AI and Ethics
https://doi.org/10.1007/s43681-023-00331-3

ORIGINAL RESEARCH

Operationalising AI ethics through the agile software development lifecycle: a case study of AI-enabled mobile health applications

Lameck Mbangula Amugongo[1] · **Alexander Kriebitz**[1,2] · **Auxane Boch**[1] · **Christoph Lütge**[1,2]

Received: 31 May 2023 / Accepted: 2 August 2023

Abstract
Although numerous ethical principles and guidelines have been proposed to guide the development of artificial intelligence (AI) systems, it has proven difficult to translate these principles into actionable practices beyond mere adherence to ethical ideas. This is particularly challenging in the context of AI systems for healthcare, which requires balancing the potential benefits of the solution against the risks to patients and the wider community, including minorities and underserved populations. To address this challenge, we propose a shift from one-size-fits-all ethical principles to contextualized case-based ethical frameworks. This study uses an AI-enabled mHealth application as a case study. Our framework is built on existing ethical guidelines and principles, including the AI4People framework, the EU High-Level Expert Group on trustworthy AI, and wider human rights considerations. Additionally, we incorporate relational perspectives to address human value concerns and moral tensions between individual rights and public health. Our approach is based on "ethics by design," where ethical principles are integrated throughout the entire AI development pipeline, ensuring that ethical considerations are not an afterthought but implemented from the beginning. For our case study, we identified 7 ethical principles: fairness, agility, precision, safeguarding humanity, respect for others, trust and accountability, and robustness and reproducibility. We believe that the best way to mitigate and address ethical consequences is by implementing ethical principles in the software development processes that developers commonly use. Finally, we provide examples of how our case-based framework can be applied in practice, using examples of AI-driven mobile applications in healthcare.

Keywords Ethics by design · Ethics AI · AI-enabled mobile applications · Software development lifecycle

1 Introduction

The availability of routine clinical, granular user-generated data combined with advanced computing makes artificial intelligence (AI) attractive in healthcare. According to the [1], there is an increase in the number of authorized AI-enabled medical devices. Similarly, AI-enabled mobile applications are attracting a lot of interest in healthcare. This is a result of significant improvement when it comes to the hardware resources on mobile devices. In healthcare, AI-powered mobile applications analyse large amounts of patient-generated data, leading to improvements in disease surveillance, early diagnosis and treatment management. Consequently, AI-enabled application can improve healthcare professionals' productivity, enhance clinician' decision-making capabilities, and potentially reduces healthcare costs. For example, in Low- or Middle-Income Countries (LMICs) with no or limited experts, telemedicine can improve access to quality healthcare allowing users to assess their physiological, psychological and behavioural data in real-time [2].

The growing application of machine learning (ML) techniques in healthcare has increased awareness of the ethical issues that arise in the design, deployment and use of AI systems. Ethical issues, such as privacy, accountability,

✉ Lameck Mbangula Amugongo
lameckmbangula.amugongo@tum.de

Alexander Kriebitz
a.kriebitz@tum.de

Auxane Boch
auxane.boch@tum.de

Christoph Lütge
uetge@tum.de

[1] Technical University of Munich, School of Social Sciences and Technology, Institute for Ethics in Artificial Intelligence, Marsstrasse 40, 80335 Munich, Germany

[2] Peter Löscher Chair of Business Ethics, Technical University Munich, Munich, Germany

Published online: 15 August 2023

transparency, fairness, robustness, safety, and trust, have been widely reported and discussed in the literature, and if not considered, ethical concerns can pose a threat to equitable health delivery and human rights [3].

There have been growing calls from institutions and researchers to apply ethics to address ethical concerns and harness the potential of AI technologies in a responsible way. For instance, Calo [4] proposed a roadmap to address major policy questions that arise when AI is applied, to aid policymakers, technologists and scholars understand the contemporary policy environment around AI. However, only a few ethical frameworks are specific to healthcare. Reddy et al. [5] proposed a governance model to tackle both the ethical and regulatory concerns that arise from the implementation of AI in healthcare. Similarly, a framework for ideal algorithms in healthcare was proposed, using a checklist to assess adherence to ethical principles [6]. In another study, a consensus guidance framework outlining six principles for the ethical use of AI in healthcare [7]. However, there remain a few practical recommendations that developers can use throughout the development lifecycle [8].

Critical questions have been raised about whether ethical principles can significantly influence the decision-making processes of humans working in the field of AI and ML because AI ethics lack the means to reinforce its own normative beliefs. Efforts are being made to bridge this gap. Hagendorff [8] proposed recommendations to translate AI ethics from mere discussions into concrete actions that developers and users of AI for healthcare can utilize throughout the AI development life cycle. Another study proposed a framework to operationalise ethics based on existing guidelines that provide actionable solutions [9]. The authors assert that they organized the framework by the AI development life cycle. However, the AI development life cycle mostly follows the agile software development life cycle (SDLC). Therefore, it would make more sense for developers to incorporate ethics in phases of the agile SDLC.

Furthermore, existing ethical principles and guidelines have several shortcomings. Firstly, they predominantly focus on theoretical adherence to ethical concepts without providing practical guidance for their implementation in the SDLC. Secondly, most frameworks and guidelines assume a one-size-fits-all approach, disregarding the unique characteristics of specific sectors or applications. This limitation is particularly important in the field of healthcare, where complex health conditions vary significantly from one ailment to another. As a result, ethical principles designed to address concerns related to the application of one healthcare domain may not be applicable to another. For instance, the ethical challenges associated with implementing AI in medical imaging for cancer diagnosis differ from those encountered when AI is utilized in the realm of mental health. Consequently, a universal ethical framework cannot adequately address all the ethical concerns arising from the use of AI in healthcare across all use cases.

To address these concerns, we identify ethical issues that arise when AI systems are applied to AI-enabled mobile health (mHealth) applications for healthcare. In this paper, we define AI-enabled mHealth applications as software applications that leverage AI techniques to provide information to users and other related services to patients via mobile platforms, such as smartphones, tablets and wearables (watches/Fitbit). These applications use ML techniques to understand and respond to user input, enabling them to provide personalized health recommendations, track health data, and offer remote monitoring and diagnosis. Because developers employ agile development methods to create mHealth applications. We believe that it is logical to integrate AI ethics into the SDLC for the effective implementation of ethical principles.

The principles we outlined are grounded in existing principles and relational theories. We emphasize the relational aspects because the integration of ethical principles in AI systems for healthcare is a complex and persistent issue, involving a balance between the potential benefits of the solution and risks to patients and the wider community, including minorities and vulnerable populations. Finally, we provide practical examples of how these ethical principles can be operationalized throughout the development lifecycle using AI-enabled mHealth application examples.

2 Review of AI-enabled mHealth applications

In this section, we examine the latest AI-based methodologies that have been employed for dietary assessment, patient monitoring, mental health management, and healthcare administration, with a specific focus on frameworks accessible on mobile devices.

2.1 Dietary assessment

Manual dietary assessment techniques such as 24-h dietary recall have been widely used in nutritional epidemiology studies to collect detailed information about participants' food intake to understand dietary behaviour and aid them in selecting healthier alternatives for their food consumption. However, this self-reporting technique heavily relies on users' subjective judgement for reporting food types and portion sizes, which can potentially introduce bias and inaccuracies in the dietary intake analyses. Consequently, semi-automated and automated visual-based approaches have been proposed in the literature. These methods have been explained in detail in several comprehensive reviews

on the state-of-the-art (SOTA) methodologies for food recognition, volume and calorific estimation [10, 11].

The advent of data privacy and protection laws such as the European Union's general data protection regulation (GDPR), combined with powerful mobile devices with capabilities to perform on-device inference makes mobile or edge computing ideal. In recent years, the number of vision-based methodologies accessible on mobile devices for food recognition and volume estimation proposed in the literature has been on the rise, as illustrated in recent reviews [12, 13]. Generally, the methods for food recognition are divided into two categories: manual hand-crafted and deep learning-based. Several studies have explored support vector machines (SVMs), using features, such as colour, texture, and local features, for classification. In their work, Luo, Ling and Ao [14] introduced, developed, and assessed an efficient food classification tool, leveraging mobile computing and predictive models to assist type2 diabetes (T2D) patients in making informed dietary choices. The tool incorporates a comprehensive food database, enabling convenient recording and tracking of patients' daily diets. They achieved a 93% accuracy in prescribing the best meal scenario, showing effectiveness in supporting T2D patients. In another study, handcrafted features like colour and texture were used to develop a novel approach using multi-segment SVM for food recognition [15]. They demonstrated the effectiveness of the model to work on smartphones.

Traditional manual handcrafted approaches encounter several challenges such as the tedious feature engineering process, which can become impractical when the database grows. Another challenge, manual approaches may not consider contextual information, such as ingredients and cooking methods, which are useful for accurate dish recognition and volume estimation. To overcome these challenges, deep learning (DL) techniques such as convolutional neural networks (CNNs) are increasingly being used. CNNs automatically extract learning features from data and have demonstrated good performance and adaptability in various food recognition and volume estimation tasks. Merchant and Pande [16] proposed a CNN-based algorithm for food recognition to help users fight obesity and help diabetes patients eat healthily. In their methodology, they applied transfer learning and fine-tuning techniques, achieving notably higher accuracy compared to other approaches that have used the Food-101 dataset. Another study applied a SOTA CNN using a multi-label predictor capable of learning recipes based on the ingredients list. They demonstrated the ability to predict the list of ingredients associated with a given picture, even when the corresponding recipe has not been previously seen by the model [17].

