Reevaluating Human Values for Patient Care in the Age of Artificial Intelligence:

A Human-Centred Approach to Mobile Digital Health Technology Regulation in the United States

*Monica Lopez**

Artificial Intelligence (AI)-based systems are rapidly revolutionising the process of healthcare delivery, introducing both opportunities and challenges. Innovation in the United States (US) to maximise the opportunities of improved diagnosis, treatment and management of disease has, however, created a gap between utility and responsible AI practices. There are several challenges that require determined efforts to solve, including data diversification and non-biased models; model explainability and algorithmic transparency; healthcare provider and patient education of AI systems; and overall human-centred considerations of ethics, fairness and human well-being. In particular, as patient-facing decision support smart systems like mobile digital applications and online sources integrate with clinical decision support systems, 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 such technologies. In this article, I argue that building responsible mobile digital AI-based health technologies across the US healthcare pipeline necessitates a revision of human values, ie one grounded in a human-centred framework whereby humans unequivocally remain at the centre of the AI lifecycle, to underpin a regulatory approach that augments both current broad US initiatives in AI ethics development and integration, and more specific Food and Drug Administration efforts.

Keywords: human-centred AI; responsible AI; healthcare; mobile digital health technology

I. Introduction

Artificial intelligence (AI) systems have demonstrated remarkable capabilities across a wide range of applications, from automating repetitive tasks[1] and generating new content[2] to enabling groundbreaking advancements in healthcare[3] and supporting a new revolution in medicine.[4] 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.

DOI: 10.21552/aire/2024/1/7

* Dr Monica Lopez, Co-Founder and CEO, Cognitive Insights for Artificial Intelligence, USA. For correspondence: <monica@ciforai.com>.

1 R Manne and SC Kantheti, 'Application Of Artificial Intelligence In Healthcare: Chances And Challenges' (2021) 40(6) Current Journal of Applied Science and Technology 78.

2 Y Cao et al, 'A Comprehensive Survey Of AI-Generated Content (aigc): A History Of Generative AI From GAN To ChatGPT' (2023) arXiv preprint arXiv:2303.04226.

3 A Bohr and K Memarzadeh, 'The Rise of Artificial Intelligence in Healthcare Applications' in *Artificial Intelligence in Healthcare* (Academic Press 2020) 25-60.

4 EJ Topol, 'High-performance medicine: the convergence of human and artificial intelligence' (2019) 25 Nature Medicine 44.

Chatbots, smart homes and automation technology to frequently monitor patients through the 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, personalised, and evidence-based guidance. Relying on machine learning (ML) and deep learning (DL), 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 (eg, heart rate, blood pressure), activity and movement data (eg, steps walked, calories burned), medication and treatment data (eg, medication schedules, dosages), and symptom and condition tracking (eg, pain levels, emotional state). Along with potential benefits, however, smart digital AI 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. Guided by an approach focused on identifying the gaps between AI system adoption and quality care delivery in the US, this article presents a holistic human-centred solution to the development and use of digital mobile health technology. Given the dynamic nature of AI systems, such a solution stems from the need to unequivocally place human well-being, both the patient and healthcare provider, at the core of every touch point across the US healthcare delivery timeline and suggests necessary changes within US regulation to yield success.

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.[5] Earlier in 2023 the Federal Trade Commission (FTC) issued a warning to businesses on the dangers of overpromising on their AI-enabled systems and engaging in AI hype,[6] as well as giving notice to hundreds to not make deceptive product claims.[7] Various health application developers have already paid settlements to the FTC for making false claims about their applications' performance capabilities.[8] One solution to the AI hype problem starts with due diligence by the healthcare provider (in the absence of application developer due diligence premarket development and post-market integration). Every AI system to be supported and utilised 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.[9] For example, if an AI system

5 For multiple examples, see NG Cortez, IG Cohen and AS Kesselheim, 'FDA Regulation of Mobile Health Technologies' (2014) 371 The New England Journal of Medicine 372.

6 M Atleson, 'Keep Your AI Claims in Check' (2023) Federal Trade Commission Business Guidance Blog <https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check>. All internet links were last accessed 11 March 2024.

7 Federal Trade Commission, 'FTC Warns Almost 700 Marketing Companies That They Could Face Civil Penalties if They Can't Back Up Their Product Claims' (Federal Trade Commission press release, April 2023) <https://www.ftc.gov/news-events/news/press-releases/2023/04/ftc-warns-almost-700-marketing-companies-they-could-face-civil-penalties-if-they-cant-back-their?ref=upstract.com>.

