Reevaluating Human Values in Mobile Digital Health Technology for Patient Care in the Age of Artificial Intelligence

By Monica Lopez, PhD[1]

Abstract

AI is rapidly revolutionizing healthcare to diagnose, treat and manage diseases. Innovation to maximize the opportunities of artificial intelligence (AI) in healthcare has, however, created a gap between utility and responsible AI practices. Several challenges that have arisen and require determined efforts to solve include data diversification and non-biased models; model explainability and algorithmic transparency; provider and patient education on the current limitations of the technology; and overall human-centered AI considerations of ethics, fairness and human well-being. In particular, as patient-facing decision support smart systems like mobile health applications and online sources integrate with clinical decision support systems to diagnose, treat, and/or monitor patient health outside of the hospital (Sutton *et al.*, 2020), responsible AI practices become fundamental to not just improving the healthcare outcomes of all, but protecting patients from any possible harms and violation of their rights resulting from the use of AI-enabled technologies. A recent survey underscores the lack of trust a significant number of adult patients currently have in the use of AI in their personal healthcare (Tyson *et al.*, 2023). To address these concerns, ethics and governance guidance and regulatory action plans have started to emerge from the World Health Organization (WHO, 2021), the Food & Drug Administration (FDA) (FDA 2021, 2022), and the National Institute of Science and Technology (NIST, 2023). Although important, these organizations fall short in addressing the uniqueness and entirety of the patient experience and, with the exception of the FDA, suggest high-level voluntary guidelines that have yet to gain frontline traction. In this concept paper, I argue that building responsible AI in healthcare necessitates a soft law approach that supports an end-to-end integration of clearly delineated human values across the dynamic mobile AI-enabled product development lifecycle from design and development to in-the-field integration and real-time monitoring, evaluation and education for AI system integration to succeed in the long-term. Moreover, it calls for a human-centered approach in which humans unequivocally remain at the center of the AI lifecycle to ensure that AI-enabled technologies are ethical, fair and enhance the human condition.

1.0 Introduction

Artificial intelligence (AI) systems have demonstrated remarkable capabilities across a wide range of applications, from automating repetitive tasks (Manne and Kantheti, 2021) and generating new content (Cao *et al.*, 2023) to enabling groundbreaking advancements in healthcare (Bohr and Memarzadeh, 2020) and supporting a new revolution in medicine (Topol, 2019). As AI-enabled technologies proliferate, critical to the transformation of the healthcare field is the derivation and dynamic use of new and important insights from the vast amount of data generated during the daily delivery of patient care.

Chatbots, smart homes and automation technology to frequently monitor patients through the

[1] Cognitive Insights for Artificial Intelligence (CIfAI)
Contact email: monica@ciforai.com

collection and analysis of qualitative and quantitative data in real time, for example, can support identifying when a patient is offtrack, exhibiting symptoms of illness, and in need of relevant, personalized, and evidence-based guidance. Relying on machine learning and deep learning, AI-enabled mobile health applications therefore become critical tools in contributing to accurate and correct guidance in a dynamic way by developing an increasingly granular knowledge base of inputs and outputs, like questions and responses in the case of chatbots, based on user interactions over time. Common types of data smart mobile health applications offer include biometric information (e.g., heart rate, blood pressure), activity and movement data (e.g., steps walked, calories burned), medication and treatment data (e.g., medication schedules, dosages), and symptom and condition tracking (e.g., pain levels, emotional state). Along with potential benefits, however, smart digital systems also carry risks that can manifest in ways that are harmful and even hazardous to individuals and society at large. These risks stem from various factors like inflated claims of performance, systemic biases, unintended consequences, lack of system interpretability, and ethical issues.

1.1 AI-related risks as applied to healthcare

Exaggerated claims about AI system capabilities, and therefore performance, is an area of concern. As system developers, including trusted partners across pharma, race to accelerate their product's market penetration, touting AI-enablement –whether true or not, verified for safety and reliability or not– has become a competitive advantage. Independent evaluations of mobile digital health applications, for example, have revealed that many products are inaccurate, do not work as claimed, miscalculate, and/or are not empirically backed, to name a few problems (see Cortez, Cohen, and Kesselheim, 2014 for multiple examples). Earlier in 2023 the Federal Trade Commission issued a warning to businesses on the dangers of overpromising on their AI-enabled systems and engaging in AI hype (Atleson, 2023). One solution to the AI hype problem starts with due diligence by the healthcare provider. Every AI system to be supported and utilized should have at a minimum published proof-of-concept reports, complete with transparency across the AI lifecycle regarding stakeholders involved, data used, testing and validation results, and system risk management and auditing results.