2.2 Patient monitoring

The widespread availability of mobile devices enables global accessibility to AI-powered medical applications for decision-making. These applications play a significant role in empowering patients whilst also enabling healthcare professionals to make more personalized and effective decisions, leading to positive outcomes for patients. Previous studies have demonstrated the benefits of real-time monitoring of patients using wearable devices [18–20]. Isakadze and Martin [19] investigated the ECG feature of the Apple Watch, the study suggests that the ECG feature shows promise in the detection of atrial fibrillation (AF) and may enable users to take an active role in their healthcare. In their study, Semaan et al. [20] showed that on average, individuals with AF engage in less daily physical activity and that deteriorating AF symptom severity is associated with reduced daily exercise. In another study, Bashar et al. [18] illustrated the feasibility of a wearable armband device and an algorithm that can be embedded in the device for real-time monitoring and detection of AF. The aforementioned studies focus on using wearable technology to monitor and detect cardiovascular disease. Other studies have shown promising benefits of wearables diabetes [21], health behaviour change intervention [20, 22] and cancer [23]. However, real-time monitoring can be a challenge in rural areas and countries without advanced network technologies and wireless communications. This necessitates the need for lightweight algorithms that will allow for on-device predictive analytics and effective data transmission.

2.3 Mental health management

Over the past years, AI has been applied in healthcare to aid patients with mental health. In the field of mental health, mHealth applications that incorporate AI have shown to be promising tools to support individuals dealing with anxiety and depression [24, 25]. AI for mental health, the majority of the mHealth applications are in the form of chatbots [26]. These chatbots are equipped with therapeutic techniques aimed at providing assistance and support to users. Such applications offer a non-judgmental and comfortable environment, mitigating the stigmatization surrounding seeking mental health advice [26]. In addition, AI-powered applications provide a cost-effective and scalable way to enable access to mental health support [27].

Several AI-enabled mHealth applications have been proposed in the literature, as outlined in a recent surveys [24, 25]. For instance, Liu et al. [28] and Burton et al. [29] found that using AI bots for self-help can lead to a reduction in depressive symptoms within a relatively short period. These positive findings motivate the development of AI beings for mental wellness support. Other existing chatbot-based

AI-enabled mHealth applications, such as Woebot [30], Tess [31], Wysa [32], and Ajivar [33], have shown promising results. For example, Woebot aims to support cognitive behavioural therapy (CBT) by providing users with tools and techniques to recognize and change behaviour patterns [30, 34]. The proposed chatbot offers tailored interventions to individuals experiencing depression or anxiety, resulting in significant improvements in patients' symptoms and mood regulations [34, 35]. Similarly, other AI assistants like Tess and Wysa incorporate therapeutic models, such as CBT and emotion-focussed therapy, along with mindfulness exercises [31, 32]. Ajivar, on the other hand, provides users with resilience and emotional intelligence training, real-time notifications, and gamification of positive habits through challenges, and mindfulness content. Its tailored support has been shown to enhance users' mental well-being as well [33]. These applications demonstrate the versatility and adaptability of AI agents in addressing various mental health needs, providing evidence-based interventions for users.

Although mHealth applications bring several benefits, ethical concerns persist. A major concern is safety, which necessitates the need for well-designed and closely monitored AI systems to prevent potential harm, particularly for users who may require specialized professional mental health interventions rather than general support. In the context of mental health care, the population faces an even higher risk of vulnerability to the system's suggestions or, conversely, rejecting a potentially valuable tool for support. To reduce these risks, development teams need to be transparent about the limitations of the technology. Thus, preventing users from over-relying on AI bots for complex mental health issues. Furthermore, the supervision of the systems' actions in the case of diagnosed mental health conditions by medical professionals is highly recommended. Finally, it is crucial to carefully consider the social implications of deploying AI systems for mental health. Moreover, concerns related to their efficacy, privacy, safety, and security need to be systematically addressed.

2.4 Healthcare administration

AI solutions are increasingly deployed in healthcare administration with the primary goal of reducing the administrative burden for professionals in the field. Whilst current research on AI in healthcare administration predominantly emphasizes software development rather than its application on mobile devices, the majority of these solutions have the potential to be integrated into portable devices.

The reduction of administrative burdens is needed given the immense workload of clinicians associated with merely administrative tasks [36]. Consequently, AI solutions focus on coordinating internal tasks and handling repetitive processes like prior authorization, updating patient records, and billing [37]. Likewise, the utilization of chatbots in the context of health has become more prevalent, despite major flaws in terms of quality and robustness as well as deficits in data protection [38]. Such chatbots are used, for example, to handle simple enquiries from a user about basic information regarding the hospital administration, or to schedule appointments.

A further application area for AI in healthcare administration is the handling of patient records. Natural language processing, for instance, can be used in compiling electronic health record documentation allowing clinicians to dedicate more time to patient interactions [39]. Additionally, AI solutions aim to simplify access to patient information for doctors when needed [40]. However, a major challenge lies in addressing data ethical and privacy issues coming along with the handling of sensitive patient data. A key concern would here the issue of sexually transmitted diseases. AI techniques are also being used to optimize various healthcare processes, particularly when allocating resources to patients or between different units. For example, solutions for scheduling doctor appointments have been developed using an agent-based approach early on [41]. Furthermore, AI can aid hospitals in predicting the length of patient stays during pre-admission, thereby enabling more efficient utilization of hospital resources.

Finally, agents including state actors or insurance also apply AI for health administration. Administrative solutions play a significant role when designing measures against pandemics [42], but also in the allocation of resources to patients [43]. Moreover, insurance, including health insurance, relies on AI when verifying claims and detecting fraud [44]. However, some of these algorithms have shown strong biases against vulnerable groups [43]. The tendency of including AI in administrative healthcare issues depicts therefore a larger trend in healthcare and is situated within an ethical debate on data privacy, bias and explainability, particularly when embedded in mobile solutions. The integration of AI ethics into mHealth presents a socio-technical challenge that necessitates both a technical and cultural transformation in AI development practices.

2.5 Challenges of AI application in healthcare

Despite the promises and opportunities that AI-enabled mHealth applications present. Like AI solutions in general, the adoption of AI-driven mHealth solutions for healthcare faces several obstacles. A major concern is trust, clinicians have a duty to deliver the best care for every patient. Thus, clinicians may be hesitant to adopt AI-driven solutions due to concerns about their reliability or a lack of trust in how the technology makes critical medical decisions In their study, Tucci, Saary and Doyle [45] highlight explainability, transparency, interpretability, usability, and education as

key factors that have been identified as crucial elements that influence healthcare professionals' trust in medical AI and facilitate effective clinician–machine collaboration in critical decision-making healthcare settings. A recent systematic review of mobile vision-based applications for food recognition and volume estimation found that only one out of the twenty-two survey applications attempted to provide explanations on features that influence model prediction [13]. Understanding how AI algorithms make decisions will help build trust in a new generation of diagnostic tools. Another obstacle facing the wide adoption of AI-enabled mHealth is the issue of bias and privacy. Because the effectiveness of AI algorithms relies heavily on the quality of the training data. Often AI models are not trained with inclusive and representative data that accounts for the lived experiences of all users, especially the underrepresented communities. Thus, ensuring fairness and ethical use of AI in healthcare becomes an essential and critical aspect.

The number of ethical principles and guidelines to ground ethical principles has been on the rise. A study by Jobin, Ienca and Vayena [46] highlighted that 84 guidelines and principles documents have been developed by different institutions. Whilst these principles and guidelines can be useful to guide clinical healthcare research, the extent to which these guidelines can guide the technical development of AI systems in healthcare remains to be seen. To fill the ethics operationalisation gap, we propose integrating ethical principles within the development lifecycle.

3 Methodology

Our ethical framework is built on existing ethical guidelines and principles, including the AI4People framework [47] and the findings of the EU High-Level Expert Group on trustworthy AI [48], but also wider human rights considerations [49]. Additionally, we incorporate relational perspectives to address human value concerns, as well as the moral tensions between individual rights and public health. The relational ethics will ensure that the predictions made by the AI solution are grounded in prioritizing the contextual understanding of patients, existing socioeconomic inequalities and historical biases.

Our framework is conceived and applied using the "ethics by design" approach, where ethical principles are incorporated iterative throughout the entire AI development pipeline (from requirements elicitation to deployment and maintenance) in the agile SDLC. Therefore, our approach ensures that ethical considerations are not just an afterthought but practically integrated into the development of AI solutions. We posit the best way to mitigate and address ethical consequences is by applying ethical principles in the SDLC that developers commonly use. Finally, we provide an example of how our case-based framework can be applied in practice, using practical examples of AI-enabled mHealth applications in image recognition, administrative AI and conversational AI.

4 Ethical principles and values for AI-enabled mHealth applications

Conventionally, software development is concerned with the designing, building, and testing of computer programmes. Thus, developers of software systems are predominantly occupied with solving the technical complexity inherent in software applications. In the context of AI applications, the complexity is wider involving the teams that make decisions on the data and algorithms to be used. This composition of AI systems generates ethical concerns. Moreover, the deployment and use of AI systems in critical areas such as healthcare can bring about normative tensions when human values are not upheld. Therefore, it is important to address a group of ethical challenges arising at the interface of technology and human values.

In this section, we discuss ethical concerns that have been discussed in the literature to address the ethical issues that arise when AI is applied in healthcare. It is important to note that there have been many ethical principles and guidelines developed in recent years, particularly in the field of AI ethics. The AI4People recommendations are considered a significant source of ethical principles for AI in the Western world, and they are largely based on bioethical principles. According to Floridi et al. [47], the bioethical principles remain relevant and can be adapted to address the unique challenges posed by AI applications in healthcare. As a result, AI4People's recommended ethical principles are comprised of five key values: (1) Autonomy, (2) Beneficence, (3) Non-Maleficence, (4) Justice, and (5) Explicability. Essentially, the AI4People recommendations added transparency and explainability to bioethics principles in healthcare. Transparency refers to the ability of users to understand how the AI system is developed and works. Whilst explainability is concerned with the ability of the AI system to provide explanations to affected users on how the system arrived at a particular decision.

In this paper, we identified seven ethical issues associated with AI in healthcare, using AI-enabled mHealth applications examples such as food recognition, blood pressure management, cardiovascular diseases and mental health management. The identified ethical concerns are fairness, agility, precision, safeguarding humanity, respect for others, trust and accountability, and robust and reproducibility. We admit that there is an overlap between the aforementioned principles and AI4People's recommendations [47] and the HLEG on trustworthy AI's recommendations [48]. However,

we rephrase our ethical values to encompass relational and communal aspects because we believe that healthcare is intrinsically a matter that affects all of society. Moreover, we posit that the relational aspects will ground mHealth AI systems to be informed by the lived experiences of all patients, particularly those who are disproportionally affected by algorithmic injustices.

