"I have to believe that leaders in healthcare will make decisions for the good of the people, right?": a qualitative study eliciting perspectives on ethical issues for AI in healthcare

*Melissa D McCradden, PhD[1]
Ami Baba, MRes[1]
Ashirbani Saha[1]
Sidra Ahmad[1]
Kanwar Boparai[1]
Pantea Fadaiefard[1]
Michael D Cusimano, MD, PhD[1]

[1]St. Michael's Hospital, Division of Neurosurgery

*Melissa D. McCradden (corresponding author)
St. Michael's Hospital, University of Toronto
250 Yonge St, 6-117
Toronto ON
injuryprevention@smh.ca

Funding statement: This study was conducted as part of a project funded by Cancer Care Ontario through the Government of Ontario. SA was supported by the "Health Grand Challenge" scholarship from Princeton University Center for Health and Wellbeing. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this manuscript.

Competing interests declaration: MDM and AB receive salary support from the Cancer Care Ontario grant.

ABSTRACT

Background: As artificial intelligence (AI) approaches in research increase and it becomes more integrated into medicine, there is a need to understand perspectives from members of the Canadian public and medical community. The aim of this project was to reveal current perspectives on ethical issues surrounding AI in healthcare.

Methods: Thirty AI-naïve participants sat for an interview involving three vignettes describing relevant but hypothetical issues to potential AI applications in healthcare. Vignettes were constructed following a non-systematic literature search to identify ethical issues relevant to AI in healthcare to engage in the vignette scenario, including: informed consent, trust, public good, accountability, responsibility, unintended consequences. Responses were transcribed and coded for themes, frequency of response types, and larger themes emerging from the interview.

Results: Seventy-six percent endorsed non-consented use of health data, but advocated for disclosure transparency. A minority of patients and caregivers felt computerized output could be allowed to allocate resources, while an equal minority stated it was inappropriate to delegate such decisions. The large majority (80%) felt purchasing health data should be prohibited, with all providers in agreement. Patients and caregivers appealed to potential benefits of industry-developed applications, some stating less privacy is acceptable for the goal of improving health.

Interpretation: Patients and caregivers reported a mixture of hopefulness and concern about AI's potential while providers were generally more skeptical. These results are consistent with previous reports but with the added component of AI-specific considerations and point to a need for public education on health data research and AI.

Artificial intelligence (AI) holds immense promise for tailoring healthcare toward individuals in effort to improve all patient outcomes [1–3]. The field is evolving rapidly due largely to big data, facilitated by electronic medical records which are now nearly universally utilized by hospitals. However, utilization of big data can bring about ethical concerns regarding accountability, responsibility, and trust, among others. The speed of progress and potential for benefit is juxtaposed against catastrophizing popular perceptions of AI which undermine public trust in this technology. Nonetheless, the views of the public are essential to supporting institutions' approaches to adopting AI as well as guiding important education initiatives that may be crucial to maintaining support and trust of the public.

Worldwide efforts have been undertaken to elicit public perceptions regarding usage of health data for research [4–11], but only limited work specific to AI methodologies and interventions [12–14]. A consistent and remarkably global finding is that members of the public value the benefit to be gained from medical research but are concerned about the privacy of personal health data. However, even prior to the era of big data, individuals described concerns about use of health data for research particularly as the data is made available to more individuals and groups [4]. Paprica et al[11] recently conducted a focus group on the use of health data with Canadian stakeholders and illuminated a strong suspicion of private industry, which has been echoed by others as well [6,9,10]. The importance of understanding stakeholder perspectives that speak to bioethical concepts that characterize medicine is important to guiding the design, implementation, and integration of AI initiatives in a clinical setting.

The present study sought to expand the scope of inquiry provided by this prior work by utilizing vignettes to elicit perspectives on a non-exhaustive set of ethical concepts that are central to AI applications in healthcare.

METHODS

Design and setting

Participants were consecutively recruited at the Neurosurgical Unit at St. Michael's Hospital, as were their caregivers. Healthcare providers caring for these patients were recruited through snowball sampling. The interviews were conducted (MDM, AB) in a private clinic room at the hospital lasting, on average, approximately 15 minutes (to a maximum of 45) with participants showing a remarkable level of engagement. All participants provided informed consent and the study was approved by the Unity Health Toronto Research Ethics Board (REB).

