A User-Focused Transdisciplinary Research Agenda for AI-Enabled Health Tech Governance

by
David Arney, Max Senges, Sara Gerke, Cansu Canca, Laura Haaber Ihle, Nathan Kaiser, Sujay Kakarmath, Annabel Kupke, Ashveena Gajeele, Stephen Lynch, Luis Melendez

WORKING PAPER
January 2019

The views expressed in this report are those of the authors alone and do not reflect those of Harvard University, the Berkman Klein Center for Internet & Society at Harvard University, or the Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School.

Permalink: https://cyber.harvard.edu/publication/2019/user-focused-transdisciplinary-research-agenda-ai-enabled-health-tech-governance

Suggested Citation: Arney, et al; "A User-Focused Transdisciplinary Research Agenda for AI-Enabled Health Tech Governance." January 2019. Available at: https://cyber.harvard.edu/publication/2019/user-focused-transdisciplinary-research-agenda-ai-enabled-health-tech-governance

About the Authors

David Arney, Nathan Kaiser, Annabel Kupke, and Ashveena Gajeelee are part of the Berkman Klein Center for Internet & Society at Harvard University, AI-Health Working Group.

David Arney Medical Device Plug and Play Interoperability Program at Massachusetts General Hospital at Harvard Medical School.

Ashveena Gajeelee Research Fellow Global Access in Action, Berkman Klein Center for Internet and Society at Harvard University.

Max Senges Stanford CDDRL. Although Max Senges is employed by Google, this paper is written entirely in his personal and academic capacities and does not reflect the opinion of his employer.

Sara Gerke Research Fellow, Medicine, Artificial Intelligence, and Law; Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School.

Cansu Canca Director, AI Ethics Lab.

Laura Haaber Ihle Harvard Department of Philosophy; Sant'Anna School of Advanced Studies.

Sujay Kakarmath Research Scientist, Pivot Labs at Partners HealthCare | Massachusetts General Hospital; Instructor, Harvard Medical School.

Stephen Lynch Director, Innovation Learning Program, Massachusetts General Hospital | Ambulatory Practice of the Future.

Luis Melendez Co-Founder, MDii, LLC.

Layout: Daniel Dennis Jones

A User-Focused Transdisciplinary Research Agenda for AI-Enabled Health Tech Governance

by
David Arney, Max Senges, Sara Gerke, Cansu Canca, Laura Haaber Ihle, Nathan Kaiser, Sujay Kakarmath, Annabel Kupke, Ashveena Gajeele, Stephen Lynch, Luis Melendez

WORKING PAPER
January 2019

Abstract

AI-enabled health technology holds significant promise for improving health outcomes and clinical workflows. However, it also generates challenges for health data governance and security. More specifically, apps that involve AI health coaching (e.g., an AI-mediated dialogue between the user and healthcare providers) evoke concerns about medical paternalism and privacy as well as the need for encompassing a broad range of individual understandings of what constitutes a good life. Leveraging a transdisciplinary approach, this paper sets forth a research agenda for stakeholders to proactively collaborate and design AI technologies that work with users to improve their health and wellbeing.

Introduction

Daily life has been radically transformed through the arrival of omnipresent connected devices that can answer questions, help with wayfinding as well as identification and evaluation of options for restaurants, groceries, activities, and commerce in general. In healthcare, in particular, inexpensive wearable sensors and home health devices, ubiquitous smartphone apps and continuous data collection are enabling broad shifts in care dissemination and delivery. Such technologies offer the potential for care to be personalized to an individual's goals and mediated through, and in some cases delivered by, autonomous systems. Artificially-intelligent devices could manipulate us in subtle, but effective ways or enable us to obtain relevant information in order to promote rational choices aligned with our short- and long-term health goals. The information age offers us many freedoms particularly insofar as it provides us with choices about who we want to be and become. While desirable on its face, it also begs questions about the value and burden of individual rationality, informed decision-making and its relation to societal good. For example, through an ongoing artificial intelligence (AI) mediated dialogue between the user and healthcare providers, user-generated data and evidenced-based health information work in tandem to define a person-specific definition of wellbeing. Importantly, the difference between a personalized AI health coach, such as the one described here, that helps a user to achieve health goals—or live "a good life"—and another app that uses this pretense to misuse her personal data for malevolent manipulation is, at times, difficult to discern.

The research agenda put forth in this paper focuses on important touchpoints in the design of AI-enabled technology ("health coaching") that promote a positive relationship between the user and various healthcare stakeholders and empower the user to maintain as much agency as possible. We explore three areas of particular interest and propose a research agenda for the space. First, in a section about user-centered design, we discuss how we can encourage individuals to choose health coaching and treatment based on their preferences and their very personal and individual understanding of what it means to lead a good life.[1] Second, we explore how increasing access to these technologies and related health data will advance user understanding and interest in personal health status and promote intrinsically-motivated health behavior change. Third, we consider the ethical dimensions at play in developing effective AI-enabled behavior change tools ("digital health nudging").

Nudging, a concept that derives from work in behavioral economics and decision-making, has become a prominent topic in behavior change research.[2] Small changes in the presentation of choices can have a strong effect on individual decision-making in a variety of fields; and importantly, nudges influence behavior (i.e., decisions) without changing the actual choices presented to a partic-

[1] Thereby, it is essential also to consider concerns regarding public health and thus, societal well-being and its potential conflicts with individual preferences.

[2] E.g., Cohen, I. Glenn., Holly Fernandez. Lynch, and Christopher T. Robertson, eds. *Nudging Health: Health Law and Behavioral Economics.* Baltimore: Johns Hopkins University Press, 2016.

ular individual.[3] Both subtle and powerful, nudges hold obvious potential for misuse and abuse as they can be used to shape some of our most sensitive decisions with respect to health-related decision-making. Hence, building in transparency and accountability is necessary;[4] users need to be sure that the systems are working in their best interest, as they have personally defined it, while also ensuring public order and public interests more generally are safeguarded.