8 Federal Trade Commission, 'FTC Charges Marketers of "Vision Improvement" App with Deceptive Claims' (Federal Trade Commission press release, September 2015) <https://www.ftc.gov/news-events/news/press-releases/2015/09/ftc-charges-marketers-vision-improvement-app-deceptive-claims>. Federal Trade Commission, 'Marketers of Blood-Pressure App Settle FTC Charges Regarding Accuracy of App Readings' (Federal Trade Commission press release, December 2016) <https://www.ftc.gov/news-events/news/press-releases/2016/12/marketers-blood-pressure-app-settle-ftc-charges-regarding-accuracy-app-readings>. J Wagner, 'The Federal Trade Commission and Consumer Protections for Mobile Health Apps' (2020) 48(1Suppl) Journal of Law and Medical Ethics 103.

9 Z Obermeyer et al, 'Dissecting Racial Bias in An Algorithm Used to Manage the Health of Populations' (2019) 366(6464) Science 447. DA Vyas, LG Eisenstein and DS Jones, 'Hidden in Plain Sight—Reconsidering the Use of Race Correction in Clinical Algorithms' (2020) 383(9) The New England Journal of Medicine 874.

is trained on historical health data that favoured 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, overgeneralisation of underrepresented characteristics, and low interpretability, to name a few issues.[10] Furthermore, identifying flaws in generated synthetic data sets using large language models, for example, require close and careful scrutiny.[11] 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 optimise 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,[12] causing Android users and lower-income patients who do not use Apple systems 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[13] and tech-savvy 'health hackers' leading patient-led system innovation[14] 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,[15] 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 behaviour, 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 built and regulated 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,

10 M Giuffrè and DL Shung, 'Harnessing The Power of Synthetic Data in Healthcare: Innovation, Application, and Privacy' (2023) 6(186) njp Digital Medicine.

11 A Taloni, V Scorcia and G Giannaccare, 'Large Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research' (9 November 2023) JAMA Ophthalmology.

12 I Sim, 'Mobile Devices and Health' (2019) 381 The New England Journal of Medicine 956.

13 Ibid.

14 T Omer, 'Empowered Citizen 'Health Hackers' Who Are Not Waiting' (August 2016) 14(1) BMC Medicine 118.

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

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 trade-off 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 utilising and are there safeguards in place to ensure long-term safety.

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 centre of decision-making. Solutions mentioned above 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. The requirements as related to the above four characteristics are as follows:

- The diversification, sourcing, and availability of data. Data is equivalent to value, 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 standardisation of norms to ensure consistency and interoperability. AI offers many opportunities, but only if suitably integrated within the patient experience and existing clinical workflow so that both patients and clinicians can benefit.
- 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 inform regulatory requirements so 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 US White House,[16] the demand for responsible AI governance is front and centre as well as the reliance on the agencies and companies that

16 United States, Executive Office of the President [Joseph Biden], 'Executive order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence' (30 October 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/>.

have been called to action. To ensure the responsible development and use of mobile digital health applications, a human-centred approach grounded in responsiveness and adaptability to context becomes a foundational principle to adhere to for a robust solution. In this article I explore how this approach offers a best solution to the immediate ethical development and adoption of AI-enabled mobile digital health technology within the US healthcare system.

II. A Human-Centred 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.[17] Emerging research has shown that verifying claims about an AI system's ethical characteristics and performance capabilities encourages responsible design building from the start,[18] fosters user confidence,[19] and facilitates broader societal acceptance.[20] Indeed, various organisations and governments have proposed ethics guidelines for AI.[21]

Central to a human-centred approach is the designing and developing of AI systems with a primary focus on human experiences, needs, and overall well-being.[22] To place humans at the centre of AI design entails considering their values, preferences, and behaviours to create a technology that addresses real-world relevant contexts and enhances user outcomes. Recognising that technology should serve as a tool to augment human capabilities rather than replace human roles,[23] 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 prioritising users' perspectives when making design editing choices across the AI lifecycle. The end result being the maximisation 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:

– 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 optimise value across all experiences?
– From a regulatory perspective, how can this human-centred AI approach inform best practices and ultimately necessary regulatory requirements across the AI lifecycle?