4.1 Fairness

Traditionally, fairness has been associated with the ethical principle of justice, which is concerned with ensuring that AI systems treat all individuals equally and do not advantage the privileged few. Furthermore, the ethical principle of justice emphasizes the importance of equality to prevent any form of discrimination against vulnerable groups and to uphold an individual's right to challenge decisions made by AI systems [46]. Interestingly, a recent study outlined that the concept of fairness needs to also account for empowerment [50]. So, fairness entails equal treatment, no discrimination, equity and empowerment. Birhane [51] argues that relationality demands focussing on the disproportionately impacted, i.e. the most marginalised and underrepresented communities. This does not imply equal treatment of humans as individuals, but recognizing that achieving the best community health means focussing on the weakest [52]. For example, in the case of an AI visual application for food recognition, marginalized people could be blind users. The application can empower these users by providing easy alternative navigation that will enable blind users to effectively use the solution via voice-based interaction.

In the AI lifecycle, bias arises because of training data and the choice of model. Therefore, to reduce the risks of biassed datasets, training data must be representative of the population that the AI system is intended to serve. Again, relational ethics challenge the traditional view of the concept of fairness by emphasizing that fairness should not only be about the data used to train the model but should also include concerns related to data governance [53]. Viljoen argues that social relations are inevitably enforced or magnified by data relationships, resulting in harm to society. For example, a well-known study by Obermeyer et al. [43] found that commercial algorithms for predicting patients who need extra healthcare support were biassed against black patients. This is because the algorithm used healthcare costs as a proxy for healthcare needs, on average black patients had lower healthcare costs than white patients, even when they had the same healthcare needs. As a result, the algorithm predicted that black patients were less likely to need extra healthcare support, even when they did. Subsequently, black patients were less likely to receive necessary healthcare support, resulting in worse outcomes. A recent study has shown that bias is not only a data problem but the choice of model [54]. For example, Bayesian models are susceptible to generating false correlations and reinforcing socially constructed stereotypes [55].

Finally, fairness is meant to reduce societal bias and injustice as well as empower users, especially the vulnerable by ensuring equity. To ethically address this concern, development teams should take a human-centred approach to implementing fairness-aware algorithms to address concerns related to bias, fairness, and stereotypes whilst promoting equitable healthcare solutions.

4.2 Agility

Most AI systems in healthcare, assume health conditions as static. Therefore, use data points collected at selected time points to make predictions. In so doing, these algorithms do not account for the dynamic changes that happen during treatment or as the medical condition progresses over time. Here, agility refers to the ability of the algorithm to capture temporal changes in clinical events that occurs. In their work, Loftus et al. [6] reviewed 20 of the most cited healthcare algorithms in medical AI. They found that none of the algorithms they reviewed exhibited a level of agility. Agile predictive analytics that captures trends over time are particularly suitable for analysing the vast amount of electronic health records (EHRs) data [56]. Other studies have shown that physiological time series data, for instance, can be used to predict mortality and particular conditions like acute kidney injury [56, 57]. Therefore, it is important for AI applications to fully capture and account for complex clinical and lifestyle changes over time to enhance real-time clinical decision-making. Lastly, the AI system should make dynamic predictions by utilizing new data as it becomes available.

4.3 Precision

Precision refers to the ability of the model to accurately perform prediction tasks. Precision Score is calculated as the ratio of correctly predicted positive cases also known as True Positives (TP) to the total predicted positive cases, which includes both correct predictions and false positives (FP), see Equation (1).

$$Precision = \frac{TP}{(TP + FP)} \tag{1}$$

In healthcare, high accuracy can be achieved by accounting for multi-modality data and the complex nature of the disease. Because of the non-linear nature of medical conditions, simple models have been found to perform poorly [58]. Nevertheless, the advantage of simple models lies in the fact that they are easy to understand and interpret.

Therefore, trade-offs need to be made between accuracy and interpretability. In the context of an AI-enabled mHealth solution, accuracy is crucial. Similarly, it is also important that clinicians and users of such can easily understand how these algorithms make decisions. For example, a computer vision-based application for food volume estimation should use comparative approaches to the manual assessment of meals that dietitians and patients currently use. This will make it easier for users to validate the accuracy of the model.

An AI-driven mHealth application for blood pressure management will have many input variables, such as continuous blood measurement, and physiological and clinical events data. This can reduce the model's ability to generalize well because of overfitting. To achieve the best results, a balance between the complexity of input data and predictive accuracy should be considered to determine the minimum number of variables required to maintain high performance. In the literature, a novel framework has been proposed for balancing model complexity with descriptive ability whilst avoiding overfitting [59].

4.4 Safeguarding humanity

Safeguarding humanity is related to the ethical principles of beneficence (do good) and non-maleficence (do no harm). The goal is to ensure that the AI solution operates as intended and does not cause harm. Prevention of harm means that the developers or other crucial decision-makers are considering potential harm to patients and the community as a whole. The communal aspect is important because healthcare is a matter that intrinsically affects all members of society. Therefore, individual human rights are also upheld by proving equitable care for all patients [52].Additionally, as emphasized by relational ethics such as Ubuntu, communal relationships involve constructing solidarity, across diversity, achieved through empathy and support, with the aim of enhancing the overall well-being of everyone involved [60]. Hence, the design, deployment and use of AI-enabled mHealth systems in healthcare must be compatible with maintaining the bonds of solidarity amongst people and generations. Finally, the design, development and use of AI applications in healthcare should comply with existing legal regulations, such as the General Data Privacy Regulation (GDPR) [61] and alike that aim to safeguard the well-being of society.

4.5 Respect for others

The AI system should exhibit compassion and care for all users especially the most vulnerable. This implies respect for diversity of what means to be human. AI systems should protect the human oversight and privacy of all users.

From classical AI, autonomous algorithms refer to the ability of the system to operate with limited human interference. This implies that the algorithm should be able to incorporate new data with minimal user involvement, including capturing longitudinal data from diverse sources. However, when it comes to decision-making, trade-offs should be made between automation and human oversight. For instance, during complex prediction or classification tasks, such as food recognition or volume estimation, human oversight may be crucial. This is because it is difficult to automatically detect ingredients, such as salt, ingredient of soup or drink [62, 63]. In this case, the user should provide additional information. In addition, users of AI-enabled mHealth applications must be able to anticipate the outcomes of the system.

In terms of privacy, patients' health data are sensitive and classified as confidential information in many jurisdictions [3]. AI algorithms require a lot of data for them to generalize well. For an AI algorithm to personalize prediction to each individual, personal data is required. Traditional ML requires data to be stored at the central server, such an approach requires personal data to be transmitted to external servers. This presents a risk that such personal data may be misused, leading to identity theft, cyberbullying, or other malicious activities. The right to privacy is a human right concerned with ensuring that individual' information is protected and securely managed [3].

Privacy issues are a subject of ongoing debate as more AI algorithms are being applied in healthcare. As a solution, researchers have proposed collaborative ML approaches such as federated learning (FL), whereby a global model is created by consolidating locally trained models [64]. Though with FL there is no explicit data sharing, as local models train, they send back insights (coefficients and gradients) that are incorporated in the global model. However, even with this distributed approach, the disclosure of private information can happen when adversaries deduce whether a particular attribute is part of the model's training data or infer class representatives from collaborative models [65, 66]. For mHealth AI applications, it is essential that user personal data remains on the device and inferences are performed on the edge device. Finally, when privacy leakages occur, they should be quantified.

4.6 Trust and accountability

Trust and ethics are indivisibly linked. Trust plays a crucial role in ethical living as it impacts all aspects of individual, communal, and business relations. However, when it comes to AI ethics, we have been more concerned about assessing the trustworthiness of AI systems rather than establishing long-term relationships between users and these systems [52]. For example, the EU High-Level Expert Group on AI

outlined that Trustworthy AI has three components: lawful, ethical and robust [48].

Whilst each of these three components is essential, none of them alone is adequate to accomplish AI that can be considered trustworthy. Trust is much broader than just assessing the technical aspects, such as explainability and transparency, to determine whether an AI system is trustworthy or not. As argued by Cotterrell [67], trust encompasses the belief in the benevolence and abilities of others, as well as the belief that common expectations in similar social situations will not be frustrated. Relational ethics such as Ubuntu root trust in the interconnectedness and interdependence of individuals within a community, where trust is established through long-term relationships with the community.

When the algorithm fails, decisions made by the model can harm patients. In this case, there should be mechanisms to determine who is responsible and prescribe the right actions to remedy the harm caused. For example, if the AI algorithm predicts the amount of insulin wrongly and the patient goes into a state of hyperglycemia. There should be mechanisms to hold the company or developers of the system accountable for such a failure. This can be achieved by establishing robust mechanisms to monitor and address the negative consequences of AI. Additionally, it is important to have a clear mapping of responsibilities throughout the lifecycle of AI, from requirements elicitation to deployment and maintenance, to ensure accountability for any potential negative impacts and to promote the ethical and responsible use of AI. Accountability can be realized and enforced by an impartial oversight body.

Finally, to attain ethical AI, trust must be established at all levels of society. The AI community must adopt a broad perspective on trust and acknowledge that it is a long-term objective that requires community involvement. Building trust is a time-intensive process, and AI developers should not wait until they have collected data to start trust-building with communities. Institutions and individuals creating AI solutions in healthcare must continuously engage with communities regarding the issues they want to address and how AI can be utilized to solve them.

4.7 Robust and reproducibility

Robustness is concerned with ensuring that the AI system operates reliably throughout its entire lifecycle. For example, given the same inputs, the AI system should always produce the same results as expected. AI systems should generate logs that will keep track of the processes, datasets and decisions. This will enable the analysis of the outcomes produced by the AI system. Thus, the AI system's outcomes are consistent with design ideas and healthcare providers. In the context of AI-enabled mHealth Applications, robustness refers to the ability of the application to perform effectively and reliably under various conditions and scenarios. A robust AI system is designed to withstand and adapt to uncertainties, adversarial inputs, and unexpected situations without compromising its performance or accuracy. Robustness in AI-enabled mHealth systems is crucial because healthcare environments are often dynamic and complex.