The digitally-driven emergent healthcare space will continue to be shaped by the innovators and technologists, physicians, policymakers and users who adopt and iterate on AI-enabled health technology. For this reason, we argue that transdisciplinary research—applied research that includes knowledge and feedback from all relevant sources—is the most appropriate method to explore this area, inform the research and development (R&D) process and communicate findings with and between relevant stakeholders,[5] such as:

- Patients and patient advocates, relatives and other social supports like community members;

- Healthcare practitioners including doctors, nurses, home health workers, and other administrative and care staff;

- Researchers from medicine, biology, psychology, economics, business, law, and philosophy (to name the most prominent);

- Experts from companies developing and providing health tech or technologies that influence health and well-being in the broader sense;

- Experts from healthcare insurance companies and from public relevant institutions (including policymakers and regulators).

Transdisciplinary approaches are not only useful when addressing problems of understanding and discovering cures for health issues, but also when the interest is more broadly centered on understanding how a problem is embedded in societal practices and challenges. In doing so, it addresses not only the "what" of a given medical treatment but also the "how." Transdisciplinary research outcomes are both holistic—in that they address the given challenge from various relevant perspectives—and stratified—in that they address concrete outcomes and practices for each area. Namely, transdisciplinary health research has the potential to result in new medical and therapeutic capabilities and identify the next frontier of mental and physical health research needs. On the other hand, it needs to produce insights and guidance regarding the ethical issues related to this technology that may facilitate the development of ethical framework, as well as legal and policy recommendations.

[3] See SECTION C: Nudging for introduction and context about nudging.

[4] As well as to allow for a continuous public discourse that leads to fair and attractive (i.e., usable and effective) practices.

[5] For the term "stakeholder" see e.g., Hansen, Solveig Lena, Tim Holetzek, Clemens Heyder, and Claudia Wiesemann. "Stakeholder-Beteiligung in der klinischen Forschung: eine ethische Analyse [Stakeholder Engagement in Clinical Research: An Ethical Analysis]." *Ethik in Der Medizin*, 2018. https://doi.org/10.1007/s00481-018-0487-7.

Illustration 1 depicts the interplay of inputs, collaboration, and outcomes.

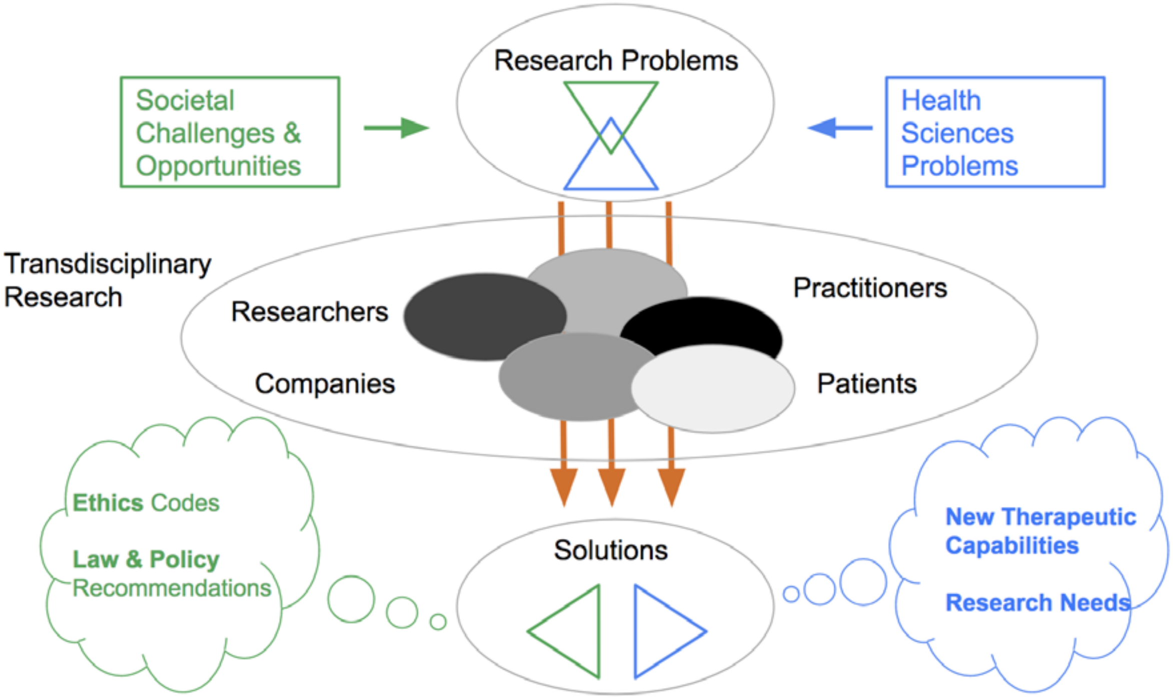
Research Problems
Societal Challenges & Opportunities
Health Sciences Problems
Transdisciplinary Research
Researchers
Practitioners
Companies
Patients
Ethics Codes
Law & Policy Recommendations
Solutions
New Therapeutic Capabilities
Research Needs

Illustration 1: Transdisciplinary Health Research[6]

One eminent question for stakeholders to consider is how to design an ethical framework within which individual "coaching programs" can be developed. There is a large body of work on health coaching, and a variety of approaches. While we have a chance to develop this ethical framework in a universal context, the legal and policy dimensions are inherently national or supranational in nature. This increases the difficulty of providing practical usability[7] as well as marketability[8] of solutions.