1. Evaluating Existing Guidelines and Standards

Most pertinent to this discussion is the Institute of Electrical and Electronics Engineers' (IEEE) Ethically Aligned Design (EAD) framework,[24] 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, prioritise human well-being, and avoid potential harms. The five guiding general principles proposed are:

1. *Human rights*: systems must not infringe on human rights.
2. *Well-being*: system design and use must prioritise metrics of well-being.
3. *Accountability*: system designers and operators must be held responsible and accountable.

17 High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy Artificial Intelligence' (2019) <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

18 M Brundage et al, 'Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims' (2020) arXiv preprint arXiv:2004.07213.

19 TA Bach et al, 'A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective' (2022) International Journal of Human–Computer Interaction.

20 H Choung, P David and A Ross, 'Trust and Ethics in AI' (2023) 38(2) AI & Society 733.

21 Eg, OECD, 'OECD AI Principles Overview' (OECD.AI Policy Observatory, May 2019) <https://oecd.ai/en/ai-principles>.

22 D Schiff et al, 'IEEE 7010: A New Standard for Assessing the Well-Being Implications of Artificial Intelligence' (2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC) 2746-2753 IEEE, October 2020).

23 B Schneiderman, 'Human-Centered AI' (Winter 2021) 37(2) Issues in Science and Technology 56.

24 The IEEE Global Initiative, 'Ethically Aligned Design' (2017) The Institute of Electrical and Electronics Engineers (IEEE) <https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf>.

4. *Transparency*: systems must operate in a transparent way.
5. *Awareness of misuse*: systems must be designed to minimise risks of their misuse.

Central to the focus of these principles is the prioritisation of human values (eg, respect of human rights) as they consider not only the technical aspect of AI and autonomous systems (eg, system explainability), but also their broader societal implications (eg, effect on well-being, legal issues of culpability, and reflection on potential consequences). The outcome is a necessary expansion of the cognitive behavioural 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-centred 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 personalisation 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 digital mobile health technology is a problem.[25]
   – Quality of data is a primary concern of varied and inconsistent patient-generated data.[26,27]
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.[28]
3. Multiple systems pose particular challenges for the optimisation of use on the healthcare provider's side because of the systems' distinct functional purposes as well as their various data collection and visualisation methods. This highlights the need for 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.[29]

A pertinent question arises: what can we do now in practice to uphold the viewpoint of a holistic human-centred 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-centred AI approach are FIRE or fairness, integrity, resilience and explainability.[30]

Specifically, as applied to healthcare:

– 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 inclusion 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 individualised care.
– Integrity refers to data stability and algorithmic validity. Accurate and appropriate acquisition and use of data, and the underlying assumptions made to characterise 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 lon-

25 S Simblett et al, 'Barriers to and Facilitators of Engagement with Remote Measurement Technology for Managing Health: Systematic Review and Content Analysis of Findings' (2018) 20(7) Journal of Medical Internet Research e10480.

26 L Howie et al, 'Assessing The Value of Patient-Generated Data to Comparative Effectiveness Research' (2014) 33(7) Health Affairs 1220.

27 W B Nowell, 'Information Patients Can Provide Will Strengthen the Real-World Evidence That Matters to Them' (2019) 106(1) Clinical Pharmacology & Therapeutics 49.

28 M Milne-Ives, MH van Velthoven and E Meinert, 'Mobile Apps for Real-World Evidence in Healthcare' (2020) 27(6) Journal of the American Medical Informatics Association 976.

29 Howie et al (n 26).

30 OO Garibay et al, 'Six Human-Centered Artificial Intelligence Grand Challenges' (2023) 39(3) International Journal of Human-Computer Interaction 391.

gitudinal experience with illness and treatment, including the healthcare provider's understanding of social determinants surrounding the patient,[31] becomes paramount. Moreover, any integration of synthetically AI-generated data necessitates rigorous scrutiny to ensure the plausibility of such data given known cases of fake AI-generated data sets.[32]
  - Benefit to patient: not misunderstood and better health outcomes.
  - Benefit to healthcare provider: confidence in providing holistic, individualised care.
- Resilience refers to technical robustness and compliance. The world and technological advancement are dynamic, so maintaining interoperable agility across systems and data integration, resistance against attack and environmental sustainability becomes fundamental to supporting the inevitably rapid evolution of any AI-enabled system while minimising the carbon footprint of large-scale computing.
  - Benefit to patient: safe and secure use and better health outcomes.
  - Benefit to healthcare provider: confidence in using a safe system and maintaining their patient's data privacy.
- 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 utilised 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,[33] AI system transparency, data sourcing inspection, 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 optimum individualised care.