Another key problem with AI systems is their susceptibility to biases. AI models learn from vast amounts of data, reflecting the societal biases present in those data. As a result, these biases can be perpetuated or even amplified, leading to unfair and discriminatory outcomes (Obermeyer *et al.*, 2019; Vyas, Eisenstein, and Jones, 2020). For example, if an AI system is trained on historical health data that favored certain demographics, it may extend those biases and prejudices when making delivery of care recommendations, reinforcing inequality and exclusion. While synthetic data, an attractive alternative to address data scarcity problems and privacy concerns has become a focus of interest, it remains a contentious topic due to its potential for bias augmentation, overgeneralization of underrepresented characteristics, and low interpretability, to name a few issues (Giuffrè and Shung, 2023). One solution to the bias problem starts at the identification and acknowledgement of systemic bias and the multiple confounding factors involved, followed by rigorous diverse data sourcing and continuous algorithmic testing to ensure, at minimum, a less biased and more accurate system.

Unintended consequences represent another peril associated with AI systems. Complex algorithms

can produce unexpected and undesirable outcomes due to their interactions with real-world scenarios and the inevitable dynamic nature of such day-to-day interactions. For instance, an AI system built and deployed to optimize delivery of care might inadvertently allocate further resources to those already with the means to access care options via the Internet and smart technology. For example, the Apple operating system iOS is known for systematically more effective mobile health applications (Sim, 2019), causing Android users and lower-income patients to not receive any of the benefits of such technology and therefore not be represented in the data necessary to improve the system's prediction capabilities for all patients. Moreover, the digital divide between low digital literacy skills (Sim, 2019) and tech-savvy 'health hackers' leading patient-led system innovation (Omer, 2016) widens the knowledge gap even more and generates further unknowns around efficiency and safety. One solution to this problem of health disparity begins with working with and understanding patients' particular socio-economic and health circumstances fundamental to supporting their healthcare needs (Ada Lovelace Institute, 2023), followed by resource allocation through the development of patient context-specific interventions to ensure parity and equity of care options available.

The lack of straightforward interpretability in AI decision-making also poses a significant risk. Deep learning models, which are at the core of AI-enabled systems, often operate as black boxes as a result of the complexity and scale of their structures, making it difficult to fully understand how they arrive at their conclusions. This lack of model interpretability can hinder the identification of errors, biases, and/or malicious behavior, and can necessitate further assessment, delaying treatment in time-critical contexts and thus increasing the risk of endangering the patient's well-being. Moreover, the absence of explainability, or a high-level explanation for all system users of how the model works, can increase the risk of blind and unverified trust in and unintentional misuse of the system. One solution to both problems is to methodically use global and local explanations, ensuring that variables driving model predictions are clinically plausible and evidence-based, and that the conjunction of variables used to provide insight behind a model's specific prediction are clearly delineated, all the while acknowledging the limitations of such explanations. Fundamental to the success of this solution is comprehensive user training in the system prior to use of the system.

Ethical concerns add yet another layer of risk to AI-enabled systems. As these systems become ubiquitous and necessitate evermore streams of data, they raise dilemmas related to privacy, autonomy, and accountability. For example, as AI-enabled products like mobile applications from external third-party vendors integrate with in-house generated systems like mobile medical applications and enter into the clinical pipeline, both protecting patient privacy and safety and delegating specific tasks to automation become even more important as medical information is dynamically exchanged between patients, physicians, and the overall care team involved. Moreover, given third-party patient data collection, storage and security with the introduction of wearable technology developed by industry partners and that which is further transformed by tech-savvy patients, due diligence becomes yet another aspect to consider in accessing AI product function and necessity, as well as performance responsibility and culpability of unintended harmful outcomes. One solution to this problem is to critically assess the tradeoff between the patient's quality of care and the necessity of the system and its weaknesses (like the lack of model transparency), including mitigation strategies considered in the event of system failure. Essentially, what is the value add of this system considering the known and unknown risks of

utilizing and are there safeguards in place to ensure long-term safety.

1.2 Opportunities and challenges to address for mobile health

As AI development advances and outpaces our capacity to fully comprehend and mitigate its potential hazards and AI-enabled mobile digital healthcare applications become increasingly integrated into medical practice, we stand at a critical juncture in the growth of healthcare. The following four intersecting characteristics stand out:

- First, technological innovation in conjunction with the availability of dynamic streams of big data is paramount to improving patient care at a level not seen before.

- Second, such a massive opportunity thereof where scaling is the end goal means it must be balanced with responsible design, development, integration and continuous monitoring in the field, keeping in mind that scaling the system's use may not be appropriate.

- Third, medicine is a safety-critical field that must put human well-being at the center of decision-making. Solutions mentioned underscore the centrality of the human's role, as both patient and health provider, to the healthcare process.

- Fourth, the increasing growth and integration of prediction models and generative AI-enabled technology and their diverse applications introduce a monumental task for both regulatory agency review and clinical care integration, which, in the absence of available capabilities and insufficient interoperable options, underscores the need to validate every system used for the highest level of performance efficacy and safety and to facilitate effective data selection and presentation within clinical workflow, respectively.