While most of this whitepaper is concerned with questions around research and development of AI-enabled health technology, including AI-enabled or AI-supported health coaching, we do not want to neglect the importance of a regulatory framework that allows for quick implementation of AI solutions while protecting the rights of the individuals. This has proved to be particularly challenging given the fast pace of evolution of AI-driven health technologies.

Many countries have already committed to join the AI race. Over the past months, countries such as

[6] Adapted from Jahn, Thomas. *Technikfolgenabschätzung – Theorie Und Praxis* 14, no. 2 (June 2015). https://www.tatup-journal.de/downloads/2005/tatup052_jahn05a.pdf.

[7] Users moving between jurisdictions.

[8] The need for providers to customize their offers for various legal contexts.

Canada, Japan, Singapore, China[9], the UAE, Finland, Germany[10], Denmark, France, the UK, the EU Commission, South Korea, and India have all released national initiatives on their strategy to develop and optimize the use of AI.[11] It has been forecast that global GDP will be 14% higher by 2030, amounting to some $15.7 trillion potential contribution due to AI. The countries to benefit the most will be China (26% boost to GDP in 2030) and North America (14.5% boost), equivalent to a total of $10.7 trillion and accounting for almost 70% of the global economic impact.[12]

Policymakers are already working on regulations and enforcement mechanisms to prevent malpractices and protect individual-level data. In the US, for example, a bill of the "FUTURE of Artificial Intelligence Act of 2017" was introduced into Congress in December 2017 to require the Secretary of Commerce to establish a Federal Advisory Committee whose objectives are, in addition to encouraging investment, to examine privacy issues, review legal and regulatory framework, the use of data by organizations, and how AI might make the healthcare system more efficient and cost-effective.[13] However, given the transnational nature of the opportunity and challenges, an international framework anchored in Human Rights would be desirable.[14]

The healthcare industry is already heavily regulated in many countries and the call for a new set of laws and policies to enable contribution of AI to the economy has been addressed on a national level. International organizations like the World Health Organization (WHO) were latecomers to this conversation. However, given the global impact of AI and the cross-border movement of data, the use of AI in healthcare is likely to be transnational in many cases. The need for an international regulatory framework led to the call for the setting up of an International Artificial Intelligence Organization to serve as a standard-setting body[15] and AI being a regular agenda item at the UN General Assembly.[16] AI is now considered as a potential game changer in tackling global health

[9] "China's New Generation of Artificial Intelligence Development Plan." Foundation for Law & International Affairs. July 30, 2017. Accessed September 29, 2018. https://flia.org/notice-state-council-issuing-new-generation-artificial-intelligence-development-plan/.

[10] "AI-Hub Europe Exclusive: German AI-Strategy Paper in English." AI-Hub Europe. July 26, 2018. Accessed September 29, 2018. http://ai-europe.eu/exclusive-german-ai-strategy-paper-in-english/.

[11] "An Overview of National AI Strategies." Medium. July 30, 2017. Accessed September 29, 2018. https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd/.

[12] Rao, Anand, and Gerard Verweij. 2017. "Sizing the Prize What's the Real Value of AI for Your Business and How Can You Capitalise?" https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf.

[13] "H.R.4625 - 115th Congress (2017-2018): FUTURE of Artificial Intelligence Act of 2017 (Bill)." December 12, 2017. Accessed September 29, 2018. https://www.congress.gov/bill/115th-congress/house-bill/4625/text.

[14] See Raso, Filippo A., Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, and Levin Kim. "Artificial Intelligence & Human Rights: Opportunities & Risks." September 25, 2018. Accessed October 03, 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3259344.

[15] Erdelyi, Olivia, and Judy Goldsmith. "Regulating Artificial Intelligence Proposal for a Global Solution." *AAAI/ACM Conference on Artificial Intelligence, Ethics and Society*, February 1, 2018. https://par.nsf.gov/biblio/10066933-regulating-artificial-intelligence-proposal-global-solution.

[16] "AI for Good Global Summit 2018." Accessed September 29, 2018. https://www.itu.int/en/ITU-T/AI/2018/Pages/default.aspx.

challenges[17] with little demonstrated progress and to help the WHO reach its 'triple billion' target: One billion more people benefitting from universal health coverage, getting better protection from health emergencies, and enjoying better health and well-being by 2023.[18]

There remain many challenges which we could not address in this paper but which nevertheless remain inherent to the issues at hand. Given the heavy reliance on data sets for AI solutions and that only a few select countries have embarked in the AI race, societies need (both national and international) safeguards to mitigate biases in data collection, prevent discrimination and generalization of AI technologies and assure equitable access to data. Those safeguards need to factor in cultural context specificities to ensure AI norms and standards are not imposed by a select few on the rest of the world's population.

Section A: Technology and a Healthy Good Life

While this paper lays out a research agenda for AI-enabled health technology, it is important to, at least, raise the more fundamental context of how science and technology can be made to inform each individual's quest for a meaningful and "good life."[19] As, for example, argued in the framework of logotherapy, asking questions about the creation and pursuit of meaning are crucial elements of a healthy life. Similarly, continuous learning about what and how behaviors and life's circumstances can be changed[20] to achieve one's life goals is key to calibrating AI support.[21]

AI-enabled technology has the potential to activate and engage the user by collecting and providing meaningful data about their practices as they relate to their physical and mental health and well-being, connecting their data to their behaviors. The

[17] The WHO Director General has stated that "digital technologies and artificial intelligence will be vital tools in achieving all three of these targets". See Adhanom Ghebreyesus, Tedros. "Artificial Intelligence for Good Global Summit." World Health Organization. May 15, 2018. Accessed September 27, 2018. http://www.who.int/dg/speeches/2018/artificial-intelligence-summit/en/.

