers, No. 349, OECD Publishing, 2023.
5. US Food & Drug Administration (FDA). Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. [Online] December 06, 2023. [Cited: February 25, 2024.] <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>.

6. Global Approach to Software as a Medical Device. [Online] September 27, 2022. [Cited: February 25, 2024.] <https://www.fda.gov/medical-devices/software-medical-device-samd/global-approach-software-medical-device>.
7. Artificial Intelligence and Machine Learning in Software as a Medical Device. [Online] September 22, 2021. [Cited: February 25, 2024.] <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>.
8. Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan. [Online] September 22, 2021. [Cited: February 25, 2024.] <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>.
9. US FDA. Artificial Intelligence and Machine Learning in Software as a Medical Device. *US FDA Digital Center of Excellence*. [Online] March 15, 2024. [Cited: May 05, 2024.] <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>.
10. European Commission. Liability Rules for Artificial Intelligence. [Online] [Cited: May 03, 2025.] https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en.
11. Shaping Europe's Digital Future. [Online] March 06, 2024. [Cited: April 05, 2024.] <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.
12. European Medicines Agency. Medical devices. *Human regulatory: overview*. [Online] [Cited: May 03, 2024.] <https://www.ema.europa.eu/en/human-regulatory-overview/medical-devices#medical-device-legislation-13035>.
13. European Medicines Agency (EMA). Artificial intelligence workplan to guide use of AI in medicines regulation. [Online] December 2023. [Cited: May 05, 2024.] <https://www.ema.europa.eu/en/news/artificial-intelligence-workplan-guide-use-ai-medicines-regulation>.
14. NIST. AI Risk Management Framework. *NIST Information Technology Laboratory*. [Online] [Cited: May 03, 2024.] <https://www.nist.gov/itl/ai-risk-management-framework>.
15. AI Risk Management Framework. *Information Technology Laboratory*. [Online] [Cited: May 03, 2024.] <https://airc.nist.gov/docs/NIST.AI.600-1.GenAI-Profile.ipd.pdf>.
16. An Ethics Impact Assessment (EIA) for AI uses in Health & Care: The correlation of ethics and legal aspects when AI systems are used in health & care contexts. Elsa Papadopoulou, Themis Exarchos. Corfu : ACM Digital Library, 2022. <https://doi.org/10.1145/3549737.3549755>.
17. Medical Protection. The four principles of medical ethics. *Medical Protection*. [Online] February 15, 2024. [Cited: May 05, 2024.] <https://www.medicalprotection.org/uk/articles/essential-learning-law-and-ethics>.
18. World Medical Association (WMA). WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects . *World Medical Association (WMA)*. [Online] September 06, 2022. [Cited: May 05, 2024.] <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>.
19. UNESCO. Universal Declaration on Bioethics and Human Rights. *UNESCO Legal Affairs*. [Online] October 19, 2005. [Cited: May 05, 2024.] <https://www.unesco.org/en/legal-affairs/universal-declaration-bioethics-and-human-rights?hub=66535>.
20. Kennedy Institute of Ethics at Georgetown. What is Bioethics? *Bioethics Research Showcase*. [Online] 2015. [Cited: May 05, 2024.] <https://bioethicsarchive.georgetown.edu/bioethicsshowcase2015/what-is-bioethics/index.html>.
21. *Mixed-Methods Design in Biology Education Research: Approach and Uses*. AM, Warfa. s.l. : CBE Life Sci Educ, 2016, Vols. 15,4 (2016): rm5. <https://doi.org/10.1187/cbe.16-01-0022>.
22. Davies, Nahla. How To Calculate Algorithm Efficiency. *KDNuggets*. [Online] September 20, 2022. [Cited: May 05, 2024.] <https://www.kdnuggets.com/2022/09/calculate-algorithm-efficiency.html>.
23. Lynch, Shalini S. Drug Efficacy and Safety. *MERCK Manual* . [Online] May 2022. [Cited: May 05, 2024.] <https://www.merckmanuals.com/professional/clinical-pharmacology/concepts-in-pharmacotherapy/drug-efficacy-and-safety>.
24. *Pharmacovigilance and Its Importance for Primary Health Care Professionals*. Hamid, A. A. A., Rahim, R., & Teo, S. P. Sept. 2022, s.l.: Korean journal of family medicine, 2022, Vols. 43(5), 290–295. <https://doi.org/10.4082/kjfm.21.0193>.
25. Stephen M. Walker II, Co-Founder / CEO. What is algorithmic efficiency? *KLU*. [Online] [Cited: May 05, 2024.] <https://klu.ai/glossary/algorithmic-efficiency>.
26. Mayank Kejriwal. Embrace the unexpected: To teach AI how to handle new situations, change the rules of the game. *The Conversation*. [Online] April 07, 2021. [Cited: May 05, 2024.] <https://news.yahoo.com/embrace-unexpected-teach-ai-handle-122915705.html>.
27. *Illusory generalizability of clinical prediction models*. Chekroud, Adam M et al. 383, New York : Science, 2024, Vol. 6679. doi:10.1126/science.adg8538.

28. Good engineers, bad engineers, and evil engineers — an anecdote for data leaders. *Medium*. [Online] November 29, 2023. [Cited: May 05, 2024.] <https://towardsdatascience.com/good-engineers-bad-engineers-and-evil-engineers-an-anecdote-for-data-leaders-6cceb82c8c3e>.
29. Oviedo Convention and its Protocols. *Council of Europe, Human Rights and Biomedicine*. [Online] December 01, 1999. [Cited: 05 05, 2024.] <https://www.coe.int/en/web/bioethics/oviedo-convention>.
30. Mittelstadt, Brent. The impact of artificial intelligence on the doctor-patient relationship. *Council of Europe, Human Right and Biomedicine, Artificial Intelligence*. [Online] December 2021. [Cited: May 05, 2024.] <https://www.coe.int/en/web/bioethics/report-impact-of-ai-on-the-doctor-patient-relationship>.
31. H. Kissinger, E. Schmidt and D. Huttenlocher. *The Age of AI*. London : John Murray, 2022.
32. Teepa, Vaishali. Data Science Certifications: Why Use Them for Professional Development? *LinkedIn*. [Online] April 19, 2024. [Cited: May 05, 2024.] <https://www.linkedin.com/pulse/data-science-certifications-why-use-them-professional-vaishali-teepa-yybfc/>.
33. Indeed Editorial Team. Accreditation vs Certification (Including Their Benefits). *Indeed Career Guide*. [Online] April 26, 2024. [Cited: May 05, 2024.] <https://au.indeed.com/career-advice/finding-a-job/accreditation-vs-certification>.
34. US FDA. Artificial Intelligence and Machine Learning in Software as a Medical Device. [Online] September 22, 2021. [Cited: February 05, 2024.] <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>.
35. US FDA, Center for Devices and Radiological Health's Digital Health Center of Excellence. Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan. [Online] US FDA, January 2021. [Cited: February 05, 2024.] <https://www.fda.gov/media/145022/download>.
36. Janneke van Oirschot, Gaby Ooms. *Interpreting the EU Artificial Intelligence Act for the Health Sector*. Amsterdam : Health Action International, 2022.
37. Exploring Human Values in the Design of a Web-Based QoL-Instrument for People with Mental Health Problems: A Value Sensitive Design Approach. Maathuis, Ivo, et al. 2020, *Science and Engineering Ethics*, pp. 871-898.