FIRE principles, like the EAD framework, places the human user at the centre 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. 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:[34]

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 standardisation frameworks.'

---

31 National Academy of Medicine; National Academies of Sciences, Engineering, and Medicine; Committee on the Future of Nursing 2020-2030; MK Wakefield et el (eds), 'The Future of Nursing 2020-2030: Charting A Path to Achieve Health Equity' (2021) <https://nap.nationalacademies.org/catalog/25982/the-future-of-nursing-2020-2030-charting-a-path-to>.

32 Taloni, Scorcia and Giannaccare (n 14).

33 Giuffrè and Shung (n 10).

34 Consumer Technology Association, 'ANSI/CTA Standard. The Use of Artificial Intelligence in Health Care: Trustworthiness' (2021) ANSI/CTA-2090.

Like FIRE principles and the EAD framework, the human user is the centre 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 behavioural expectations of a declaration of assurance.

3. World Health Organisation (WHO) Ethical Principles

Realising 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.[35] I highlight their relevancy to a human-centred approach.

- Human agency: protect autonomy whereby humans 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 sophisticated and reliable as systems may become, they are tools to be used for beneficial outcomes, not wielded to influence or coerce.
  - This is in favour of a human-centred 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-based 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 favour of a human-centred approach in *prioritising human well-being*.
- Human understanding: ensure transparency, explainability and intelligibility whereby the AI system 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 to improving its capabilities.
  - This is in favour of a human-centred 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 AI 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 (eg, manufacturers, hospitals, clinicians).
  - This is in favour of a human-centred approach in *requiring human responsibility and accountability over all AI systems*.
- Non-biased and equitable systems: ensure inclusiveness and equity whereby AI systems are 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 favour of a human-centred 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 systems that are responsive and sustainable whereby designers, developers and users continuously, systematically and transparently examine an AI-based technology to determine whether it is

35 World Health Organisation, *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance* (World Health Organisation 2021) Licence: CC BY-NC-SA 3.0 IGO.

responding adequately, appropriately and according to communicated expectations and requirements in the context in which it is used. Responsiveness also requires that AI-based technologies be consistent with wider efforts to promote health ecosystems 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 favour of a human-centred 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 harmonised guidelines for the development, use and implementation of AI systems that also respect regional and cultural differences.

4. The European Union Artificial Intelligence Act

Leading as the world's first piece of legislation to regulate AI and prevent its misuses, the European Union (EU) AI Act deserves mention for its relevancy to the human-centred approach espoused. Likely to come in force by summer 2024, the soon-to-be EU law is notable for its risk-based approach which identifies the level of risk posed by an AI system based on its use case. Use cases are classified as unacceptable-risk, high-risk, low-risk, or minimal-risk, and each impose a set of different rules and obligations.[36]

In the case of medical devices, for example, that are already subject to certain EU regulation, the EU AI Act provides for AI systems constituting such products to be considered high-risk AI systems in that they 'have a significant harmful impact on the health, safety and fundamental rights of persons in the Union.'[37] As such, high-risk AI providers must:
- 'Establish a **risk management system** throughout the high-risk AI system's lifecycle.
- Conduct **data governance** to ensure that training, validation and testing datasets are relevant, sufficiently representative and, to the best extent possible, free of errors and complete according to the intended purpose.
- Draw up **technical documentation** to demonstrate compliance and provide authorities with the information to assess that compliance.
- Design their high-risk AI system for **record-keeping** to enable it to automatically record events relevant for identifying national level risks and substantial modifications throughout the system's lifecycle.
- Provide **instructions for use** to downstream deployers to enable the latter's compliance.
- Design their high-risk AI system to allow deployers to implement **human oversight**.
- Design their high-risk AI system to achieve appropriate levels of **accuracy, robustness, and cybersecurity**.
- Establish a **quality management system** to ensure compliance.'[38]

Critical requirements for the systems themselves are in line with all the above ethical principles as they include risk management, data quality, transparency, human oversight, and accuracy. Moreover, businesses providing or deploying an AI system will face obligations around registration, quality management, monitoring, record-keeping, and incident reporting. Given the high-stakes context of healthcare, the EU AI Act justifiably sets a high bar of conduct that starts with and focuses on the user's benefit.