[18] "World Health Assembly Delegates Agree New Five-year Strategic Plan." World Health Organization. May 23, 2018. Accessed September 27, 2018. https://afro.who.int/news/world-health-assembly-delegates-agree-new-five-year-strategic-plan. "World Health Assembly Approves New Strategic Plan With Focus on "Triple Billion" Targets." Bridges. May 31, 2018. Accessed September 26, 2018. https://www.ictsd.org/bridges-news/bridges/news/world-health-assembly-approves-new-strategic-plan-with-focus-on-%E2%80%9Ctriple.

[19] The notion of "good life coaching" adds a much deeper and more long term perspective around Maslow's goal of self-actualization or Viktor Frankl's logotherapy. This is in line with the recent school of thought around positive psychology which stresses that rather than happiness, coaching should concern itself with positive emotion, engagement, relationships, meaning and accomplishment. See Seligman, Martin E. P. *Flourish: A Visionary New Understanding of Happiness and Well-Being*. New York: Free Press, 2012.

[20] While also understanding the constraints beyond an individual's control and setting realistic goals to work towards.

[21] One moral framework that could support this system could be Aristotle's virtue ethics, where taking care of one's health and thus having the capacity to act virtuously is a moral pursuit itself. In this framework, we can also appeal to Aristotle's idea of "practical wisdom" that allows one to assess the situation at hand and determine how to balance the relevant virtues. For example, one's decision to lose weight could be morally valuable if it enables her to excel in her abilities yet if one is setting unrealistic weight goals for herself, this would show a problem with her ability to use her "practical wisdom." While we propose that sound technology design is grounded in empowering the individual and taking his/her personal abilities, desires, and values into account (i.e., user-centered design), we see the need to balance this individualistic approach with broader societal and environmentalist ethical perspectives. Again, drawing from Aristotle's virtue ethics, virtuous life is not only concerned with the individual but also how individual's life relates to what is good for the community/polis. See Crisp, Roger. "Well-Being." Stanford Encyclopedia of Philosophy. September 6, 2017. Accessed September 29, 2018. https://plato.stanford.edu/entries/well-being/.

goal of AI-enabled health technology can hence be defined to help individuals lead lives that are healthier and that match their values.

Unlike the traditional health care system, the digital world, and in particular digital assistants, have the advantage of collecting “truthful” data regarding one’s everyday health behavior and all of its aspects. Having such data continuously collected and analyzed is not possible in today’s traditional healthcare setting; whereas a holistic digital world provides this opportunity.

In traditional health care settings, most patients do not have access to guidance in well-being in accordance with their preferences and values. Guidance and coaching may be unattainable or provided as a “one size fits all” guidance for healthy behavior. A broadly available system where users feel that their values considered and reflected by the coach might also be more effective in motivating the user to follow wellbeing advice.

We propose a three-pronged model in which (1) science and (2) data are enabling an AI-mediator that continuously engages in an open dialogue with the user in order to define his or her (3) well-being related goals. It would go beyond the scope of this paper and the collective expertise of the authors to describe the technical architecture of the system we believe is viable and desirable based on the components we see emerging (personal virtual assistance, small data, federated machine learning and the proliferation of health-related wearables). Let us illustrate the system through a concrete use case. When the user queries her coach for a nearby restaurant, the top response should take her preferences as well as her goals into consideration and hence show the healthiest options first, followed by a pizza shop, then the burger and fried food locale. It is in this context that we provide our constructive analysis in order to shape a research agenda aimed at informing responsible technology and governance innovation.

Research and science of health and well-being evolve our collective understanding of what constitutes a healthy life as well as psychological well-being. However, many unhealthy practices are perceived as pleasant. AI can help the user make better decisions on their overall well-being by providing an interface to the body of insights on what science finds to be healthy. Consider the case of a recovering alcoholic who occasionally smokes rather than drinks. There are many examples where people may consciously choose to engage in objectively harmful behavior; it is not the role of technology to make these choices for the user, but rather to make them aware of what their choices entail and how they fit their overall goals regarding their individual well-being.[22] Where to draw the line on enabling harmful behavior is one of the difficult practical and ethical questions for developers in this space.

This technology enhances user agency by directly and easily enabling users to choose between different providers for guidance in well-being. By doing so, the user has the freedom to follow the advice of those providers whose value system fits theirs. In this setting, it is crucial that a gatekeeper function exists to ensure sound medical advice. Similar

[22] I.e., support them in living their personal definition of “a good life.”

to the traditional healthcare setting, where doctors and healthcare personnel have the function of providing expert guidance and thereby helping patients make informed and rational decisions, the marketplace of well-being services provided in this tool must also ensure that the advice is of a sound medical nature. Since health issues are of great importance, and a scenario where the gatekeeper function disintegrates could have grave consequences both for the individual and for the society, it is not sufficient for the tool to simply rely on 'informed consent' and it is necessary to employ an internal system of certification of legitimate providers.[23]

Emergent ubiquitous connected technology also allows for continuous passive data collection about our life practices and, hence, the AI-enabled coach can serve the user by providing insights about the goal-striving progress. Because the interface is intelligent, AI-enabled tools can help the user to understand the science as well as to "negotiate" personal health goals, training plans or behavior change strategies. The AI coach is not one monolithic entity but rather an intermediary between the user and health and well-being experts who offer training or interventions aimed at improving health and well-being. Naturally, this marketplace for health/well-being services must provide reasonable governance to allow for innovation while also ensuring users are exposed to healthy and medically sound content and interventions. Allowing for innovation without permission, a formally defined "beta phase," effectively a clinical trial, which allows for testing the efficacy and safety might be useful to complement the current liabilities and certification requirements in place for commercial and certified medical devices and services.