According to a survey conducted by the scientific journal Nature, more than 70% of researchers have made unsuccessful attempts to replicate experiments conducted by other scientists [68]. Reproducibility is an important aspect not only for scientific research but for ML as it helps foster trust and increase the credibility of the solution [6]. Before an AI solution goes into the clinical trial phase, it should be externally validated with external data cohorts and prospectively. Overall, we need AI standards to improve the reproducibility of ML. Thus, ensure that AI systems are developed ethically and do not pose any risks to patients. In the next section, we discussed operationalizing the aforementioned ethical principles iteratively across all phases of the agile development process.

5 Framework for operationalising AI ethics

Creating AI systems that meet normative standards is a challenging task that cannot be solved simply by urging developers to be more "ethical" in their work. Due to the complex nature of AI systems, we cannot know all of their outcomes in advance. Additionally, some unexpected interactions and motivations may only become apparent after a product has been applied over time. Some researchers have argued that issuing developers with codes of ethics and conduct allows organisations that develop AI systems to self-regulate and provide guidance to developers [48].

Though helpful, codes of ethics and conduct alone are insufficient for effectively addressing AI systems development values and principles. Therefore, comprehensive measures are necessary such as operationalizing AI ethical values and principles throughout the agile AI system development lifecycle. The agile development lifecycle comprises four phases: requirements elicitation, design and development, testing, and deployment. Moreover, we recommend that development teams consider techniques to assess how the developed mHealth application aligns with the 12 principles of agile software development, which include (1) user satisfaction; (2) accommodating requirements changes; (3) frequent delivery of working software; (4) collaboration with stakeholders; (5) support, trust and motivation; (6) Effective communication; (7) Measure progress through functional delivery of system; (8) Consistency in development and maintenance; (9) Attention to technical detail and improve design agility; (10) Simplicity; (11) Self-organising teams;

(12) Regular reflections on how to improve effectiveness in the team [69].

Figure 1 illustrates how the identified ethical values and principles in section 3 can be practically incorporated into the agile SDLC. The principle we have identified should be addressed in different phases of the SDLC throughout the AI lifecycle. However, because ethical concerns are not uniformly distributed across the AI lifecycle, we propose that developers shift their focus from a single ethical principle, as suggested in certain guidelines [70], to contextualized ethical principles most relevant to the particular phase. Finally, trade-offs between ethical principles should be thoughtfully considered and well-documented [71], utilising frameworks such as the one proposed by [72] to assess ethical tensions.

5.1 Requirement elicitation

During the requirements elicitation phase, ethical principles should be identified and aligned with the design of the solution, including considering specific human values, biases and lived experiences of users. Ethical principles and values are dependent on the specific use case. In the context of AI-enabled mHealth applications, we identified 7 ethical principles and values: fairness, agility, precision, safeguarding humanity, respect for others, trust and accountability, and robust and reproducibility. In terms of fairness, developers of AI-enabled mHealth solutions should ensure that AI systems treat all patients equally, by considering both the current privilege and potential empowerment of individuals.

Bias can occur at different stages, including the data used to train the model, the algorithms used to build it, and how it is implemented and used. In the context of AI-enabled mHealth applications, bias concerns can be addressed by studying cultural values and perspectives, as well as curating a representative dataset. For example, a dataset for food image recognition should include images of diverse kinds of food, representative of the diverse communities the application intends to serve. Similarly, an algorithm to personalize blood pressure management should include diverse patients, in terms of race, gender, age and physical fitness. Since the impact of AI on different demographic groups can vary, it is essential to consider the potential biases that may arise in the development process.

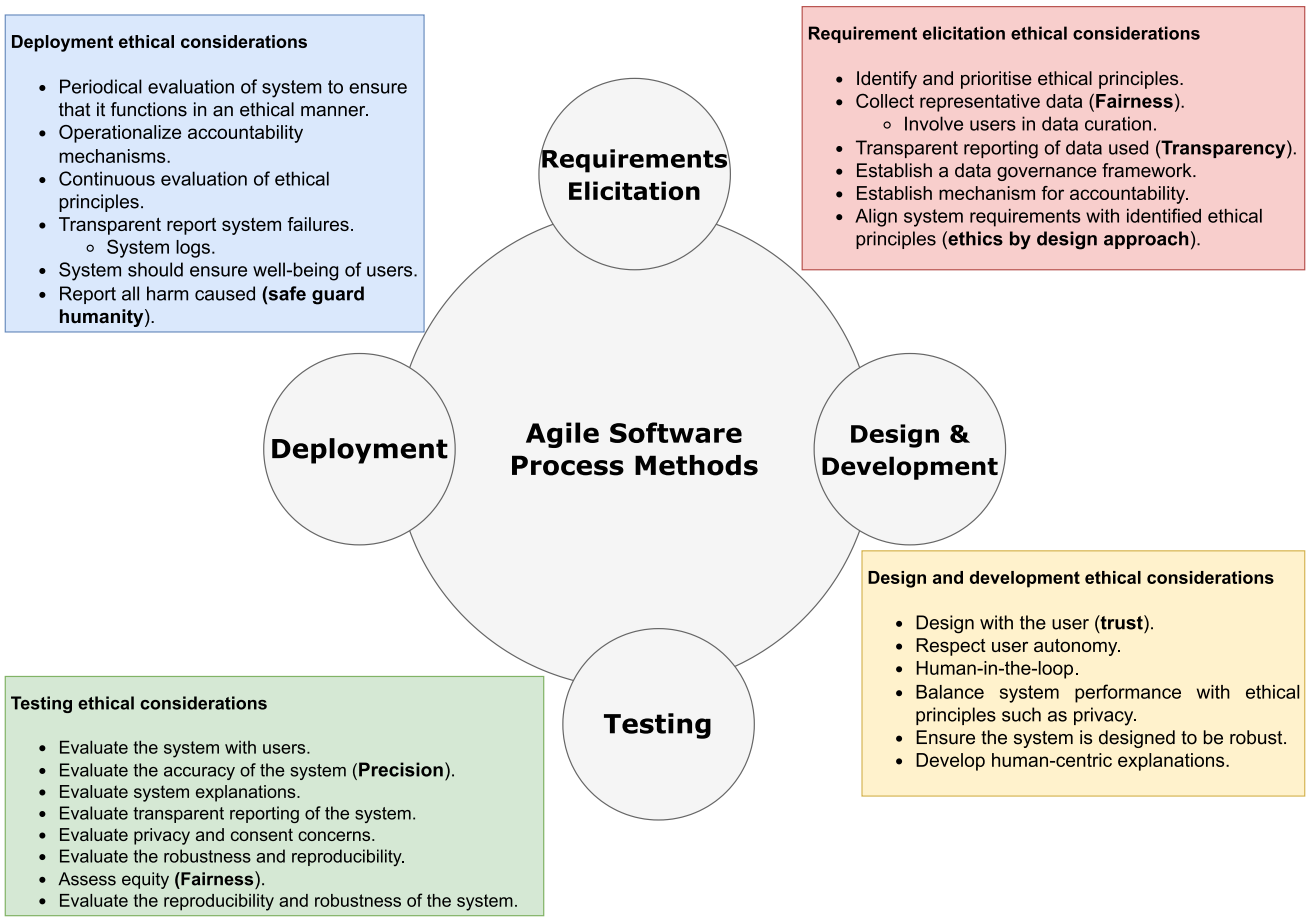

Fig. 1 Embedding ethical principles into the agile software development process

During data collection, it is essential for the development team to not only ensure that the data is fairly representative but also to establish a robust data governance strategy that empowers users, particularly the most vulnerable. Data governance plays a crucial role in managing data relations that inherently amplify social relations, as highlighted by [53]. If not managed effectively, these data relationships can cause harm to users of the system. In addition, the comprehensive data governance strategy should include the protection of data workers, including in cases where data cleaning is outsourced. When data curation is outsourced, a data governance strategy should provide clear guidelines and procedures that prioritize the well-being and fair treatment of content moderators, including fair compensation and conducive working conditions. Finally, a data governance strategy should outline proactive measures to prevent exploitation by implementing strict monitoring of working conditions.

Concerning privacy concerns, it is important that from the onset developers of AI systems engage different communities where they intend to deploy the system to understand how communities view privacy and to identify cross-cultural differences. These cross-cultural views are important to inform the privacy design for AI technologies [73]. During the requirements elicitation phase, developers of mHealth applications should robustly engage communities using collaborative activities, such as ideation, hackathons or workshops, to understand individual differences in privacy.

Formal methods should be applied to ensure the precision, correctness, and safety of software systems [74]. Formal specifications provide a rigorous approach to specifying software systems, especially mission-critical applications such as healthcare solutions. Consider an AI-based mental health application that uses natural language processing (NLP). Formal methods in this context entail establishing a formal language for input and output, applying mathematical models for NLP processes, performing formal verification for precision, defining safety properties, and incorporating error-handling mechanisms. By leveraging formal methods, developers can ensure the creation of robust and reliable solutions.

Given the complex nature of healthcare systems and the involvement of various stakeholders with different needs and priorities, such as providers, patients, insurers, and regulators. It is important to identify and prioritize ethical principles to ensure that they are effectively addressed. For AI-enabled mHealth applications, patient-centricity with an equity goal is recommended to prioritize the interests of patients, especially vulnerable populations. Lastly, effective communication between AI developers and healthcare stakeholders is critical to ensure understanding and alignment on the objectives of the system. It may require bridging the gap between technical language and healthcare domain-specific language, managing expectations, and conducting user experience research to ensure understandability on every level.