III. Policy Initiatives in the US

One of the most human-centred frameworks published in the US 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*

36 European Commission, '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' (January 2024) <https://artificialintelligenceact.eu/wp-content/uploads/2024/01/AI-Act-FullText.pdf>

37 Ibid.

38 'AI Act Consolidated Text' (January 2024) <https://artificialintelligenceact.eu/wp-content/uploads/2024/01/AI-Act-Overview_24-01-2024.pdf>.

for the American People.[39] While a human-centred approach is not explicitly mentioned, the blueprint identifies five principles and practices to guide the design, use and deployment of AI systems as a way to meaningfully and most beneficially impact the US 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 in particular, the US Food and Drug Administration (FDA) plays a crucial role in regulating medical devices and ensuring their 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 labelling and documentation requirements, monitoring devices for safety and effectiveness through post-market surveillance, and
- alignment among regulatory bodies.

The FDA has taken significant steps since 2017 to address the regulation of ML-based medical devices and digital health applications (ie medical software)[40] 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 and integration of AI systems has been the need to actively include diverse perspectives –eg healthcare professionals and stakeholders from industry and academia– across the AI lifecycle, as well as to continuously monitor system performance in real-world contexts to identify and address potential risks and ensure the system's ongoing safety and effectiveness. Multiple actions have been taken. I highlight their relevancy to a human-centred approach.

1. 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 ML-based medical devices. The program focused on evaluating the developer's organisational excellence, product quality, and post-market performance tracking. The goal was to prioritise oversight based on the software developer's demonstrated commitment to producing safe and effective products. It was completed September 2022.[41] While the safety and effectiveness of the product was underscored, the following proposed points would more robustly support a human-centred 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

39 Office of Science and Technology Policy, 'Blueprint for An AI Bill of Rights: Making Automated Systems Work for The American People' (October 2022) <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>.

40 US Food and Drug Administration, 'FDA Releases Artificial Intelligence/Machine Learning Action Plan' (12 January 2021) <https://www.fda.gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan>. US Food and Drug Administration, 'Policy for Device Software Functions and Mobile Medical Applications' (September 2022) <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications>. US Food and Drug Administration, 'Digital Health Policy Navigator' (14 December 2022) <https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-policy-navigator>.

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

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.

2. Proposed Regulatory Framework for Modifications to AI/ML-Based Software as a Medical Device (SaMD)

Defined as software intended for a medical purpose without being part of a hardware medical device,[42] the FDA released in April 2019 a discussion paper outlining a potential regulatory framework for AI and ML-based SaMD 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 proposed points would more robustly support a human-centred approach moving forward:
– Clarification on the definition of 'improving performance' in the context of modifications made to an AI/ML-based SaMD. This becomes critical when dealing with different types of devices.
– Clarification on the criterion used for transparency of an AI/ML-based SaMD and real-world performance monitoring of an AI/ML-based SaMD. This becomes essential in understanding how the FDA will review, assess, and approve the real-world performance monitoring the manufacturer will engage in.
– Transparency on the criterion used for the FDA's Pre-Cert TPLC (total product life cycle) approach across all sizes, types, and statuses of organisations, considering that those features affect system development and integration.

3. AI/ML-Based SaMD Action Plan

In January 2021, the FDA unveiled an action plan for AI and ML-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 ML practices (GMLP); support of a patient-centric approach; development of methods for evaluation and improvement of ML algorithms; and advancement of pilots for real-world performance monitoring. While the action plan acknowledges the pertinent input from stakeholders to guide simultaneous innovation and oversight practices, the following proposed points would more robustly support a human-centred approach moving forward, particularly on the topic of AI system transparency:
– Gaining insights directly from patients on factors that impact their trust in AI-enabled technologies is a step forward. However, such cannot be the only way to inform transparency demands on manufacturers given that patients come from many different backgrounds and may or may not understand the risks inherent to AI systems. Moreover, users should not be burdened with identifying what information they should or should not be receiving when an AI system is being utilised.
– Addressing transparency for users is a multifaceted problem. Users include both patients and healthcare providers. Transparency requirements must meet the unique needs of all users.
– Choosing the type of labelling for AI/ML-based devices is a major question in focus. Multistakeholder research efforts have been done and continue to be enhanced to create nutrition-like labels for datasets.[43] Consensus-based standardisation of features necessary to have transparent datasets can serve as the foundation for consensus-based standardisation of features necessary to have transparent mobile AI-enabled healthcare systems more broadly.