Lastly, our model assumes, perhaps crudely, that most users are willing to contribute their data to scientific research to allow scientists to deepen our knowledge about health and the impact of interventions. Data formats and metadata need to be standardized to allow for semantic interoperability enabling required underlying functions. As data about oneself becomes increasingly holistic, there are deep questions about privacy and the value of sharing data with the scientific community to improve scientific insights. Ownership of individual data as well as access and the right to copy data is of paramount importance to make this scenario possible.

We envision the relationship between the user and AI to take the form of a health coach which serves to augment the user. Whenever appropriate, this personal intelligence will engage in open-ended dialogue in which the AI serves as a portal to the world's knowledge while knowing the user's expertise and preferences.

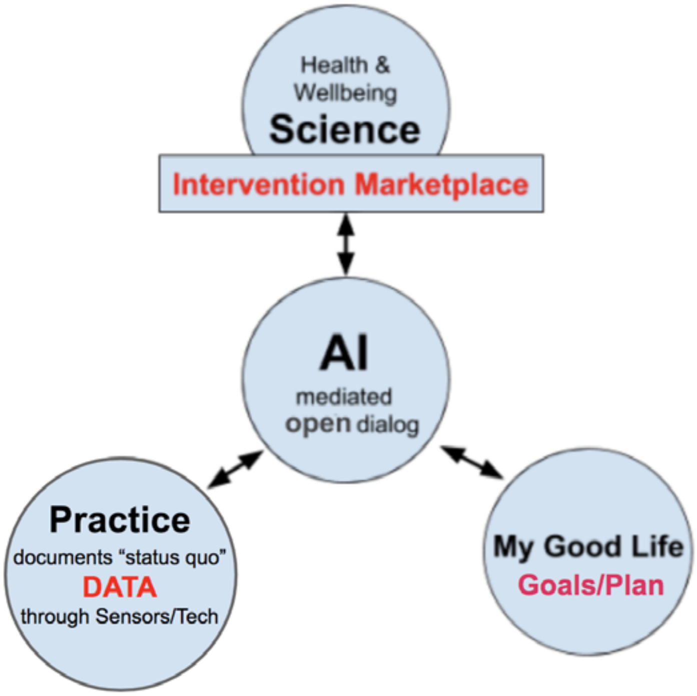

Illustration 2: AI-mediated Health and Well-being coaching

[23] The anti-vaccination movement is an example of misinformation that grew viral on media and social media with grave health consequences.

Section B: Data

Data, broadly speaking, is at the core of all AI applications. Data involves a wide range of issues around collection, standardization, ownership, and consent. We include in scope not only health data from traditional sources such as electronic health records (EHRs) but also well-being data from fitness devices and other applications that can be used to assess or impact health and well-being.[24] This casts a wide net—most personal data is relevant in this way. For instance, browser history could be mined for indications of depression, and location data can be used to estimate the amount of exercise an individual is getting. Of course, these raise important questions regarding privacy, transparency, and consent. That said, some types of data are particularly relevant for health and well-being and, in many cases, especially sensitive. These include EHRs with medications, allergies, records of vital signs, diagnoses and clinician notes, as well as demographic data about the patient.

Voluminous and contextually 'rich' healthcare data from EHRs is a major force driving the enthusiasm around using AI in healthcare. However, the primary purpose of structured EHR data collection has been to inform care and administrative needs, rather than serve as a digital representation of the person's health status. EHR data is notorious for containing unintended artifacts and incomplete data. The completeness of health status representation in an EHR for individuals is related inseparably to factors that are independent of their health. For example, younger individuals with employer-sponsored health insurance may have easier access to care than an older person on Medicaid living alone, resulting potentially in a more complete health status representation for the former.

Well-being data such as motion and heart rate from wearables is not covered by the same privacy protections as medical records but can be used to infer things about patients that, if they were stored in the patient's EHR, would be covered by the Health Insurance Portability and Accountability Act (HIPAA). We question whether HIPAA's scope should be expanded in order to include those new data sets. Instead, a separate regime for health-relevant data that is not covered by HIPAA is required and preferable to ensure adequate data protection.[25]

It seems useful to divide data more generally into three categories: non-sensitive, sensitive, and 'grey area,' as an initial classification aimed at giving the user/patient agency over how their personal data is used to the greatest extent.[26] There is a great need for work[27] on making data use agreements clear and comprehensible to participants, and for applying protections to sensitive data more broadly than just to traditional health records. It is not clear, for instance, that any data could be classified as "non-sensitive" for all people.

[24] Especially mental health analysis can benefit significantly from information consumption and communication data.

[25] For more information on this issue and possible options see Cohen, I. Glenn, and Michelle M. Mello. "HIPAA and Protecting Health Information in the 21st Century." *Jama* 320, no. 3 (July 17, 2018): 231-32. doi:10.1001/jama.2018.5630.

[26] See also Cohen, I. Glenn, and Michelle M. Mello. "HIPAA and Protecting Health Information in the 21st Century." *Jama* 320, no. 3 (July 17, 2018): 231-32, p.232.

[27] Wilbanks, John. "Design Issues in E-Consent." *The Journal of Law, Medicine & Ethics* 46, no. 1 (2018): 110-18. doi:10.1177/1073110518766025.; "Intended Use." Accessdata.fda.gov. Accessed September 29, 2018. https://www.accessdata.fda.gov/scripts/cder/training/OTC/topic2/topic2/da_01_02_0040.htm.