Creation of the AI vignettes

To guide the development of the vignettes, we conducted a non-systematic literature search guided by the question, "what are the relevant ethical issues for AI in healthcare?" The full protocol is described in Supplementary file 1. The final set of ethical issues that were operationalized in the vignette were: informed consent, accountability, privacy, confidentiality, responsibility, public engagement, unintended consequences, and trust. We assessed these through three scenarios: 1) data-driven approaches to healthcare research; 2) use of machine-learning in clinics; 3) commercialization of data. The vignettes consisted of hypothetical scenarios and participants were asked to divine the perspectives of a hypothetical public as a means of assessing their own opinions while minimizing reporting bias[14-15]. Participants were also asked about their current knowledge of AI. Interviewers refrained from providing additional information beyond the details identified in the interview script (Supplementary File) and did not provide feedback on participants' responses.

Cohort

The sample of 30 participants included: 18 patients with a history of brain tumour; 7 caregivers of these patients; 5 healthcare professionals. Previous work has indicated that 12 interviews reaches thematic saturation, and while the providers were remarkable consistent in their responses, patients and caregivers held more diverse views. Participant demographics are presented in Table 1. No participants had formal experience with AI systems or methodologies.

Table 1: Participant demographic information presented by group

		Patients n = 18 (%)	**Caregivers** n = 7 (%)	**HCPs** n = 5 (%)
Gender	Female	10 (56)	6 (86)	5 (100)
	Male	8 (44)	1 (14)	0 (0)
Age	20-30	0 (0)	1 (14)	1 (20)
	31-40	2 (11)	0 (0)	1 (20)
	41-50	3 (17)	2 (29)	2 (40)
	51-60	5 (28)	1 (14)	0 (0)
	61-70	4 (22)	2 (29)	1 (20)
	71-80	1 (6)	0 (0)	0 (0)
	81-90	3 (17)	0 (0)	0 (0)
	Not disclosed	0 (0)	1 (14%)	0 (0)
	Average	*60.5*	*50.8*	*43.6*
Highest level of Education	High School	2 (11)	1 (14)	0 (0)
	College/University	13 (72)	4 (57)	2 (40)
	Master's Degree/Doctorate	3 (16)	1 (14)	3 (60)
	Not disclosed	0 (0)	1 (14)	0 (0)
Ethnicity	Caucasian	11 (61)	4 (57)	2 (40)
	Black	1 (6)	0 (0)	1 (20)
	Asian	3 (17)	1 (14)	1 (20)
	Middle Eastern	1 (6)	0 (0)	0 (0)
	Central American	0 (0)	1 (14)	0 (0)
	European	2 (11)	1 (14)	1 (20)

Outcomes and analysis

Interviews were audio-recorded and transcribed verbatim. As the vignettes involved both categorical and open-ended response types, we utilized a directed content analysis that allowed data interpretation under the umbrella of our predefined themes [15]. We (MDM, AB, PF) quantified the frequency of response types to closed-ended questions. Open-ended questions were codified based on prevailing reasoning for the answer(s) given to a particular question.

RESULTS

Participants were extremely engaged in the challenging vignettes and gave thoughtful, considered answers given their lack of technical expertise. Table 2 includes a synopsis of definitions of relevant of ethical concepts and the AI-specific responses provided by our participants. Representative quotes are presented in Table 3.

Table 2: Key concepts and key participant responses

	Definition	Key Points
Consent	Requires agreement free of coercion, with a strong understanding of the anticipated benefits and risks	· 50% felt each individual must consent to allow their data used in research · 82% were against private companies obtaining data without individual consent · 80% said even de-identified data should not be sold to private companies without consent
Privacy	Control over one's personal interests (e.g., personal health information)	· De-identification believed to be removal of name, SIN, birthdate, address, health card number · Some felt privacy should be given up for the sake of medical benefit to others
Confidentiality	Obligation of institutions to safeguard entrusted information	· Conditions under which individuals provide consent ought to be respected (e.g., use of data for health research versus marketing purposes) · Privilege of receiving health information due to its highly sensitive nature
Responsibility	Taking ownership of a decision	· 27% accepted computerized output as sufficient for allocating resources · 27% felt it inappropriate to delegate responsibility to computers · One provider likened this to inappropriate treatment; others advocated for shared decisions
Accountability	Assigning blame, answerability, liability, and proper accounting	· Media as a key mechanism for accountability · Minority indicated skepticism of institutions and companies being held accountable

Unintended consequences	Outcomes unforeseen, generated without purposeful action	· Most participants accepted that mistakes happen · All stressed the need for transparency, disclosure and reparations · Transparency and publication prevent others from repeating the mistake
Trust	Reliability, consistency in words and actions, guardianship	· Healthcare institutions are highly trusted organizations · Physicians and healthcare providers are entrusted with carrying out research with health data
Public collaboration	Supporting the meaningful participation of societal members	· Many felt that the public had a duty to be involved in research in some way · Unsure of how specifically to have a voice in medical research