These four intersecting characteristics further highlight a central focus point: the human user, and the requirements that imposes to support. The requirements as related to the above four characteristics are as follows:

- The diversification, sourcing, and availability of data. Data is gold, but only if useful and meaningful to each and every user.

- The integrity and effectiveness of AI-enabled systems. AI offers many benefits, but only if responsibly developed and used.

- The establishment and standardization of norms to ensure consistency and interoperability. AI offers many opportunities, but only if suitably integrated within the patient experience and existing clinical workflow.

- The responsibility of developers to develop AI-enabled systems to the highest ethical standards, and users to demand thereof. AI has the potential to be a game-changer for the better, but only if there is consistent and transparent pre-market system assessment and post-market system oversight for system safety, reliability and outcome benefit for all system users.

The opportunities and challenges of mobile health options are not new, they have increasingly gained steam for at least the past decade. What is novel here is the proposal of the underlying approach needed from which to build industry ethical guidelines and best practices to ensure that opportunities are taken advantage of and challenges are met. Moreover, with the landmark release of the Executive Order (EO) on the 'Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence' by the White House (United States, Executive Office of the President [Joseph Biden] 2023), the demand for responsible AI governance is front and center as well as the reliance on the agencies and companies that have been called to action. To ensure the responsible development and use of mobile digital health applications, a human-centered approach grounded in responsiveness and adaptability to context becomes a foundational principle to adhere to for a robust solution. In this concept paper I explore how this approach offers a best soft law solution to the immediate ethical development and adoption of AI-enabled mobile digital health technology within the U.S. healthcare system.

2.0 A Human-Centered Framework

One significant way to address the risks of AI-enabled systems is to develop them in such a manner that we can confidently rely on their performance. Such concept of system reliance has introduced the now commonly used term of trustworthy AI (High-Level Expert Group on AI, 2019). Emerging research has shown that verifying claims about an AI system's ethical characteristics and performance capabilities encourages responsible design building from the start (Brundage *et al*., 2020), fosters user confidence (Bach *et al.*, 2022), and facilitates broader societal acceptance (Choung, David, and Ross, 2023). Indeed, various organizations and governments have proposed ethics guidelines for AI (e.g., OECD, 2019).

Central to a human-centered approach is the designing and developing of AI systems with a primary focus on human experiences, needs, and overall well-being (Schiff, 2020). To place humans at the center of AI design entails considering their values, preferences, and behaviors to create a technology that addresses real-world relevant contexts and enhances user outcomes. Recognizing that technology should serve as a tool to augment human capabilities rather than replace human roles (Shneiderman, 2021), emphasis lies in empathy for the user and collaboration with the user to ensure that the AI-enabled system aligns with the expectations of the human. Methods to achieve such alignment include understanding users' specific needs, engaging users in the decision-making of the system's design, development, integration and in-field monitoring, and prioritizing users' perspectives when making design editing choices across the AI lifecycle. The end result being the maximization of the system's potential impact because of its accessibility, intuitiveness and overall usability across a diversity of users.

Two questions guide the following analysis:

I. How do we best merge the strengths of AI-enabled systems like speed and data collection and storage efficiency with the strengths of being human like curiosity, learning and adaptability, and thus optimize value across all experiences?

II. From a soft law approach, how can this human-centered AI perspective inform best practices across the AI lifecycle?

2.1 Evaluating existing guidelines and standards

Most pertinent to this discussion is IEEE's Ethically Aligned Design (EAD) framework (IEEE, 2017), which is aimed at guiding the ethical development and deployment of AI and autonomous systems. At its core is a structured approach for incorporating ethical considerations into system design, implementation and operation, offering guidance across various dimensions to ensure that AI and autonomous systems align with human values, prioritize human well-being, and avoid potential harms. The five guiding general principles proposed are:

- Human rights: Systems must not infringe on human rights.
- Well-being: System design and use must prioritize metrics of well-being.
- Accountability: System designers and operators must be held responsible and accountable.
- Transparency: Systems must operate in a transparent way.
- Awareness of misuse: Systems must be designed to minimize risks of their misuse.

Central to the focus of these principles is the prioritization of human values (e.g., respect of human rights) as they consider not only the technical aspect of AI and autonomous systems (e.g., system explainability), but also their broader societal implications (e.g., effect on well-being, legal issues of culpability, and reflection on potential consequences). The outcome is a necessary expansion of the cognitive behavioral aspects of system usability in order to capture the legal and environmental aspects of embedded systems. I will refer to this framework as a holistic human-centered approach. This approach stands as uniquely fitted to the domain of healthcare, in particular the use of mobile smart digital health applications, for three reasons:

1. Patient engagement is fundamental to the personalization and continuity of care outside of the hospital setting, and therefore necessitates an intuitive user interface that does not sacrifice ease of use over quality of data collected.
    - Consistency of use and long-term patient engagement with mobile health technology is a problem (Simblett *et al*., 2018).
    - Quality of data is a primary concern of varied and inconsistent patient-generated data (Howie *et al.*, 2014; Nowell, 2019).