5.2 Design and development

During this phase, participatory design with stakeholders, including developers, clinicians, the community and patients (users) of the system must be all involved and the goal of the solution should be highlighted, i.e. specify the aim and expected outcomes. For an AI-driven mHealth application, design decisions need to be made to balance trade-offs between system performance and privacy. Whilst precision and robustness are important for mHealth applications, it is also crucial that users' personal information is protected. To achieve this, developers should explore various techniques to model training, such as traditional ML (where model training and inference take place at a central server) [75] and decentralized approaches (collaborative training without data leaving the user's mobile device) [64, 65]. Traditional ML requires personal data to be transmitted to a central server to perform inferences. Centralized learning is usually well-resourced in terms of processing power, i.e. models are better trained using high-performance computational servers. This approach also allows the models to be deployed and used at scale. However, health data is confidential in nature and when such data is shared onto the cloud, data privacy can be compromised. To overcome the privacy challenges, collaborative approaches such as FL have been proposed.

Decentralized collaborative approaches are good for preserving user privacy by keeping user data on the mobile device. This ensures that patients maintain ownership of their data. Such an approach adheres to the principle of data minimization in compliance with the General Data Protection Regulation (GDPR). This approach aligns with the GDPR's principles of function and storage limitation. However, performing local training and inferences on the user's mobile device comes with challenges. For example, in large-scale development, client heterogeneity in terms of data and devices is inevitable. This can have an impact on the quality of model training in terms of accuracy, fairness, and time. Other challenges include expensive communication, data leakages and battery consumption. To address the energy consumption concern, a power-aware algorithm was developed [76]. The algorithm selects clients with higher battery levels and uses them to optimize system efficiency. Stakeholders must participate in these decisions, including the choice of the model to be applied because these decisions ultimately affect them. An ideal algorithm for mHealth applications should be federated to protect user data and conform to existing regulations such as GDPR.

In the design and development phase, developers need to ensure that the "Ethics by design" approach is thoroughly followed to align expected outcomes with human values.

This can be achieved by human-centred approaches, such as "human in the loop" and model cards. In addition, the choice of model is also important because some models are prone to make biassed causal inferences [55], and other models are known to be "black boxes". In high-risk applications such as mobile healthcare, it is better to use models that are easily explainable and interpretable. The explainability of AI is essential to building trust and increasing the adoption of AI systems in healthcare. Creating awareness amongst AI developers about the level of explainability in ML models can help build more explainable AI systems. However, a recent review study on computer vision applications for food recognition, volume estimation and calorific estimation found that only one out of 22 studies attempted to provide explanations on how the model makes decisions to the end users [13]. This illustrates that we still have a long way to go to create explainable AI applications for mobile health. Ideally, AI developers need to ensure that explanations provided are human-centred, by linking characteristics in the data to domain knowledge that will allow experts to understand a given output and factors that have influenced the given outcome [77]. For example, if a food recognition application predicts the kinds of food comprising a given dish. An explanation could be an annotation that outlines all items that make up the dish, an example is provided in Fig. 2. As illustrated in Fig. 2B, a system for food recognition needs to keep the human in the loop, for instance, by asking the user if recognition of the different kinds of food on a given plate is correct.

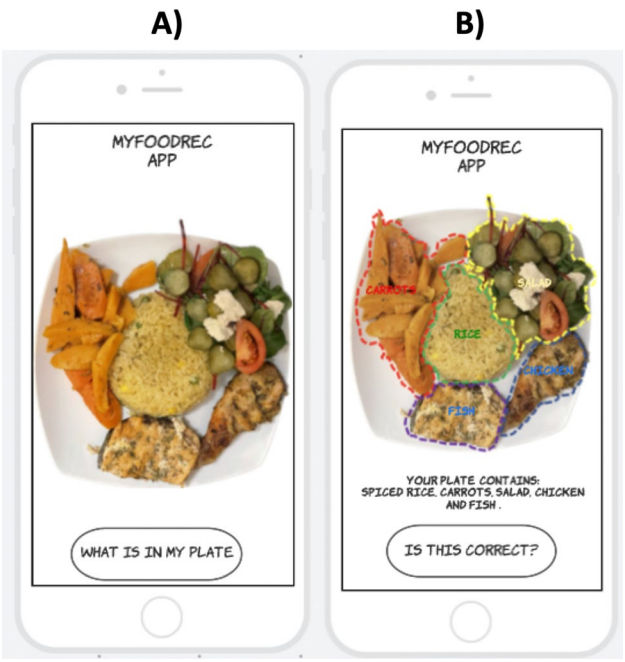

Fig. 2 An example illustrating a possible explanation for a food recognition case. **A** Show a food plate with different types of food. **B** A possible explanation by segmenting and recognising each type of food on the plate. Human autonomy is very important; thus, the system needs to ask the user if the classification of the types of food on the plate is correct

During the design and development phase, developers should ensure that the system is designed to be agile, enabling it to capture clinical and physiological changes over time. This can be realised by incorporating dynamic information from diverse sources, such as EHRs and other sources, which will enable the system to capture complex changes that occurs or conditions that changes rapidly. For instance, a mHealth application that helps patients monitor their blood pressure should use temporal evolution of blood pressure measurements generated by the application to predict a patient's blood pressure. The continuous measure is essential to accurately evaluate the blood pressure trending ability, allowing accurate prediction of blood pressure [78].

The rise of generative AI also poses risks to mHealth applications. First, generative AI tools can generate images or videos of food that is similar to real dishes. Therefore, developers of vision-based mHealth applications must anticipate the possibility of their solutions being misused in the real world and ensure that vision-based mHealth applications can deal with deep fakes. Second, the increased use of large language models (LLMs) in our society through conversational AI, such as chatGPT and Bard, accessible via mobile devices, such as smartphones and tablets, poses potential dangers. Whilst LLMs may be used to assist with diagnosis and treatment recommendations, they are not always accurate [79]. Inaccurate diagnoses or recommendations can have serious irreversible consequences for users. For example, a Belgian man was reported to have killed himself after a conversation with an AI chatbot [80, 81]. Thus, these tools should not be accessible to users with different mental disorders, such as generalized anxiety disorder and hypochondria. Lastly, LLMs are trained with large datasets of text from the internet is dominated by hegemonic viewpoints and encode biases and inequalities, disproportionally affecting marginalized communities [82]. Thus, LLMs for healthcare need to be developed differently to ensure they enable clinicians to provide equitable care. Additionally, developers should avoid developing general-purpose AI tools for healthcare but rather focus on developing specialized AI tools to aid with specific health problems.

Finally, the use of AI-driven mHealth solutions raises ethical questions about the delegation of authority and autonomy in healthcare decision-making and the potential loss of human empathy and connexion in patient care. Therefore, developers should ensure ethical considerations are operationalized in the design/development phase and that AI-driven systems foster mutually beneficial relationships.

5.3 Testing

After the development, different versions of the AI-driven mHealth applications should go through various testing phases with real users. There should be a feedback loop to ensure that users' feedback is incorporated into the next iteration of the system. Thus, ensure the system meets stakeholder expectations whilst prioritizing patient-centricity and equity for the marginalized. During the testing phase, we suggest a systematic integration of the ethical framework and the testing process. This will enable developers to guarantee that both black box and white box AI-based algorithms align with ethical principles, leading to the development of reliable and socially responsible AI systems. For white box models, developers should test the internal workings and structures of the system. However, the majority of the algorithms in healthcare employ DL techniques, thus are Black box algorithms. For black box models, developers should apply model interpretability methods such as Local Interpretable Model-agnostic Explanations (LIME) [83], SHapley Additive Explanation (SHAP) [84] and Textual Explanations for Visual Models [85] to gain insight into the decision-making process and inner logic of a given model. This makes it easy to transparently interpret when the model makes mistakes about a particular prediction.

Moreover, testing scenarios should be created to assess whether ethical requirements are met. For example, the system can be tested to determine how it responds to different input data (robustness), how it handles sensitive patient data and whether it makes fair and unbiased decisions. In addition, given the same inputs, the output of the system should be the same (reproducibility). Test whether there are any biases associated with age, race and gender and higher or lower likelihoods of certain conditions. Not accounting for these can lead to results that reinforce existing inequities in health.

Causality is increasingly being applied to ML approaches in healthcare [77, 86]. Causality is concerned with establishing the relationship between the cause and effects in the data being analysed [86]. Causality was first developed in the context of econometrics, social sciences and epidemiology where the variables being studied are usually scalars [87]. In the context of healthcare, the use of diverse data (structured and unstructured), such as images and text, is a challenge for the extraction of meaningful information for ML [86]. For example, a mHealth application for remote blood pressure management is complex with many confounding factors, such as diet, exercise, stress levels, blood pressure measurements, medications and hormones [88]. Establishing causal inferences between these different confounding factors can be a challenge. To address this, developers must design in silico realistic scenarios to determine the causal relationship between blood pressure and other variables whilst controlling for potential confounding variables, which can help establish causality and ensure that the AI system is making accurate predictions.

Assessing the accuracy of AI-powered conversational mHealth applications is crucial to ensure their reliability and safety when interacting with patients and providing medical information. Thus, development teams should conduct real-world scenario simulation experiments using expert evaluations, including clinicians and patients. Additionally, AI-powered mHealth systems should be comprehensively benchmarked against clinical gold standards.

Data security is another challenge, developers should perform robust testing under various conditions to identify security vulnerabilities such as data leakages. Thus, ensuring that user-sensitive health data is protected. Explainable AI (XAI) tools should be applied to test the transparency and understandability of the system, particularly in healthcare, where potential biases and weaknesses in accuracy and fairness can have serious implications for patient health and safety. However, development teams need to be careful about reducing complex medical conditions to a number. Finally, to determine the degree to which human values are incorporated into the AI solution, a combination of qualitative and quantitative metrics must be developed, tested, and validated. Human-centred explanation evaluations need to be performed with the potential users.

5.4 Deployment

During the deployment phase, mechanisms for continuous monitoring must be established. This will ensure that the developed AI system continues to meet ethical requirements and that it is functioning as envisaged. Developers should also ensure that there is a continuous feedback loop, which will enable users to provide feedback and flag issues, such as biases, inaccurate predictions and security issues. Developers should ensure that there is transparent ethical and technical reporting of system failures, e.g. system logs and sustainability reports. Such reporting can help the public understand the rationale behind the AI system's decisions, how it works, and its limitations, reducing the risk of misunderstandings or mistrust. In addition, transparent reporting demonstrates the developers' commitment to adhering to ethical principles and can provide reassurance to the public that the AI system has been developed and deployed responsibly.