Table 3: Illustrative participant responses

Protection of health data	*"The world that we live in, there's all kinds of access to information even though it's protected, but you hear all kinds of scenarios where sensitive information gets leaked, so yeah. I would have some concerns"* [18-042, Patient]
Skepticism of accountability mechanisms	*"As a member of the public my opinion doesn't count"* [18-004, Patient]
Computers allocating treatment	*"It's ethically incorrect, as you are picking and choosing who gets treatment. You need to give them options and have conversations with the patients"* [18-008, Provider]
Allowing data sale to private industry	*"You should have to give up some [privacy]... you want to be cured and [the company is] providing you with this cure, so you balance it out"* [18-012, Caregiver]
Computer-based predictions	*"Before [the brain tumour] I might say yes, because I would say, you know, it's the survival of the fittest [...] But, you can never underestimate the fight that belongs in a person, even with a disease. And they can far surpass the expectations that are set out in these kinds of statistics"* [18-001, Patient]

Trust and confidentiality	*"I mean, I think for, you know in a democratic society for members of the public to have faith in the healthcare system, I think individuals need to believe that what they believe to be confidential is held confidential, and not shared. But also for me to have confidence in healthcare systems, I have to believe that leaders in healthcare systems will make decisions for the greater good of people, right?"* [18-032, Patient]
Health data vs. other data	*"It's a privilege to be told this information – patients don't even tell their family what they tell us"* [18-054, Provider]

Conditions of the use of data for healthcare research

There was a nearly unanimous agreement that health data is a valuable resource that can be directed for the purpose of improving health and disease treatment through research, but disagreement as to the threshold for requiring consent for its use. Many of those who advocated for consent initially (with 15% endorsing consent under any and all circumstances) did feel that in an urgent, disastrous situation (e.g., disease outbreak), the circumstances were sufficiently compelling to warrant an "accelerated process" [18-008, Provider] or complete bypassing of consent. Many advocated for disclosure nonetheless, suggesting social media, phone calls, text messages, or media.

Most participants cited "de-identification" as a satisfactory condition for non-consented use of health data for research (76%). When asked about what "de-identification" meant, respondents referred to removing names, social insurance number, birthdate, address, and/or healthcare number. These perspectives were explicitly connected with the use of data by researchers in healthcare for the purpose of improving medical care.

Deference to computer outputs?

A minority of respondents (27%) readily accepted the idea that an output from a "computer" should allocate patients to treatment or no treatment based on a prediction from a computer regarding their probability of benefiting. The lone provider in this group likened this to the obligation to not offer inappropriate treatments that are unlikely to benefit a patient. The 50% who resisted this notion appealed to: fairness or equality ("trying is more important" [18-008, Provider]); fair opportunity ("everyone deserves the chance to be treated" [18-017, Provider]); evidential uncertainty ("should do more research" [18-015]); and individual factors influencing prognosis. Providers (but for 1) rejected the notion of AI allocating treatment, appealing to the need for these decisions to be made in collaboration with patients.

Upon the reveal that there was a mistake in the computer system, many declared they anticipated such a mistake and nearly all were accepting of the notion that mistakes do happen. They almost universally supported disclosure of the mistake (though one participant disagreed, fearing repercussions to the developers) and reparations, including lawsuits ("they should suck it up and pay" [18-007, Caregiver]) and efforts to financially compensate and medically treat the individuals who were excluded from treatment. Some, however, were less forgiving ("fire them and hire new researchers" [18-049, Patient]). When asked who was to blame, the majority pointed to those who developed the algorithm with a few more specifically blaming the people who input the data into the computer. One participant said that the person most in charge was responsible for the outcome. One provider described the need to publish and report the negative results so that others would not repeat the same mistake.

Secondary use/sale of data

The large majority (82%) strongly felt that selling health data to private companies should be entirely prohibited. Those who disagreed (7%) described reasoning to the effect of privacy as an acceptable sacrifice for the prospect of benefit to the larger population. They appealed to the potential benefit as justification, indicating that as long as the product being developed would help people it was acceptable practice and if its side effects were minimal. Others described the difficulty in not knowing what kind of product would be developed, where one participant astutely noted "every company thinks they're honourable but it depends on your perspective" [18-002, Patient]. Overwhelmingly and regardless of their view, participants advocated for transparency about how their data would be used, communicated openly by a trusted institution or custodian of health information.