De-identification poses another difficulty. De-identification is not binary. Simply removing the 18 named data types listed by HIPAA[28] is not sufficient in all cases as many other health-related insights can be gleaned from data not covered under HIPAA. Moreover, re-identification poses another problem. Corporations that maintain large datasets about individuals have the means to re-identify records in a way that those with fewer resources cannot. In assessing the adequacy of a de-identification scheme, we have to consider the resources that an entity can bring to bear on the problem of re-identifying.

There are certain aspects of health and wellness, where more data on individual behavior can help people make better decisions with respect to their goals. Sleep quality is a good example, where individuals might not realize the problem until data about their sleep pattern shows that they need to improve their sleeping habits if they want to feel more rested. In addition, understanding the effects of lifestyle choices could allow individuals to exercise more agency with regard to their well-being and make decisions that reflect their personal values.

Data standardization is key to developing algorithms for diagnosis and treatment of disease as well as supporting behavioral management of health and wellness goals. Most algorithms will integrate data from multiple sources. Devices that make data available using standardized terminology with well-defined semantics, on open interfaces, and including metadata are needed to allow downstream consumers of the data to determine whether it is suitable for their intended use. In the health and wellness domain, many of these devices are regulated as strict as medical devices intended to support life-critical applications.

These requirements—standardized terminology, open interface protocols, adequate metadata—are the same ones that apply to regulated medical devices used by clinical professionals in formal healthcare environments and also to home use or health and wellness devices such as weight scales and heart rate monitors.

[28] 45 C.F.R. § 164.514.

SECTION C: Nudging

Daily behavior has an immense influence on our short and long-term health and well-being, and AI-enabled technology has the ability to influence individual and group-level health outcomes in explicit and implicit, and direct and indirect ways. Behavioral economists Cass Sunstein and Richard Thaler first coined the term "nudging," to describe a method of promoting one behavioral choice over another (or several others) while still permitting personal autonomy.[29] Nudging "alters people's behavior in a predictable way without forbidding any options,"[30] and thus does not impact the actor's ability to choose but rather, the direction, valence, and likelihood of a given choice.

Traditional nudges often appear in simple environmental design choices such as placing healthy snacks at eye-level in a grocery store[31] or selecting double-sided printing by default. By design, nudges are not explicit notifications, but rather a way of making it easier to engage in the promoted action or behavior. Nudging is inextricably tied to issues of power and autonomy—and thus, ethics.[32] Some questions that arise: who defines and decides the "right" end goal, particularly when a "healthy" goal can take many forms? Who is allowed to make changes to the environment that encourage certain choices?

To many, nudging evokes concerns about paternalism and manipulation, where unseen forces exert great effort to influence people. This can be seen as the opposite of health coaching. Indeed, while nudges do not take choices away from an individual's options set, they do, in fact, impose one party's (or group of individuals) decisions regarding the "right behavior" on those who are nudged by leverage "loopholes" to influence behavior. In some sense, individual autonomy is not violated since no option is restricted, but in another sense, through "choice architecture" individuals' rational decisions are indeed manipulated. Rather than a deontological approach, a consequentialist approach could provide the stronger argument for the moral permissibility of nudges: without taking away the individual's agency, nudges try to attain individual or social benefits.

The term "digital nudging" emerged only recently in engineering and computer systems literature, and is defined as the "use of user-interface design elements to guide people's behavior in digital choice environments."[33] Digital nudges can be personalized and driven by AI assistants or can be

[29] There is much controversy regarding the "manipulation" through choice architecture and its relation to individual autonomy. For more information on nudging see e.g., Patel, Mitesh S., Kevin G. Volpp, and David A. Asch. "Nudge Units to Improve the Delivery of Health Care." *New England Journal of Medicine* 378 (January 18, 2018): 214-16. doi:10.1056/nejmp1712984.; Simkulet, William. "Nudging, Informed Consent and Bullshit." *Journal of Medical Ethics* 44 (2018): 536-42. doi:10.1136/medethics-2017-104480; Aggarwal, Ajay, Joanna Davies, and Richard Sullivan. ""Nudge" in the Clinical Consultation – an Acceptable Form of Medical Paternalism?" *BMC Medical Ethics* 15 (2014): 31. doi:10.1186/1472-6939-15-31.

[30] Thaler, Richard H., and Cass R. Sunstein. *Nudge. Improving Decisions about Health, Wealth and Happiness.* New Haven: Yale University Press, 2008, p.6.

[31] Bucher, Tamara, Clare Collins, Megan E. Rollo, Tracy A. Mccaffrey, Nienke De Vlieger, Daphne Van Der Bend, Helen Truby, and Federico J. A. Perez-Cueto. "Nudging Consumers towards Healthier Choices: A Systematic Review of Positional Influences on Food Choice." *British Journal of Nutrition* 115, no. 12 (June 29, 2016): 2252-263. doi:10.1017/s0007114516001653.

[32] Daniel Hausman and Brynn Welch, "Debate: To Nudge or Not to Nudge". *Journal of Political Philosophy* 18(1), 2010, pp. 123-136

[33] Weinmann, Markus, Christoph Schneider, and Jan Vom Brocke. "Digital Nudging." *Business & Information Systems Engineering* 58, no. 6 (December 2016): 433-36. doi:10.1007/s12599-016-0453-1.

programmed into applications so they are seen the same way by all users. Because digital experiences are mediated through software, they are thus innately and comprehensively susceptible to nudging and other manipulation in ways that the physical world is not; for example, it takes time and effort to move the fruit to eye-level in a chain of supermarkets, but changing how fruit is displayed in an online shop only needs to happen once to change the user experience of millions of users.