2. Respect of the patient as a human being with individual needs and preferences and awareness of the value of their unique and personal data, and not a mere number of many, is critical to engendering the trust needed to ensure system engagement and the resulting health benefits from such.
    - Patient consent to the use of their personal data for clinical and research purposes becomes paramount and inextricably tied to data privacy and security (Milne-Ives, van Velthoven, and Meinert, 2020).

3. Multiple systems pose particular challenges for the optimization of use on the healthcare provider's side because of the systems' distinct functional purposes as well as their various data collection and visualization methods. This highlights the need interoperability across

multiple systems to effectively integrate.

- Patient-generated data is diverse given that it is collected from a variety of sources and across the patient's history of care, opening the door to disparate data points, input from various healthcare providers, and interaction with multiple stakeholders (Howie *et al*., 2014).

A pertinent question arises: what can we do now in practice to uphold the viewpoint of a holistic human-centered approach so that AI performs as expected in a consistent manner and benefits the human user –patient and healthcare provider? A guiding set of ethical principles that have been proposed for this holistic human-centered AI approach are FIRE or fairness, integrity, resilience and explainability (Garibay *et al.*, 2023).

Specifically:

- Fairness refers to developing bias-free algorithms. Bias is inevitable, but explicitly recognizing this inevitability can lead to identifying and implementing mitigation strategies— such as wider data access, greater data diversity, and active participation of those for which the system can benefit— to intentionally lower the risk of systemically biased decision-making of overestimating or underestimating the risks associated with specific clinical outcomes. High-quality datasets become paramount.
    - Benefit to patient: not discriminated against and better health outcomes.
    - Benefit to healthcare provider: trust in providing non-biased individualized care.

- Integrity refers to data stability and algorithmic validity. Accurate and appropriate use of data and the underlying assumptions made to characterize such data are fundamental to making correct predictions. As should be espoused, assumptions need to be empirically-backed, not selected unsystematically. The patient's voice or longitudinal experience with illness and treatment, including the healthcare provider's understanding of social determinants surrounding the patient, becomes paramount.
    - Benefit to patient: not misunderstood and better health outcomes.
    - Benefit to healthcare provider: confidence in providing holistic, individualized care.

- Resilience refers to technical robustness and compliance. The world and technological advancement are dynamic, so maintaining interoperable agility, resistance against attack and environmental sustainability becomes fundamental to supporting the inevitable rapid evolution of any AI-enabled system while minimizing the carbon footprint of large-scale computing.
    - Benefit to patient: safe and secure use and better health outcomes.
    - Benefit to healthcare provider: accuracy in providing individualized care.

- Explainability refers to transparency of the algorithmic decision-making process. Understanding of the inner workings of the system must cut across from the data points utilized to the machine-learning model generated. This will allow for any system failures to be quickly identified, appropriately mitigated, and effectively communicated thereof. And as synthetic data enters the context of healthcare to deal with the challenges of data scarcity and privacy (Giuffrè and Shung, 2023), AI system transparency and model interpretability become

paramount.

- Benefit to patient: holistic understanding of care and better health outcomes.
- Benefit to healthcare provider: holistic understanding of all data points and system recommendations for providing individualized care.

FIRE principles, like the EAD framework, places the human user at the center whereby the user should at the very least not be discriminated against, not be put in harm's way, and not be left unaware of negative outcomes in the event of system malfunction.

2.2 American National Standards Institute (ANSI) / Consumer Technology Association (CTA) Standard-2090. Use of AI in healthcare - Trustworthiness

The American National Standards Institute (ANSI) has played a vital role in developing standards that address the ethical, technical, and regulatory considerations related to AI in healthcare. As soft law guidance, standards provide a flexible and adaptive framework that address ethical, technical, and regulatory challenges and can promote responsible and effective deployment of AI while simultaneously allowing for the dynamic nature of innovation to thrive. Trustworthiness in this context refers to the ability of AI systems to deliver accurate, reliable, and unbiased outcomes while preserving patient safety, privacy, and autonomy. ANSI standards provide high-level guidance to promote trustworthiness of AI systems. The consensus-driven standard identifies three expressions —human, technical and regulatory— of how trust is created and maintained (Consumer Technology Association, 2021):

1. "Human trust focuses on fostering humanistic factors that affect the creation and maintenance of trust between the developer and users. Specifically, human trust is built upon human interaction, the ability to easily explain, user experience, and levels of autonomy of the AI solution.

2. Technical trust focuses on the technical execution of the design and training of an AI system to deliver results as expected. Technical trust can also be defined by considerations for data quality and integrity including issues of bias, data security, privacy, source and access.