Comprehensive mechanisms and clear guidelines should be established to foster accountability when system failures occur. If a system failure occurs, the development team should: (1) identify the root cause of the failure should be established. (2) Implement corrective actions to remedy the erroneous action. This can be done by updating the model. 3) Transparent communication to stakeholders, including an

explanation about the failure, correction steps taken and the potential effects on patients and their ecosystem.

Lastly, organisations deploying AI should implement ethics-based auditing using structured processes [9]. This allows them to evaluate the extent to which their AI systems align with ethical principles and identify any gaps that require corrective actions. In addition, put mechanisms in place to address identified gaps, therefore, improving user satisfaction and fostering trust in AI systems [45, 89].

6 Discussion

In this study, we identify ethical concerns associated with AI-enabled mHealth applications. Moreover, we illustrate how these ethical issues can be addressed iteratively within the agile SDLC using practical examples of AI-driven mHealth systems. We suggest that the AI ethics community in healthcare should focus on contextualized case-based principles rather than general ethical frameworks and guidelines. This will enable developers to focus on and address ethical issues most relevant to the problem throughout the development lifecycle from requirements elicitation to deployment. By operationalizing AI ethics in a familiar process, developers can better reason for trade-offs between performance and ethical principles throughout the development lifecycle.

Despite the growing body of literature mapping ethical principles and guidelines in healthcare. Principles alone cannot guarantee ethical AI [90]. There is a need to operationalize AI ethics. A study in software engineering has found that incorporating human values into software development can be achieved through an evolutionary approach, rather than a revolutionary one, by building upon existing practices [91]. They found that organizational culture plays a role in addressing human values and ethical principles. Furthermore, they suggested agile methods can be modified to incorporate a greater emphasis on human values, allowing businesses to gradually integrate values-based thinking into their current processes rather than undergoing a complete overhaul. AI development lifecycle is not only closely related to software development, developers of AI solutions follow the SDLC and draw a lot of practices from the software engineering domain [92, 93]. Lastly, AI-enabled mHealth applications are developed following software development processes such as agile methods. Therefore, addressing ethical implications through an iterative and continuous process from the start of development could effectively integrate ethical considerations into the practical development of mobile medical AI solutions. In this study, we propose integrating ethical principles within the SDLC will enable developers of mobile AI systems in healthcare better address its ethical concerns whilst accounting for the technical capabilities of the solution.

6.1 Limitations

Our study is not without any challenges. First, we acknowledge that we lack oversight mechanisms to effectively align AI development in the healthcare sector. We believe that regulations and governance mechanisms play a crucial role in ensuring the alignment of AI with ethical values. Second, we provide a practical demonstration of how ethics can be integrated into the agile software development process using a specific use case. Further studies are needed to showcase how AI ethical principles can be operationalized to other healthcare problems, such as medical imaging. Third, we understand that there may be differences between the conceptual and practical implications of our proposed framework. Thus, we provide an unambiguous way to operationalise ethical principles for mHealth applications within the agile process which developers of mobile applications are familiar with and trained to use. Lastly, we demonstrate how our framework can be operationalised using mHealth examples. Though the identified principles and guidelines are specific to AI-enabled mHealth solutions. Our approach to operationalizing AI ethics in the SDLC throughout the AI development pipeline can be applied to other domains.

6.2 Future work

This study illustrates how ethical principles can be integrated into the agile SDLC, addressing ethical concerns iteratively across all phases. Future work should explore how to standardize the operation of ethical principles. This would enable consistent implementation of guidelines and principles. In the literature, the use of checklists to assess the integration of guidelines in AI development has been proposed [6, 70, 94]. However, checklists have limitations, such as developers and organizations, working on AI systems can become reliant on checklists and erode critical thinking [70]. Additionally, other authors have argued that the technology industry cannot self-regulate, and voluntary ethical compliance is a strategic effort by big tech to prevent legally enforceable regulations [95]. Therefore, the is a need to translate ethical principles into inclusive regulations. When it comes to regulations, the European Union with the EU AI Act [96] (which aims to categorize AI applications based on their risk status) is leading. However, we are yet to see how this law will be enforced because it is difficult to know how existing AI systems are being used. One way to uncover how AI systems work is to have mandatory public vetting, where communities are given a chance to interact with the system and uncover biases and errors before the system is approved for deployment. Ideally, we recommend AI

regulations become agile to be able to keep up with technological advances. Finally, there is a need to establish meaningful global coordination that safeguards the development of AI in critical sectors such as healthcare.

In this study, we propose a shift from one-size-fits-all ethical frameworks to contextualized case-based principles to assist AI developers in values-conscious development, the aim is to identify specific ethical values that can be operationalized. Finally, incorporating ethical principles into the agile development process will increase the sensitivity of AI developers to the ethical dimensions of the technology.

7 Conclusion

The widespread availability of powerful smart mobile devices presents opportunities for global accessibility of AI-powered medical applications. These applications can play an essential role in empowering patients better manage their conditions whilst enabling clinicians to make personalized and effective clinical decisions. Therefore, improve outcomes for patients. However, the rise of AI in healthcare can lead to disproportionate effects on marginalized communities, possibly exacerbating health disparities. AI ethics has garnered significant attention to ensure the ethical development, deployment, and usage of AI systems. Presently, existing frameworks and principles primarily concentrate on adhering to ethical principles and do not provide practical approaches to operationalizing AI ethics. Whilst some studies have aimed to operationalize AI ethics across the AI lifecycle, they often assume a one-size-fits-all approach. This paper proposes a contextualized case-based framework that empowers developers to operationalize ethical principles within the agile SDLC using practical examples of AI-enabled mHealth applications. We emphasize the crucial role of community involvement in the development of human-centred AI systems for healthcare, advocating for co-designing AI systems with the local community and expert clinicians. Such collaborative efforts aim to enhance trust in AI systems within the healthcare domain. Finally, we offer examples to operationalise ethical principles throughout the entire development lifecycle, ensuring that mHealth AI systems are ethically grounded and aligned with the needs and values of the communities they intend to serve.

Acknowledgements This work was supported by the Institute for Ethics in Artificial Intelligence (IEAI) at the Technical University of Munich.

Author Contributions First author prepared the draft manuscript and analysis, whilst the co-authors contributed to the study design and provided suggestions for changes. All authors have thoroughly read and approved the published version of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Not applicable.

Availability of data and materials: Not applicable

Declarations

Conflict of interest Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Code availability Not applicable.

Consent for publication Not applicable.

References

1. U.S. Food and Drug Administration: Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices (2022). https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices
2. Albahri, O.S., Albahri, A.S., Zaidan, A.A., Zaidan, B.B., Alsalem, M.A., Mohsin, A.H., Mohammed, K.I., Alamoodi, A.H., Nidhal, S., Enaizan, O., Chyad, M.A., Abdulkareem, K.H., Almahdi, E.M., Al. Shafeey, G.A., Baqer, M.J., Jasim, A.N., Jalood, N.S., Shareef, A.H.: Fault-tolerant mhealth framework in the context of iot-based real-time wearable health data sensors. IEEE Access 7, 50052–50080 (2019). https://doi.org/10.1109/ACCESS.2019.2910411
3. Gerke, S., Minssen, T., Cohen, G.: Ethical and legal challenges of artificial intelligence-driven healthcare, 1st edn., pp. 295–336. Elsevier (2020). https://doi.org/10.1016/B978-0-12-818438-7.00012-5
4. Calo, R.: Artificial intelligence policy: A roadmap. SSRN Electron. J. (2017). https://doi.org/10.2139/ssrn.3015350
5. Reddy, S., Allan, S., Coghlan, S., Cooper, P.: A governance model for the application of AI in health care. Journal of the American Medical Informatics Association 27(3), 491–497 (2019) https://doi.org/10.1093/jamia/ocz192 https://academic.oup.com/jamia/article-pdf/27/3/491/34152225/ocz192.pdf
6. Loftus, T.J., Tighe, P.J., Ozrazgat-Baslanti, T., Davis, J.P., Ruppert, M.M., Ren, Y., Shickel, B., Kamaleswaran, R., Hogan, W.R., Moorman, J.R., Upchurch, G.R., Rashidi, P., Bihorac, A.: Ideal algorithms in healthcare: Explainable, dynamic, precise, autonomous, fair, and reproducible. PLOS Digital Health 1, 0000006 (2022) https://doi.org/10.1371/journal.pdig.0000006