No healthcare providers felt that it was appropriate to sell either identifiable or de-identified data. The sale of data was perceived to conflict with the responsibilities of health data custodians. One provider described patients as a "vulnerable population," as they are eager to support any endeavour purported to help others with the same disease even if they know they as individuals will not directly benefit [18-010, Provider]. That the language of research being able to "find a cure" was echoed repeatedly in this context by patients and caregivers, seems to support providers' views.

Trust and public engagement in research

Patients and caregivers described a high level of trust in healthcare institutions with regard to ethical practices, regulation, and being trusted to act responsibly with health data and follow regulations designed to protect the public. When asked about a duty to participate in research specifically though allowing use of their health data, 32% stated that people had a positive duty to allow such use for the specific purpose of researching health-related problems. An equal 32% indicated no one had such a duty. Nearly all participants who did not express a yes/no answer indicated that they personally felt a sense of duty to contribute their data to research but that not everyone would agree and that individuals wishes should be respected. Others described a positive duty only if the research involved de-identified data and no potential harms to participants.

Several participants described a morally significant difference between data obtained from social media versus health data. All providers stated health data was special, while most patients and caregivers indicated that in modern society people are now aware of the consequences of smartphone use and they tended to minimize privacy concerns. Despite a perception that these practices are now inevitable, the majority of participants clearly indicated discomfort with the lack of transparency regarding how their data was being used.

INTERPRETATION

In general, participants' responses reflected genuine uncertainty on many central ethical issues such as the acceptability of non-consented research on de-identified data, location of responsibility and accountability regarding unintended consequences, and secondary uses of data. Healthcare providers were generally skeptical, had far more reservations about allowing machine outputs to make resource allocation decisions and treatment recommendations, and universally condemned selling health data.

In alignment with previous reports, our participants endorsed the notion that the broad use of health data as a resource to improve health [10,11,19], generates tension with personal risks pertaining to privacy in particular [4]. Patients in particular derived a sense of altruism in providing their data, which contrasted the powerlessness in having suffered a brain tumour; this notion corresponded with providers'

statements that patients may constitute a vulnerable group. Similarly, Hastings[19] found that parents whose children frequented health institutions were highly motivated to have their child's data used for research; however, Bansal et al [8] found that poor health status (thus, presumably, use of health institutions) was negatively predictive of participation in research. The intersection of vulnerability and informed consent has been explored in the ethics literature and continues to present practical challenges to researchers[20-22]. These challenges may be exacerbated in the context of AI, since mainstream media hype currently dominates the discourse[23], which may promote idealistic thinking among potential participants (if not catastrophization, to the other extreme). Further, these considerations may differ according to a person's specific disease, socioeconomic status, and demographic characteristics and warrant further exploration[8].

There is universally a strong call for assurances about privacy and security of health data use [6,7,9,11]. Similar to Paprica[11], we found a generally more negative and mixed reaction to involvement of private companies in health data access. Use of data to improve health was sharply contrasted with profit-making [6,9–11]. Interestingly, the proportion of responses regarding selling of health data have remained remarkably consist over the past decade[9], suggesting that despite negative press surrounding breaches of data people continue to trust healthcare institutions[24] (in contrast with private companies). A substantial challenge to adequate protection of health data is the uncritical acceptance of de-identified or anonymized data [4], which may provide unsubstantiated confidence in sufficient protection controls and speaks to the strong and urgent need for better regulation of uses of health data in the context of AI[24].

While certain authors have pointed to technology as a potential disrupter of patient autonomy [25], our results and others [14] indicate that these concerns may not be applicable to the large majority of patients. One study [12] noted that most participants would blindly trust a robot to guide them through a rehabilitation protocol. However, the authors used an interactive, robot that approximated a human interaction, while our results described the guidance coming from a 'computer' (i.e., non-humanoid). Together, it is suggestive of the idea that if human-like features increased, people may lower skepticism of AI. Nonetheless, a minority of persons [5,12] may be generally distrustful, which may negatively impact health outcomes [12]. The providers in this study overwhelmingly indicated that treatment decisions require conversations with patients and families. Even among tech-savvy youth seeking treatment for highly stigmatized conditions[14] there continues to be a preference for interacting with healthcare providers, and it is noted that AI can be helpful in prompting attempts to seek healthcare.

Future directions

The views captured in this study reflect AI-naïve participants. Future studies may extend these findings by soliciting views from AI-knowledgeable persons. Additionally, though our sample was not homogenous, it is selective in that it represents persons with high levels of healthcare interactions; it will no doubt be important to capture the views of a more diverse group of individuals with varying levels of healthcare interactions.