3. Regulatory trust is gained through compliance by industry based upon clear laws and regulations. This trust can be based upon information from regulatory agencies, federal and state laws and accreditation boards and international standardization frameworks."

Like FIRE principles and the EAD framework, the human user is the center of focus. Here, however, emphasis lies in the very concept of trust —further divided into three categories— as being the fundamental glue between technological adoption and user engagement. In other words, trust in AI technology is suggested to depend as much on a human user-to-human developer relationship built on confidence, as the need for the production of reliable, safe, and approved product and service outcomes. These identified pillars of trust thus reinforce the cognitive behavioral expectations of a declaration of assurance.

2.3 World Health Organization (WHO) Ethical Principles

Realizing the need for ethical principles to provide guidance to stakeholders on how basic moral requirements should direct or constrain their decisions and actions in the specific context of developing, deploying and assessing performance of AI technology for health, the following six ethical principles have been identified by the WHO Expert Group as the most appropriate (World Health Organization, 2021). I highlight their relevancy to a human-centered approach.

- Human agency: Protect autonomy whereby humans should remain in full control of healthcare systems and medical decisions. Respect for autonomy also entails the related duties to protect privacy and confidentiality and to ensure informed, valid consent by adopting appropriate legal frameworks for data protection. Under this principle, humans should have the final word. As reliable as systems may become, they are tools to be used not wielded to influence or coerce.
    - This is in favor of a human-centered approach in *maintaining human control over AI system functioning and decision-making*.

- Human well-being: Promote human well-being, human safety and the public interest whereby AI technologies should not harm people. They should satisfy regulatory requirements for safety, accuracy and efficacy before deployment, and measures should be in place to ensure quality control and quality improvement. Under this principle, there is no excuse for bias and discrimination. Harms should be identified and removed.
    - This is in favor of a human-centered approach in *prioritizing human well-being*.

- Human understanding: Ensure transparency, explainability and intelligibility whereby the AI should be intelligible or understandable to developers, users and regulators. Under this principle black boxes are unacceptable as model explainability and model interpretability are fundamental to understanding the efficacy of the system and improving it capabilities.
    - This is in favor of a human-centered approach in *requiring human understanding of all AI systems*.

- Human responsibility and accountability: Foster responsibility and accountability whereby humans require clear, transparent specification of the tasks that systems can perform and the conditions under which they can achieve the desired level of performance. Under this principle, inflated or false claims of performance are unacceptable as human supervision and attestation of system capacity become critical to attributing accountability, including across the multiple touch points of care (e.g., manufacturers, hospitals, clinicians).
    - This is in favor of a human-centered approach in *requiring human responsibility and accountability over all AI systems*.

- Non-biased and equitable systems: Ensure inclusiveness and equity whereby AI is designed to encourage the widest possible appropriate, equitable use and access, irrespective of age, gender, income, ability or other characteristics. Under this principle, technology is for all users and their benefit. Adaptability is essential, allowing for necessary accommodations to be adopted.
    - This is in favor of a human-centered approach in *ensuring the non-discrimination*

and benefit of all humans through the designing, use, and implementation of non-biased and equitable AI systems.

- System effectiveness and reliability: Promote AI that is responsive and sustainable whereby designers, developers and users continuously, systematically and transparently examine an AI technology to determine whether it is responding adequately, appropriately and according to communicated expectations and requirements in the context in which it is used. Responsiveness also requires that AI technologies be consistent with wider efforts to promote health systems and environmental and workplace sustainability. Under this principle, AI systems should not be adopted for the sake of popularity. Rather, they should be determined relevant, necessary and effective.
    - This is in favor of a human-centered approach in *prioritizing unique human value as further supported by AI system relevancy, necessity and effectiveness*.

The WHO's six ethical principles encourage a basic moral foundation built on the inviolable right of human dignity that can support the development of consistent ethical practices across borders. As healthcare is a global issue, international collaboration becomes a next step of action to establish harmonized guidelines for the development, use and implementation of AI that also respects regional and cultural differences.

3.0 Policy Initiatives in the U.S.

One of the most human-centered frameworks published in the U.S. to date is the White House Office of Science and Technology Policy's (OSTP) *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* (OSTP, 2022). While a human-centered approach is not explicitly mentioned, the blueprint identifies five principles and practices to guide the design, use and deployment of AI as a way to meaningfully and most beneficially impact the U.S. American public's rights, opportunities, and access to critical needs. The five principles include maintaining the safety and effectiveness of AI-enabled systems, as well as independent audits to confirm thereof and publication of such; designing equitable systems and using systems equitably to mitigate against unfairness and thus prevent discrimination; protecting user privacy through limitations on data collection; providing transparency to users on the use of an automated system, as well as explanations of outcomes; and empowering users with the choice to opt out or to interact with a human as an alternative.