7. World Health Organization: Ethics and governance of artificial intelligence for health: WHO guidance (2021). https://www.who.int/publications/i/item/9789240029200
8. Hagendorff, T.: The ethics of AI ethics: An evaluation of guidelines. Mind. Mach. **30**(1), 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8
9. Solanki, P., Grundy, J., Hussain, W.: Operationalising ethics in artificial intelligence for healthcare: a framework for ai developers. AI and Ethics (2022). https://doi.org/10.1007/s43681-022-00195-z
10. Boushey, C.J., Spoden, M., Zhu, F.M., Delp, E.J., Kerr, D.A.: New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. Proceedings of the Nutrition Society **76**(3), 283–294 (2017). https://doi.org/10.1017/S0029665116002913
11. Lo, F.P.W., Sun, Y., Qiu, J., Lo, B.: Image-based food classification and volume estimation for dietary assessment: A review. IEEE J. Biomed. Health Inform. **24**(7), 1926–1939 (2020). https://doi.org/10.1109/JBHI.2020.2987943
12. Tahir, G.A., Loo, C.K.: A comprehensive survey of image-based food recognition and volume estimation methods for dietary assessment. Healthcare **9**(12) (2021) https://doi.org/10.3390/healthcare9121676
13. Amugongo, L.M., Kriebitz, A., Boch, A., Lütge, C.: Mobile computer vision-based applications for food recognition and volume and calorific estimation: A systematic review. Healthcare 11, 59 (2022) https://doi.org/10.3390/healthcare11010059
14. Luo, Y., Ling, C., Ao, S.: Mobile-based food classification for type-2 diabetes using nutrient and textual features. In: 2014 International Conference on Data Science and Advanced Analytics (DSAA), pp. 563–569 (2014). https://doi.org/10.1109/DSAA.2014.7058127
15. Oliveira, L., Costa, V., Neves, G., Oliveira, T., Jorge, E., Lizarraga, M.: A mobile, lightweight, poll-based food identification system. Pattern Recogn. **47**(5), 1941–1952 (2014). https://doi.org/10.1016/j.patcog.2013.12.006
16. Merchant, K., Pande, Y.: Convfood: A cnn-based food recognition mobile application for obese and diabetic patients. In: Shetty, N.R., Patnaik, L.M., Nagaraj, H.C., Hamsavath, P.N., Nalini, N. (eds.) Emerging Research in Computing, Information, Communication and Applications, pp. 493–502. Springer, Singapore (2019)
17. Bolaños, M., Ferrà, A., Radeva, P.: Food Ingredients Recognition through Multi-label Learning (2017)
18. Bashar, S.K., Hossain, M.-B., Lázaro, J., Ding, E.Y., Noh, Y., Cho, C.H., McManus, D.D., Fitzgibbons, T.P., Chon, K.H.: Feasibility of atrial fibrillation detection from a novel wearable armband device. Cardiovascular Digital Health Journal **2** (2021) https://doi.org/10.1016/j.cvdhj.2021.05.004
19. Isakadze, N., Martin, S.S.: How useful is the smartwatch ecg? Trends Cardiovasc. Med. **30**(7), 442–448 (2020). https://doi.org/10.1016/j.tcm.2019.10.010
20. Semaan, S., Dewland, T.A., Tison, G.H., Nah, G., Vittinghoff, E., Pletcher, M.J., Olgin, J.E., Marcus, G.M.: Physical activity and atrial fibrillation: Data from wearable fitness trackers. Heart Rhythm 17(5, Part B), 842–846 (2020) https://doi.org/10.1016/j.hrthm.2020.02.013 . Digital Health Special Issue
21. Rodriguez-León, C., Villalonga, C., Munoz-Torres, M., Ruiz, J.R., Banos, O.: Mobile and wearable technology for the monitoring of diabetes-related parameters: Systematic review. JMIR Mhealth Uhealth **9**(6), 25138 (2021). https://doi.org/10.2196/25138
22. McMahon, S.K., Lewis, B., Oakes, M., Guan, W., Wyman, J.F., Rothman, A.J.: Older adults' experiences using a commercially available monitor to self-track their physical activity. JMIR Mhealth Uhealth **4**(2), 35 (2016). https://doi.org/10.2196/mhealth.5120
23. Beauchamp, U.L., Pappot, H., Holländer-Mieritz, C.: The use of wearables in clinical trials during cancer treatment: Systematic review. JMIR Mhealth Uhealth **8**(11), 22006 (2020). https://doi.org/10.2196/22006
24. Gamble, A.: Artificial intelligence and mobile apps for mental healthcare: a social informatics perspective. Aslib Journal of Information Management 72, 509–523 (2020) https://doi.org/10.1108/AJIM-11-2019-0316
25. Milne-Ives, M., Selby, E., Inkster, B., Lam, C., Meinert, E.: Artificial intelligence and machine learning in mobile apps for mental health: A scoping review. PLOS Digital Health **1**(8), 1–13 (2022). https://doi.org/10.1371/journal.pdig.0000079
26. Abd-alrazaq, A.A., Alajlani, M., Alalwan, A.A., Bewick, B.M., Gardner, P., Househ, M.: An overview of the features of chatbots in mental health: A scoping review. International Journal of Medical Informatics 132, 103978 (2019) https://doi.org/10.1016/j.ijmedinf.2019.103978
27. Damij, N., Bhattacharya, S.: The role of ai chatbots in mental health related public services in a (post)pandemic world: A review and future research agenda. In: 2022 IEEE Technology and Engineering Management Conference (TEMSCON EUROPE), pp. 152–159 (2022). https://doi.org/10.1109/TEMSCONEUROPE54743.2022.9801962
28. Liu, H., Peng, H., Song, X., Xu, C., Zhang, M.: Using ai chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. Internet Interventions 27, 100495 (2022) https://doi.org/10.1016/j.invent.2022.100495
29. Burton, C., Tatar, A.S., McKinstry, B., Matheson, C., Matu, S., Moldovan, R., Macnab, M., Farrow, E., David, D., Pagliari, C., Blanco, A.S., Wolters, M.: Help4Mood Consortium: Pilot randomised controlled trial of help4mood, an embodied virtual agent-based system to support treatment of depression. J. Telemed. Telecare **22**(6), 348–355 (2016). https://doi.org/10.1177/1357633X15609793. (**PMID: 26453910**)
30. Fitzpatrick, K.K., Darcy, A., Vierhile, M.: Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial. JMIR Ment Health **4**(2), 19 (2017). https://doi.org/10.2196/mental.7785
31. Klos, M.C., Escoredo, M., Joerin, A., Lemos, V.N., Rauws, M., Bunge, E.L.: Artificial intelligence-based chatbot for anxiety and depression in university students: Pilot randomized controlled trial. JMIR Form Res **5**(8), 20678 (2021). https://doi.org/10.2196/20678
32. Inkster, B., Sarda, S., Subramanian, V.: An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth **6**(11), 12106 (2018). https://doi.org/10.2196/12106
33. Sturgill, R., Martinasek, M., Schmidt, T., Goyal, R.: A novel artificial intelligence-powered emotional intelligence and mindfulness app (ajivar) for the college student population during the covid-19 pandemic: Quantitative questionnaire study. JMIR Form Res **5**(1), 25372 (2021). https://doi.org/10.2196/25372
34. Prochaska, J.J., Vogel, E.A., Chieng, A., Baiocchi, M., Maglalang, D.D., Pajarito, S., Weingardt, K.R., Darcy, A., Robinson, A.: A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the covid-19 pandemic. Drug and Alcohol Dependence 227, 108986 (2021) https://doi.org/10.1016/j.drugalcdep.2021.108986
35. Darcy, A., Daniels, J., Salinger, D., Wicks, P., Robinson, A.: Evidence of human-level bonds established with a digital conversational agent: Cross-sectional, retrospective observational study. JMIR Form Res **5**(5), 27868 (2021). https://doi.org/10.2196/27868

36. Berg, S.: "Nudge theory" explored to boost medication adherence (2018). https://www.ama-assn.org/delivering-care/patient-support-advocacy/nudge-theory-explored-boost-medication-adherence Accessed 27-07-2023
37. Hussain, A., Malik, A., Halim, M.U., Ali, A.M.: The use of robotics in surgery: a review. International Journal of Clinical Practice 68(11), 1376–1382 (2014) https://doi.org/10.1111/ijcp.12492 https://onlinelibrary.wiley.com/doi/pdf/10.1111/ijcp.12492
38. Utermohlen, K.: Robotic Process Automation (RPA) Applications in the Healthcare Industry (2018). https://medium.com/@karl.utermohlen/4-robotic-process-automation-rpa-applications-in-the-healthcare-industry-4d449b24b613 Accessed 27-07-2023
39. Leventhal, R.: How Natural Language Processing is Helping to Revitalize Physician Documentation (2017). https://www.hcinnovationgroup.com/policy-value-based-care/article/13029202/how-natural-language-processing-is-helping-to-revitalize-physician-documentation Accessed Accessed 27-07-2023
40. Saria, S.: A $ 3 trillion challenge to computational scientists: Transforming healthcare delivery. IEEE Intell. Syst. **29**(04), 82–87 (2014). https://doi.org/10.1109/MIS.2014.58
41. Huang, J., Jennings, N.R., Fox, J.: Agent-based approach to health care management. Appl. Artif. Intell. **9**(4), 401–420 (1995). https://doi.org/10.1080/08839519508945482
42. Whitelaw, S., Mamas, M.A., Topol, E., Spall, H.G.C.V.: Applications of digital technology in covid-19 pandemic planning and response. The Lancet Digital Health 2, 435–440 (2020) https://doi.org/10.1016/S2589-7500(20)30142-4
43. Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S.: Dissecting racial bias in an algorithm used to manage the health of populations. Science 366(6464), 447–453 (2019) https://doi.org/10.1126/science.aax2342 https://www.science.org/doi/pdf/10.1126/science.aax2342
44. Kapadiya, K., Patel, U., Gupta, R., Alshehri, M.D., Tanwar, S., Sharma, G., Bokoro, P.N.: Blockchain and ai-empowered healthcare insurance fraud detection: an analysis, architecture, and future prospects. IEEE Access 10, 79606–79627 (2022) https://doi.org/10.1109/ACCESS.2022.3194569
45. Tucci, V., Saary, J., Doyle, T.E.: Factors influencing trust in medical artificial intelligence for healthcare professionals: a narrative review. Journal of Medical Artificial Intelligence 5, 4–4 (2022) https://doi.org/10.21037/jmai-21-25
46. Jobin, A., Ienca, M., Vayena, E.: The global landscape of ai ethics guidelines. Nature Machine Intelligence 1, 389–399 (2019) https://doi.org/10.1038/s42256-019-0088-2
47. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E.: Ai4people-an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. Minds and Machines 28, 689–707 (2018) https://doi.org/10.1007/s11023-018-9482-5
48. European Commission and Directorate-General for Communications Networks, Content and Technology: Ethics Guidelines for Trustworthy AI. Publications Office, Brussels (2019). https://data.europa.eu/doi/10.2759/346720
49. Kriebitz, A., Lütge, C.: Artificial intelligence and human rights: A business ethical assessment. Business and Human Rights Journal **5**(1), 84–104 (2020). https://doi.org/10.1017/bhj.2019.28
50. Pendse, S.R., Nkemelu, D., Bidwell, N.J., Jadhav, S., Pathare, S., De Choudhury, M., Kumar, N.: From treatment to healing:envisioning a decolonial digital mental health. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3491102.3501982
51. Birhane, A.: Algorithmic injustice: a relational ethics approach. Patterns 2, 100205 (2021) https://doi.org/10.1016/j.patter.2021.100205
52. Amugongo, L.M., Bidwell, N.J., Corrigan, C.C.: Invigorating ubuntu ethics in ai for healthcare: Enabling equitable care. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. FAccT '23, pp. 583–592. Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3593013.3594024
53. Viljoen, S.: A relational theory of data governance. Yale Law J. **131**, 573–654 (2021)
54. Hooker, S.: Moving beyond "algorithmic bias is a data problem". Patterns 2, 100241 (2021) https://doi.org/10.1016/j.patter.2021.100241
55. Pager, D., Karafin, D.: Bayesian bigot? statistical discrimination, stereotypes, and employer decision making. The Annals of the American Academy of Political and Social Science 621, 70–93 (2009). Accessed 2023-05-25
56. Meyer, A., Zverinski, D., Pfahringer, B., Kempfert, J., Kuehne, T., Sündermann, S.H., Stamm, C., Hofmann, T., Falk, V., Eickhoff, C.: Machine learning for real-time prediction of complications in critical care: a retrospective study. The Lancet Respiratory Medicine 6, 905–914 (2018) https://doi.org/10.1016/S2213-2600(18)30300-X
57. Beaulieu-Jones, B.K., Yuan, W., Brat, G.A., Beam, A.L., Weber, G., Ruffin, M., Kohane, I.S.: Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? npj Digital Medicine 4, 62 (2021) https://doi.org/10.1038/s41746-021-00426-3
58. Kim, S., Kim, W., Park, R.W.: A comparison of intensive care unit mortality prediction models through the use of data mining techniques. Healthc Inform Res 17(4), 232–243 (2011) https://doi.org/10.4258/hir.2011.17.4.232 http://www.e-hir.org/journal/view.php?number=599
59. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences 113(15), 3932–3937 (2016) https://doi.org/10.1073/pnas.1517384113 https://www.pnas.org/doi/pdf/10.1073/pnas.1517384113
60. Metz, T.: An african theory of social justice. In: Distributive Justice Debates in Political and Social Thought: Perspectives on Finding a Fair Share, pp. 171–190. Routledge, Abingdon, Oxfordshire, UK (2016)
61. THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,: General Data Protection Regulation (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
62. Mezgec, S., Koroušić Seljak, B.: Nutrinet: A deep learning food and drink image recognition system for dietary assessment. Nutrients **9**(7) (2017) https://doi.org/10.3390/nu9070657
63. Park, H., Bharadhwaj, H., Lim, B.Y.: Hierarchical multi-task learning for healthy drink classification. In: 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2019). https://doi.org/10.1109/IJCNN.2019.8851796
64. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R.M., Trask, A., Xu, D., Baust, M., Cardoso, M.J.: The future of digital health with federated learning. npj Digital Medicine 3, 119 (2020) https://doi.org/10.1038/s41746-020-00323-1
65. Nasr, M., Shokri, R., Houmansadr, A.: Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In,: IEEE Symposium on Security and Privacy (SP). IEEE **739–753**,(2019). https://doi.org/10.1109/SP.2019.00065, (2019)