The term "digital nudging" has not, to the best of our knowledge, been explicitly associated with AI technology in the healthcare context. There are manifest "digital choice environments" along the healthcare continuum—in both preventative and diagnostic settings—where AI-enabled nudging could be introduced. For example, machine-learning based health and healthcare mobile apps supporting nutrition and fitness goals and medication adherence plans have been and are being developed. Reminders to book appointments and complete web-based hospital forms are also means of preventative health measures in which an AI could step in. Other examples that fall more squarely in the healthcare or hospital environment include AI-powered clinician "assistants" to support disease diagnosis (e.g., through digital imaging) and patient monitoring (e.g., through clinical alerting interfaces) with the goal of improving health outcomes and reducing hospital staff burden. On an administrative level, "digital nudging" may aid EHR completion and consistency.

The full potential of AI derives to a great extent from its ability to make complex calculations with a speed that is several orders of magnitude faster than human capabilities. In other words, it may be impossible for a human recipient of a particular AI-enabled nudge to assess its veracity in a short span of time. To make matters more complicated, much of the technology underlying AI in its current state (e.g., deep learning) does not enable even the experts developing the technology to explain how a particular conclusion or recommendation is arrived at. In the background of such limitations, AI tools are assessed today by comparing the degree to which their performance mimics the 'reality' in the large dataset used for training, without any regard for the rationale of such a decision. It is difficult, therefore, to assume that an AI-enabled technology has learned the statistics and the "scientific algorithm" underlying a particular decision that impacts health without learning to incorporate human biases. Furthermore, it is difficult to imagine that the performance of a data-dependent technology will not differ based on the availability of data, directly and indirectly, describing a person's health status. Clinicians and other recipients of AI nudges may be compelled to trust them based on heuristics rather than sound insights.

The increasing adoption of AI-powered technology in healthcare especially has been driven, at least in part, by the digitization of health data. Large-scale data collection and machine-learning-enabled automatic validation permit the collection of more detailed patient information, potentially supporting treatment practices that are difficult and expensive to perform today. Compared to such technology, clinicians are much more limited in how often they can assess patient information and modify treatments; AI can collect data and provide guidance as frequently as once per second for weeks, involving clinicians only when their input and skills are necessary. According to Frost & Sullivan, "AI has the potential to improve outcomes by 30 to 40 percent while cutting treatment costs by as much as 50 percent".[34] Reducing healthcare costs could operate through a number of channels with one being AI-powered digital nudging of patients and clinicians.

Digital nudging flows from an AI system to an end recipient. Numerous healthcare stakeholders can leverage this AI-enabled technology to promote health among their users, consumers, citizens or patients. For example, pharmaceutical companies, insurance companies, physicians, and governments all have stakes in the health of individuals and populations. These interests are at times competing, and mediating between multiple competing nudges is likely to be difficult.

On the other hand, nudging of clinicians may be able to support them in their goal of providing care to their patients, for instance by decreasing preventable adverse events. Clinicians navigate a sea of information very quickly during most patient care encounters. Adding data from home health and wellness devices increases the amount of information they must process. Nudging is one means for algorithms to focus the clinician's attention on the important pieces of information. To do this, algorithms need contextual data about the patient and their health goals and enough knowledge of the clinician's practices and workflows to be able to promote the right piece of information at the right time. Otherwise, we risk bombarding clinicians with irrelevant or unactionable notifications.

While this technology might help individuals exercise more agency for their well-being by alerting them to the abnormalities in their data, research must be done to understand the right balance in user communication. A system that notifies users too often might result in users ignoring notifications after a while and thereby becoming ineffective in providing useful guidance. Similarly, an overflow of notifications regarding abnormalities in one's well-being related data might have the counter effect of causing anxiety and guilt for not being able to lead a perfectly healthy life and thereby negatively affecting the user well-being.

It must also be noted that while this technology aims to capture all aspects of user life, it might fall short on not easily trackable aspects of one's life. Thus, a possible difficulty here is balancing the traceable and untraceable health effects of individual behavior. For example, while the system might rightly notify the user of the negative effects of smoking or drinking alcohol, it might not be able to take into account the possible positive effects of such behavior on a particular individual, whose only way to control herself from further self- or other-harming behavior is through these substances and for whom quitting them with no further help could cause more harm.

Nudging is, by design, not obvious to the person being nudged. Thus it is difficult or impossible to ignore or filter. In physical environments, it is not possible for individuals to opt out of nudging. Like advertising, nudging aims to change behaviors. Ethical nudging should respect the individual's

[34] Belcher, Kayla. "From $600 M to $6 Billion, Artificial Intelligence Systems Poised for Dramatic Market Expansion in Healthcare." https://ww2.frost.com/news/press-releases/600-m-6-billion-artificial-intelligence-systems-poised-dramatic-market-expansion-healthcare. January 5, 2016. Accessed September 29, 2018.

autonomy and choices and support their goals. In healthcare, nudging could help people to meet the health and wellness goals they have identified and agreed to.

One key difference between nudging in the physical world[35] and a digital nudge is that—at least, theoretically—it is possible to opt-in or explicitly consent to a nudging system through personally-owned devices like web browsers or phones. As our experience of the world is increasingly mediated through technology, we are concerned about how digital nudging might distort (individual) experiences to achieve societal ends.

Furthermore, transparency in design can allow users to learn about nudging in general and concrete nudges being applied to them. When a person uses technology, a phone or other computer, it should allow the user to ask for information about the nudging happening, give them control over the amount and methods of nudging, and to ask for specific examples, for instance clarifying ‘this is meant to nudge you to smoke less in accordance with your request’.

Research Agenda

Research agendas are expressions of intellectual interests and priorities. They are inherently tied to the time and place of their creation and to the groups of people who write them. We have attempted to distill the essential issues and questions from a wide-ranging discussion among a group with very different research interests and hope that this research agenda will be of interest to the broader community.