Regarding the healthcare domain specifically, the U.S. Food and Drug Administration (FDA) plays a crucial role in regulating medical devices and ensuring their safety and effectiveness. The FDA has taken significant steps since 2017 to address the regulation of machine learning-based medical devices and digital health applications (e.g., US FDA 2021, 2022, 2023) to tackle the unique challenges AI systems inherently pose. Efforts support the FDA's recognition that ensuring the safety and effectiveness of these technologies is equally important as fostering innovation in the healthcare industry. Critical to the management of potential risks resulting from the development of AI systems has been the need for including diverse perspectives across the AI lifecycle, as well as continuous monitoring of system performance in real-world contexts to identify and address potential risks and ensure the system's ongoing safety and effectiveness.

The primary concern of the FDA is to protect public health through:

- The effectiveness and safety of AI-enabled medical devices,
- consistency of evaluation criteria,
- elimination of uncertainties and thoroughly reviewed medical devices,
- effective and safe integration of new AI system advancements into medical devices,
- high quality and representative-of-the-intended-use-population data used for training and validation, including the respect of privacy and confidentiality,
- clear labeling and documentation requirements, monitoring devices for safety and effectiveness through post-market surveillance, and
- alignment among regulatory bodies.

To gather input and insights into the regulation of AI and machine learning systems in healthcare from a diversity of perspectives, the FDA has engaged with healthcare professionals and stakeholders from industry and academia. Moreover, the FDA has emphasized the importance of real-world performance monitoring. Multiple actions have thus been taken. I highlight their relevancy to a human-centered approach.

- Digital Health Pre-Certification Pilot Program (Pre-Cert): In 2017, the FDA introduced the Digital Health Pre-Certification Pilot Program to streamline the regulatory process for digital health technologies, including machine learning-based medical devices. The program focused on evaluating the developer's organizational excellence, product quality, and post-market performance tracking. The goal was to prioritize oversight based on the software developer's demonstrated commitment to producing safe and effective products. It was completed September 2022 (US FDA, 2023). While the safety and effectiveness of the product was underscored, the following points would more robustly support a human-centered approach moving forward:
    - Identification of the cutoff point where product safety and effectiveness are not achieved, with specific redress strategies in place.
    - Clarification of the types of harms predicted and assessed beyond probability of occurrence and degree of severity, including how they will be determined and handled.
    - Inclusion of explainability of the factors and logic that lead to an outcome from use of the product in addition to transparency of the product's intended use, known limitations and hypothesized constraints, user interface interpretation, and clinical workflow integration.
    - Specification of methods and metrics used to obtain user feedback on product quality and performance, and alternative methods and metrics additionally used, if applicable.

- Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning-Based Software as a Medical Device (SaMD), defined as software intended for a medical purpose without being part of a hardware medical device (IMDRF SaMD Working Group, 2013): In April 2019, the FDA released a discussion paper outlining a potential regulatory framework for AI and machine learning- based Software as a Medical Device for use in clinical settings. The framework included considerations for premarket review, post-

market surveillance, and ongoing modifications to AI algorithms to ensure safety, effectiveness, and transparency. While the paper underscored support in facilitating a rapid cycle of product improvement while providing effective safeguards, the following points would more robustly support a human-centered approach moving forward:

- Clarification on the definition of "improving performance" in the context of modifications made to an AI/ML-based SaMD.
- Clarification on the criterion used for transparency of AI/ML-based SaMD and real-world performance monitoring of AI/ML-based SaMD.
- Transparency on the criterion used for the FDA's Pre-Cert TPLC (total product life cycle) approach across all sizes, types, and statuses of organizations, considering that those features affect system development and integration.

• AI/Machine Learning-Based Software as a Medical Device Action Plan: In January 2021, the FDA unveiled an action plan for AI and machine learning-based systems. The plan, which was an update from the FDA discussion paper released in April 2019, included five actions to address the unique challenges posed by these systems. They included further development of the proposed regulatory framework along with draft guidance on a predetermined change control plan to take into consideration the software's learning over time; support of good machine learning practices; support of a patient-centric approach; development of methods for evaluation and improvement of machine learning algorithms; and advancement of pilots for real-world performance monitoring.

• Digital Health Policy Navigator: In December 2022, the FDA developed a tool to help product developers determine whether their product's software functions are potentially the focus of the FDA's oversight.

In summary, the FDA's role in regulating machine learning-based medical devices and digital health applications is to ensure patient safety, device effectiveness, and the responsible integration of these technologies into the healthcare ecosystem. Through their oversight, they provide a framework that encourages innovation while maintaining high standards of quality and safety.

Given the distinctive nature of mobile digital health technology and the need to build these systems with the recognition that they owe a duty of care, with manufacturer responsibility, to patients, the following recommendations are proposed for the following two user groups: healthcare providers and patients. For healthcare providers, the demands most pertinent to establishing trust and providing the needed confidence to use AI systems as a complementary tool to their expertise and human capabilities are:

1. The development of system transparency requirements and the standardization of and education on such. Transparency must cover the entire gamut of concepts usually associated with explainability including model interpretability, data set visibility, model capabilities and limitations, and contextual usability.