66. Wang, Z., Song, M., Zhang, Z., Song, Y., Wang, Q., Qi, H.: Beyond inferring class representatives: User-level privacy leakage from federated learning. In: IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, pp. 2512–2520. IEEE Press,(2019). https://doi.org/10.1109/INFOCOM.2019.8737416
67. Cotterrell, R.: Trusting in law: Legal and moral concepts of trust. Current Legal Problems 46, 75–95 (1993) https://doi.org/10.1093/clp/46.Part_2.75
68. Baker, M.: 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016) https://doi.org/10.1038/533452a
69. Beck, K., Grenning, J., Martin, R.C., Beedle, M., Highsmith, J., Mellor, S., Bennekum, A., Hunt, A., Schwaber, K., Cockburn, A., al.: Principles behind the Agile Manifesto. Agile Alliance (2001). https://web.archive.org/web/20100615234816/http://agilemanifesto.org/iso/en/
70. Madaio, M.A., Stark, L., Wortman Vaughan, J., Wallach, H.: Co-designing checklists to understand organizational challenges and opportunities around fairness in ai. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, pp. 1–14. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3313831.3376445
71. Arrieta, A.B., Díaz-Rodrìguez, N., Ser, J.D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F.: Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. Information Fusion 58, 82–115 (2020) https://doi.org/10.1016/j.inffus.2019.12.012
72. Xafis, V., Schaefer, G.O., Labude, M.K., Brassington, I., Ballantyne, A., Lim, H.Y., Lipworth, W., Lysaght, T., Stewart, C., Sun, S., Laurie, G.T., Tai, E.S.: An ethics framework for big data in health and research. Asian Bioethics Review 11, 227–254 (2019) https://doi.org/10.1007/s41649-019-00099-x
73. Li, Y.: Cross-cultural privacy differences. In: Knijnenburg, B.P., Page, X., Wisniewski, P., Lipford, H.R., Proferes, N., Romano, J. (eds.) Modern Socio-Technical Perspectives on Privacy, pp. 267–292. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-82786-1_12 . https://doi.org/10.1007/978-3-030-82786-1_12
74. Woodcock, J., Larsen, P.G., Bicarregui, J., Fitzgerald, J.: Formal methods: Practice and experience. ACM Comput. Surv. **41**(4) (2009) https://doi.org/10.1145/1592434.1592436
75. Bishop, C.M.: Pattern Recognition and Machine Learning **4**, 738 (2006). https://doi.org/10.1117/1.2819119www.library.wisc.edu/selectedtocs/bg0137.pdf
76. Wang, C., Wei, X., Zhou, P.: Optimize scheduling of federated learning on battery-powered mobile devices. In: 2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 212–221 (2020). https://doi.org/10.1109/IPDPS47924.2020.00031
77. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H.: Causability and explainability of artificial intelligence in medicine. WIREs Data Mining and Knowledge Discovery 9(4), 1312 (2019) https://doi.org/10.1002/widm.1312 https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1312
78. Hofmann, G., Proença, M., Degott, J., Bonnier, G., Lemkaddem, A., Lemay, M., Schorer, R., Christen, U., Knebel, J.-F., Schoettker, P.: A novel smartphone app for blood pressure measurement: a proof-of-concept study against an arterial catheter. Journal of Clinical Monitoring and Computing 37, 249–259 (2023) https://doi.org/10.1007/s10877-022-00886-2
79. Xue, V.W., Lei, P., Cho, W.C.: The potential impact of chatgpt in clinical and translational medicine. Clinical and Translational Medicine **13** (2023) https://doi.org/10.1002/ctm2.1216
80. Cost, B.: Married father commits suicide after encouragement by AI chatbot: widow (2023). https://nypost.com/2023/03/30/married-father-commits-suicide-after-encouragement-by-ai-chatbot-widow/
81. Bharade, A.: A widow is accusing an AI chatbot of being a reason her husband killed himself (2023). https://www.businessinsider.com/widow-accuses-ai-chatbot-reason-husband-kill-himself-2023-4
82. Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.: On the dangers of stochastic parrots: Can language models be too big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21, pp. 610–623. Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3442188.3445922
83. Ribeiro, M.T., Singh, S., Guestrin, C.: "why should i trust you?": Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16, pp. 1135–1144. Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2939672.2939778
84. Lundberg, S., Lee, S.-I.: A Unified Approach to Interpreting Model Predictions (2017)
85. Rao, V.N., Zhen, X., Hovsepian, K., Shen, M.: A first look: Towards explainable textvqa models via visual and textual explanations. In: NAACL 2021 Workshop on Multimodal Artificial Intelligence (2021). https://www.amazon.science/publications/a-first-look-towards-explainable-textvqa-models-via-visual-and-textual-explanations
86. Sanchez, P., Voisey, J.P., Xia, T., Watson, H.I., ONeil, A.Q., Tsaftaris, S.A.: Causal Machine Learning for Healthcare and Precision Medicine (2022)
87. Imbens, G.W., Rubin, D.B.: Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press (2015). https://doi.org/10.1017/CBO9781139025751
88. Xiang, Y., Li, S., Zhang, P.: An exploration in remote blood pressure management: Application of daily routine pattern based on mobile data in health management. Fundamental Research **2**(1), 154–165 (2022). https://doi.org/10.1016/j.fmre.2021.11.006
89. Mökander, J., Floridi, L.: Ethics-based auditing to develop trustworthy ai. Minds and Machines 31, 323–327 (2021) https://doi.org/10.1007/s11023-021-09557-8
90. Mittelstadt, B.: Principles alone cannot guarantee ethical ai. Nature Machine Intelligence 1, 501–507 (2019) https://doi.org/10.1038/s42256-019-0114-4
91. Hussain, W., Perera, H., Whittle, J., Nurwidyantoro, A., Hoda, R., Shams, R.A., Oliver, G.: Human values in software engineering: Contrasting case studies of practice. IEEE Trans. Software Eng. **48**(5), 1818–1833 (2022). https://doi.org/10.1109/TSE.2020.3038802
92. Serban, A., Blom, K., Hoos, H., Visser, J.: Adoption and effects of software engineering best practices in machine learning. In: Proceedings of the 14th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM). ESEM '20. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3382494.3410681 . https://doi.org/10.1145/3382494.3410681
93. Washizaki, H., Uchida, H., Khomh, F., Guéhéneuc, Y.-G.: Studying software engineering patterns for designing machine learning systems. In: 2019 10th International Workshop on Empirical Software Engineering in Practice (IWESEP), pp. 49–495 (2019). https://doi.org/10.1109/IWESEP49350.2019.00017
94. Nebeker, C., Bartlett Ellis, R.J., Torous, J.: Development of a decision-making checklist tool to support technology selection in digital health research. Translational Behavioral Medicine 10(4), 1004–1015 (2019) https://doi.org/10.1093/tbm/ibz074https://academic.oup.com/tbm/article-pdf/10/4/1004/33852267/ibz074.pdf
95. Ochigame, R.: The invention of "Ethical AI": how big tech manipulates academia to avoid regulation (2019). https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/

96. European Parliament: Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206