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

43 S Holland et al, 'The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards' (2018) arXiv:1805.03677. K S Chmielinski et al, 'The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence' (2022) arXiv:2201.03954.

4. 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.[44]

5. 21st Century Cures Act

In effect December of 2016, digital health and wellness applications that track, monitor and maintain the concept of a healthy lifestyle are considered of low risk and are not regulated by the FDA.[45] Without the need to apply for premarket review and register applications, system safety and risk assessment are left to developers and application marketplaces.[46] While these applications are not intended for medical use, the following proposed points would more robustly support a human-centred approach moving forward:

- Expansion of the operational definition of digital mobile health applications 'not intended for medical use,' acknowledging that digital health and wellness applications may be used, at least by patients, to self-diagnose, cure, mitigate, prevent, and/or treat a disease or condition and thus necessitate regulatory oversight.
- Rigorous evaluation of product claims and risk profiles from AI-enabled health and wellness applications, including appraisal of presented empirical support of promised health benefits and assessments of harm from use.
- Inclusion of existing premarket evaluation practices on AI-enabled health and wellness applications.
- Inclusion of data privacy and security protections, given that already hundreds of documented mobile applications for mental health, for example, have been identified as not having privacy policies in place.[47]

The FDA's role in regulating ML-based medical devices is to ensure patient safety, device effectiveness, and their responsible integration into the US healthcare ecosystem. Post-market evaluation from the FTC, although effective in regulating 'unfair or deceptive acts or practices'[48] and expanding its powers with the recent authorisation of compulsory process for AI-related products and services,[49] is insufficient for protecting public health because after-the-fact oversight is reactive instead of proactive when users have already downloaded applications and used them for unknown lengths of time. As a result, expanding FDA oversight over mobile digital health applications currently not considered medical devices becomes an imperative for maintaining not only high standards of quality and safety of all mobile AI-enabled health systems, but for upholding a holistic human-centred approach in which FIRE principles can be implemented.

IV. Policy Recommendations

On the heels of the FDA's November 2023 update regarding AI/ML-based medical devices[50] and 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 healthcare providers and patients alike, the following recommendations for industry and healthcare organisation cooperation at minimum and regulatory action at maximum are proposed for both AI/ML-based medical devices and mobile digital health apps:

1. The development of robust data governance. As demands increase for relevant, sufficiently representative, error-free and complete datasets, precision of the intended system's purpose becomes

44 US Food and Drug Administration, 'Digital Health Policy Navigator' (n 40).

45 US Food and Drug Administration, 'General Wellness: Policy for Low Risk Devices' (September 2019) <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/general-wellness-policy-low-risk-devices>.

46 TJ Kasperbauer and DE Wright, 'Expanded FDA Regulation of Health and Wellness Apps' (2020) 34 Bioethics 235.

47 K O'Loughlin et al, 'Reviewing The Data Security and Privacy Policies of Mobile Apps for Depression' (2019) 15 Internet Interventions 110.

48 Federal Trade Commission Act 15 USC. § 45 <https://rb.gy/gxb4bj>.

49 Federal Trade Commission, 'FTC Authorizes Compulsory Process for AI-related Products and Services' (Federal Trade Commission press release, November 2023) <https://www.ftc.gov/news-events/news/press-releases/2023/11/ftc-authorizes-compulsory-process-ai-related-products-services>.

50 US Food and Drug Administration, 'Artificial Intelligence Program: Research on AI/ML-Based Medical Devices' (November 2023) <https://www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/artificial-intelligence-program-research-aiml-based-medical-devices>.

paramount. New methods are needed for AI algorithm training for limited labeled training and test data.