2. The creation of a registry of systems to make cataloguing them simple and accessible to all, including to report errors and misdiagnoses and impact assessments of harm (with a further highlight on the operational definition requirements that entails). This is inspired

by the third-party databases that track FDA-approved AI algorithms (Benjamens, Dhunnoo, and Meskó, 2020; American College of Radiology Data Science Institute, 2021).

3. The development of risk management requirements, acknowledging that NIST's AI Risk Management Framework (RMF) can be specifically tailored to healthcare. This is of noteworthy importance given U.S. President Biden's EO that significantly expands NIST's responsibilities related to responsible AI development. Relevant responsibilities include the development of guidelines and best practices for consensus-driven industry standards to ensure safe AI systems, the creation of companion resources for the AI RMF and Secure Software Development Framework, the initiation of efforts to provide guidance and benchmarks for auditing potentially harmful AI capabilities, and the identification of standards and techniques for content authentication, tracking provenance, labeling synthetic content and detecting synthetic content, among other areas.

4. The establishment of system auditing requirements, including the publication of report results that are interpretable to all.

5. The interoperable management of the rich, unstructured troves of data gathered by the patient from these mobile digital devices. This becomes a requirement for system transparency, risk management, and auditing practices.

6. The integration of privacy protections of personal health data that fall outside the ambit of the Health Insurance Portability and Accountability Act or HIPAA. A notable example is The Washington My Health My Data Act (HB 1155) 2023 meant to protect consumers' sensitive health data from being collected and shared without their consent. This aids with maintenance of data security and system robustness against adversarial attack.

7. The creation of an expert oversight board for system checks and balances. This is inspired by the role institutional review boards have to ensure the protection of human rights and the well-being of research subjects.

For the patient user of an AI-enabled system, the following recommendations are complementary to the healthcare provider's needs:

1. The notification of the use of an AI-enabled system in their care.

2. The option, when possible, to opt-in or opt-out of AI system use. This is particularly critical when dealing with smartphones, tablets, and wearable devices.

3. The inclusion of system transparency and explainability. The functionality of new visualization methods becomes increasingly pertinent and highlights just how interrelated it is with demands for transparency and auditing and impact assessment reports.

4. The development of ease of usability of the system, considering the diverse needs of the end-user.

5. The retainment of agency over self-gathered data and human-in-the-loop functionality. This entails robust end-user system education.

6. The development of a traceable mechanism for the entrepreneurial efforts of tech savvy and avid amateur "doctor" patients who tinker with applications and build better-grade applications out of the purview of oversight yet necessitate safety measures in place.

7. The integration of local governments oversight on health application operations responsive to local conditions, including testing plan of equity measures across diverse patient populations.

8. The development of appropriate capacity building to support all of the above, including upskilling in data science and machine learning.

These recommendations at the core stand on one major point: that there should be cause to pause for a moment to determine what we should automate, and when we should use an AI system, because data is gold only when it is meaningful to us. Moreover, the question remains of what we should not automate, and when we should not use an AI system. Given the known harms and risks, as builders and users of the technology, we have a responsibility to know how the technology works and to demand transparency. We humans must retain control and know our purpose, ensuring that the AI system is in the loop (Lynch, 2022). I therefore underscore that any soft law trending towards best practices as created by industry and healthcare organizations must be fundamentally guided by human-centered AI principles.

References

Ada Lovelace Institute (2023). Access denied? Socioeconomic inequalities in digital health services, https://www.adalovelaceinstitute.org/report/healthcare-access-denied/

American College of Radiology Data Science Institute. (February 1, 2021). New ACR DSI Searchable FDA-Cleared Algorithm Catalog Can Ease Medical Imaging AI Integration. https://www.acrdsi.org/News-and-Events/New-ACR-DSI-Searchable-FDA-Cleared-Algorithm-Catalog-Can-Ease-Medical-Imaging-AI-Integration

Atleson, M. (2023). Keep Your AI Claims in Check. Federal Trade Commission Business Guidance Blog, https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., and Sousa, S. (2022). A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2022.2138826

Benjamens, S., Dhunnoo, P., Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digital Medicine*, *3* (1):118. doi: 10.1038/s41746-020-00324-0.

Bohr, A., and Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. In *Artificial Intelligence in healthcare* (pp. 25-60). Academic Press.

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., ... and Anderljung, M. (2020). Toward trustworthy AI development: mechanisms for supporting verifiable claims. *arXiv preprint arXiv:2004.07213*.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., and Sun, L. (2023). A comprehensive survey of AI-generated content (aigc): A history of generative ai from GAN to ChatGPT. *arXiv preprint arXiv:2303.04226*. https://doi.org/10.48550/arXiv.2303.04226

Choung, H., David, P., and Ross, A. (2023). Trust and ethics in AI. *AI & Society*, *38* (2), pp. 733-745.