We have followed a patient-centric approach, where we emphasize the perspective and priorities of individual patients.[36] The ultimate goal of creating technology for health and wellness is to enable patients to live healthier, more meaningful lives. This includes improving patient safety and outcomes in acute care settings like hospitals as well as health and wellness concerns in everyday life.[37] We believe that caregivers, and in the medical and behavioral health spaces, the primary care team, are essential participants in successfully selecting, prescribing, personalizing, and analyzing digital health technologies and deployments.[38] The primary care team, in the role of health coaching, seeks to provide patients with options and help them choose the best technologies available to achieve their personal health goals.[39]

[35] E.g., crowd-control in stadiums or park bench design to avoid homeless sleeping on them.

[36] “Patient-Centered Medical Home (PCMH).” NCQA. Accessed September 29, 2018. https://www.ncqa.org/programs/health-care-providers-practices/patient-centered-medical-home-pcmh/.

[37] Higgins, Tricia Collins, Jesse Crosson, Deborah Peikes, Robert McNellis, Janice Genevro, and David Meyers. “Using Health Information Technology to Support Quality Improvement in Primary Care.” March 2015. https://pcmh.ahrq.gov/sites/default/files/attachments/Using%20Health%20IT%20Technology%20to%20Support%20QI.pdf.

[38] Kraschnewski, Jennifer L., and Robert A. Gabbay. “Role of Health Information Technologies in the Patient-Centered Medical Home.” *Journal of Diabetes Science and Technology* 7, no. 5 (2013): 1376-385. doi:10.1177/193229681300700530.

[39] Quinn, Charlene C., Suzanne Sysko Clough, James M. Minor, Dan Lender, Maria C. Okafor, and Ann Gruber-Baldini. “WellDoc™ Mobile Diabetes Management Randomized Controlled Trial: Change in Clinical and Behavioral Outcomes and Patient and Physician Satisfaction.” *Diabetes Technology & Therapeutics* 10, no. 3 (June 2008): 160-68. doi:10.1089/dia.2008.0283.

Given the need for applied health tech research to embrace a transdisciplinary approach, there is a substantial need to experiment, analyze and evolve research methodologies that involve all relevant stakeholder groups. Next, to the substantive research agenda listed below, the authors advocate to applying, sharing and optimizing transdisciplinary research methods.

> How can we design an AI-enabled health coach that can engage the user in an open dialog about his/her short-term and long-term goals?

> Which ethical concerns should be in focus in the development and use of AI-enabled health technology and how can an ethical framework be constructed to guide policy-makers and other stakeholders?

> How can the veracity and accuracy of personalized recommendations (made through AI-enabled health technology) be assessed meaningfully and presented to consumers?

> How can we optimize and mitigate macro effects (i.e., public health) of AI-enabled health technology?

> What data governance ecosystem allows for privacy, legal clarity[40] and innovation?[41]

> How can innovation and offerings on the health intervention marketplace be governed at the national and international level?[42]

> How do we design nudging systems that are transparent,[43] protect individual agency and promote achieving one's goals?

[40] E.g., liability and data protection.

[41] And hence interoperability and evolution.

[42] Regulatory regime and standards setting - from codes of conduct, to certification and legal approval.

[43] Unbiased and meant to educate the user.

CONCLUSIONS

We need advances in regulatory science, if not a transformation of our approach, to better manage large-scale data collection and analytics and the use of AI in diagnosing and treating disease. Consumer data privacy protections may need to expand to cover health-relevant data that is not covered by HIPAA.

In the future, AI may participate in patient care as a member of the clinical team. Team communication is complex—an art that is learned over the course of medical training and practice. If an AI assistant is to participate as a health care team member, it must be trained to understand the clinical workflow so as to avoid interrupting at critical moments or distracting other caregivers; knowledge of this kind requires a sense of proportionality (such that the AI can judge when a medical event important enough to justify an interruption occurs) as well as "social skills" so that an AI can communicate effectively while respecting the skills and judgment of the other team members.

Another key concern that we have aimed to address in this research agenda is how technologies that run on collecting and analyzing data about all aspects of user life are susceptible to measures that utilize this data to manipulate users. While we have in no way comprehensively exhausted this discussion, several design implications and research questions aim to address this issue and provide a platform for further investigations into the ethical implications of this.

Through the data collected on behavior and on notifications provided to users about the health effects of their behavior, users might be assigned personal responsibility for their health outcomes. When utilized in health policies, this system would raise a number of ethical issues as discussed in the literature of personal responsibility in health.[44]

Tools must ensure that data is not accessed and utilized by third parties. Allowing health care providers (government, employers, or insurance companies) access to this detailed data about user behavior might result in ethically problematic health policies. Even the option of disclosing one's data in return of lower premium could function as an ethically questionable incentive if it results in non-disclosure to be "punished."

Data-driven AI-enabled health technology promises to be a powerful tool for reducing health inequities and enhance well-being at scale. The enthusiasm around these innovations must be cautiously weighed alongside their ability to amplify and perpetuate social, economic, demographic, cultural and historical biases. We put forward this transdisciplinary research agenda in hopes of convening a range of stakeholders (researchers, practitioners, entrepreneurs, policy makers, etc.) who, like us, seek to deliberate and shape the ethically and technically-intricate digital health field. Please send comments as well as proposals for collaboration to ai-health@cyber.harvard.edu or petrie-flom@law.harvard.edu or contact@aiethicslab.com or contact one of the authors.

[44] Cappelen, Alexander, Ole F. Norheim. "Responsibility, fairness and rationing in health care," Health Policy 76 (2006) 312–319. https://doi.org/10.1016/j.healthpol.2005.06.013.