2. The development of AI system transparency requirements and the standardisation of such. Transparency must cover the entire gamut of concepts usually associated with explainability; at minimum, those include: model interpretability, dataset sourcing and visibility, model capabilities and limitations, real-world testing and outcomes, and contextual usability. This is inspired by proposals in the image of nutrition or medical product labels.[51]
3. The creation of a registry of AI systems to make cataloguing them simple and accessible to all. Such registry would also include a section for real-time reporting errors and misdiagnoses and impact assessments of harm (with a further highlight on the operational definition requirements that entails harmonising), effectively helping post-market system monitoring. This is inspired by third-party databases that track FDA-approved AI algorithms.[52]
4. The development of risk management requirements, acknowledging that NIST's AI Risk Management Framework (RMF) can be tailored to healthcare. This is of noteworthy importance given the US President Biden's EO that significantly expands NIST's responsibilities related to responsible AI system 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, labelling synthetic content and detecting synthetic content, among other areas.
5. The establishment of AI system auditing requirements. Agreement needs to be determined on the auditing scope, timing, and parties and organisations involved. Moreover, metrics on system performance estimation and reference standards need to be established. Furthermore, summary results should be published and made easily accessible to and interpretable by all; this is inspired by requirements of automated employment decision tools as stipulated by the Local Law 144 of 2021 of the City of New York, NY.[53]
6. The interoperable management of the rich, unstructured troves of patient-gathered data from mobile digital AI-enabled devices. This becomes a requirement for system transparency, risk management, and auditing practices.
7. 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[54] 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.
8. The creation of an expert oversight board for system checks and balances. Whether under the purview of the FDA or of the healthcare provider's organisation or a separate entity altogether at the national level, this is inspired by the role institutional review boards have to ensure the protection of human rights and the well-being of research subjects.
9. The development of a traceable mechanism to detect and track the entrepreneurial efforts of tech savvy and avid amateur 'doctor' patients who tinker with applications and build better-grade applications outside of the purview of oversight. While possibly more effective than industry or healthcare organisation-produced and monitored products, these systems necessitate safety measures in place to ensure accountability in the case of error or other unexpected negative outcomes.

---

51 MP Sendak et al, 'Presenting Machine Learning Model Information to Clinical End Users with Model Facts Labels' (2020) 3(1) NPJ Digital Medicine 41.

52 S Benjamens, P Dhunnoo and B Meskó, 'The State of Artificial Intelligence-Based FDA-Approved Medical Devices and Algorithms: An Online Database' (2020) 3(1) NPJ Digital Medicine 118. American College of Radiology Data Science Institute, 'New ACR DSI Searchable FDA-Cleared Algorithm Catalog Can Ease Medical Imaging AI Integration' (1 February 2021) <https://www.acrdsi.org/News-and-Events/New-ACR-DSI-Searchable-FDA-Cleared-Algorithm-Catalog-Can-Ease-Medical-Imaging-AI-Integration>.

53 Committee on Technology, The New York City Council, 'Local Law 144' <https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=ID%7CText%7C&Search=>.

54 Washington State Legislature, 'Washington My Health My Data Act' (2023) <https://app.leg.wa.gov/billsummary?billnumber=1155&year=2023>.

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

The following recommendations as immediate best practices are complementary to the above:

1. The notification of the use of an AI-enabled system throughout any touch point across healthcare delivery. This is particularly critical when clinicians integrate AI tools like large language models within their workflows. Until AI-enabled systems become commonplace and system transparency requirements standardised and normalised, notification of use sets the stage for dialogue between healthcare provider and patient.
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 visualisation methods becomes increasingly pertinent and highlights the interrelatedness of auditing and impact assessment reports with system transparency. Adopting proposed labels function as documentation that can enhance user trust, particularly on the health provider's side, and enhance transparency and system certification review.
4. The retainment of agency over self-gathered data and understanding of the who, what, when, where and how of system data collection.
5. Human-in-the-loop functionality and clarity over such. All AI systems must entail robust end-user system education.
6. The development of appropriate capacity building to support the utilisation of AI systems across the healthcare pipeline. User competency, including upskilling in data science and machine learning, becomes paramount to understanding when and why an AI tool is useful.

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. 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 we do not remain passive in our engagement with AI tools.[55] I therefore underscore that human-centred AI principles must fundamentally guide both AI system best practices as created by industry and healthcare organisations and the proposed regulatory actions for a robust and sustainable healthcare ecosystem.

55 S Lynch, 'AI in The Loop: Humans Must Remain in Charge' (Stanford University Human-Centered Artificial Intelligence News, 17 October 2022) <https://hai.stanford.edu/news/ai-loop-humans-must-remain-charge>.