Cortez, N. G., Cohen, I. G., and Kesselheim, A. S. (2014). FDA Regulation of Mobile Health Technologies. *The New England Journal of Medicine, 371*, pp. 372-379. DOI: 10.1056/NEJMhle1403384

Garibay, O. O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., Grieman, K., Havens, J. C., Jirotka, M., Kacorri, H., Karwowski, W., Kider, J., Koon, S., Lopez-Gonzalez, M., Maifeld-Carucci, I., McGregor, S., Salvendy, G., Shneiderman, B., Stephanidis, C., Strobel, C., Holter, C. T., Xu, W. (2023). Six Human-Centered Artificial Intelligence Grand Challenges. *International Journal of Human–Computer Interaction, 39* (3), pp. 391-437. DOI: 10.1080/10447318.2022.2153320

Giuffrè, M. and Shung, D. L. (2023). Harnessing The Power of Synthetic Data in Healthcare: Innovation, Application, and Privacy. *njp Digital Medicine, 6* (186). https://doi.org/10.1038/s41746-023-00927-3

High-Level Expert Group on AI. (2019). Ethics guidelines for trustworthy artificial intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Howie, L., Hirsch, B., Locklear, T., and Abernethy, A. P. (2014). Assessing The Value of patient-Generated Data To Comparative Effectiveness Research. *Health Affairs, 33* (7), pp. 1220-1228. https://doi.org/10.1377/hlthaff.2014.0225

IMDRF SaMD Working Group. (December 9, 2013). Software as a Medical Device (SaMD): Key Definitions, https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf

Lynch, S. (October 17, 2022). AI in the Loop: Humans must remain in charge. News, Stanford University Human-Centered Artificial Intelligence. https://hai.stanford.edu/news/ai-loop-humans-must-remain-charge

Milne-Ives, M, van Velthoven, M. H., and Meinert, E. (2020). Mobile apps for real-world evidence in health care. *Journal of the American Medical Informatics Association, 27* (6), pp. 976-980. doi: 10.1093/jamia/ocaa036. PMID: 32374376; PMCID: PMC7647309.

Nowell, W. B. (2019). Information Patients Can Provide Will Strengthen the Real-World Evidence That Matters to Them. *Clinical Pharmacology & Therapeutics, 106* (1), pp. 49-51. doi:10.1002/cpt.1460

Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science 366* (6464), pp. 447-453. DOI:10.1126/science.aax2342

Omer, T. (August 2016). Empowered Citizen 'Health Hackers' Who Are Not Waiting. *BMC Medicine 14*(1), 118. DOI: 10.1186/s12916-016-0670-y. PMID: 27530970; PMCID: PMC4988004

OSTP. (October 2022). Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Schiff, D., Ayesh, A., Musikanski, L., and Havens, J. C. (2020, October). IEEE 7010: A new standard for assessing the well-being implications of artificial intelligence. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* pp. 2746-2753. IEEE.

Schneiderman, B. (Winter 2021). Human-Centered AI. *Issues in Science and Technology 37* (2), pp. 56-61. https://issues.org/human-centered-ai/

Sim, I. (2019). Mobile Devices and Health. *The New England Journal of Medicine 381*, pp. 956-968. DOI: 10.1056/NEJMra1806949

Simblett, S., Greer, B., Matcham, F., Curtis, H., Polhemus, A., Ferrão, J., Gamble, P. and Wykes, T. (2018). Barriers to and Facilitators of Engagement with Remote Measurement Technology for Managing Health: Systematic Review and Content Analysis of Findings. *Journal of Medical Internet Research 20* (7), e10480, pp. 1-13. http://www.jmir.org/2018/7/e10480/

Sutton, R.T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N. and Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digital Medicine 3* (17). https://doi.org/10.1038/s41746-020-0221-y

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine 25*, pp. 44–56. https://doi.org/10.1038/s41591-018-0300-7

US FDA. (January 12, 2021). FDA releases artificial intelligence/machine learning action plan. https://www.fda.gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan

US FDA. (September 2022). Policy for device software functions and mobile medical applications. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications

US Food and Drug Administration. (December 14, 2022). Digital Health Policy Navigator. https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-policy-navigator#:~:text=If%20you%20are%20developing%20a,to%20FDA%27s%20oversight%20as%20devices

US Food and Drug Administration. (2023). The Software Precertification (Pre-Cert) Pilot Program: Tailored Total Product Lifecycle Approaches and Key Findings. https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-pilot-program

United States, Executive Office of the President [Joseph Biden]. Executive order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 30 Oct. 2023. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Vyas, D. A., Eisenstein, L. G., and Jones, D. S. (2020). Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine*, *383* (9), 874-882.

Washington My Health My Data Act, Washington Laws 191 (2023). https://app.leg.wa.gov/billsummary?BillNumber=1155&Initiative=false&Year=2023