Limitations

We have acknowledged that the study population is one with heavy involvement in healthcare, and thus may not adequately reflect the views of those who have fewer interactions with the healthcare system. Herein, we do not claim to provide exhaustive elucidation these ethical concepts, aiming solely to elucidate some of the particular considerations that may be relevant to AI-related research and healthcare applications. More importantly, however, is the question of how to use current views to inform the practices of healthcare institutions. While the challenge of moving from 'is' to 'ought' is commonly

acknowledged, data custodians are nonetheless compelled to act in accordance with the expectations of the data owners.

Conclusion

This study elucidated AI-relevant ethical issues among of a group of AI-naïve patients, caregivers, and healthcare providers at a large urban hospital. We illuminate the discrepancy between their expectations surrounding use of health data and emerging uses that drive data science research. Overwhelmingly, there is broad support for health data use in research not just in Canada but elsewhere as well, and all studies cite privacy as the main concern from the public. As such, AI's endorsement by society is overdue for public education initiative to earn trust.

ACKNOWLEDGEMENTS

The authors are grateful to the patients, their caregivers, and the clinicians who were involved with this study.

REFERENCES

1. Hinton, G. Deep Learning-A Technology With the Potential to Transform Health Care. *JAMA* **320**, 1101–1102 (2018).
2. Israni, S. T. & Verghese, A. Humanizing Artificial Intelligence. *JAMA* (2018). doi:10.1001/jama.2018.19398
3. Naylor, C. D. On the Prospects for a (Deep) Learning Health Care System. *JAMA* **320**, 1099–1100 (2018).
4. Robling, M. R. Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study. *Journal of Medical Ethics* **30**, 104–109 (2004).
5. Haddow, G., Bruce, A., Sathanandam, S. & Wyatt, J. C. 'Nothing is really safe': a focus group study on the processes of anonymizing and sharing of health data for research purposes. *J. Eval. Clin. Pract.* **17**, 1140–1146 (2011).
6. Lehnbom, E. C., Brien, J. E. & McLachlan, A. J. Knowledge and attitudes regarding the personally controlled electronic health record: an Australian national survey. *Intern. Med. J.* **44**, 406–409 (2014).
7. Grande, D., Mitra, N., Shah, A., Wan, F. & Asch, D. A. The Importance of Purpose: Moving Beyond Consent in the Societal Use of Personal Health Information. *Annals of Internal Medicine* **161**, 855 (2014).

8. Bansal, G., Zahedi, F. 'mariam' & Gefen, D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems* **49**, 138–150 (2010).
9. Willison, D. J. *et al.* Alternatives to project-specific consent for access to personal information for health research: insights from a public dialogue. *BMC Med. Ethics* **9**, 18 (2008).
10. Gaylin, D. S., Moiduddin, A., Mohamoud, S., Lundeen, K. & Kelly, J. A. Public attitudes about health information technology, and its relationship to health care quality, costs, and privacy. *Health Serv. Res.* **46**, 920–938 (2011).
11. Paprica, P. A., de Melo, M. N. & Schull, M. J. Social licence and the general public's attitudes toward research based on linked administrative health data: a qualitative study. *CMAJ Open* **7**, E40–E46 (2019).
12. Xu, J., Bryant, D. G. & Howard, A. Would You Trust a Robot Therapist? Validating the Equivalency of Trust in Human-Robot Healthcare Scenarios. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (2018). doi:10.1109/roman.2018.8525782
13. Hengstler, M., Enkel, E. & Duelli, S. Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change* **105**, 105–120 (2016).
14. Aicken, C. R. H. *et al.* Young people's perceptions of smartphone-enabled self-testing and online care for sexually transmitted infections: qualitative interview study. *BMC Public Health* **16**, 974 (2016).
15. Hsieh, H.-F. & Shannon, S. E. Three approaches to qualitative content analysis. *Qual. Health Res.* **15**, 1277–1288 (2005).
16. Finch, J. The Vignette Technique in Survey Research. *Sociology* **21**, 105–114 (1987).
17. Alexander, C. S. & Becker, H. J. The Use of Vignettes in Survey Research. *Public Opin. Q.* **42**, 93 (1978).
18. Hughes, R. Considering the Vignette Technique and its Application to a Study of Drug Injecting and HIV Risk and Safer Behaviour. *Sociol. Health Illn.* **20**, 381–400 (1998).

19. Hastings, T. M. Family perspectives on integrated child health information systems. *J. Public Health Manag. Pract.* **Suppl**, S24–9 (2004).

20. Levine, C. *et al.* The Limitations of 'Vulnerability' as a Protection for Human Research Participants. *The American Journal of Bioethics* **4**, 44–49 (2004).

21. Macklin, R. Bioethics, Vulnerability, and Protection. *Bioethics* **17**, 472–486 (2003).

22. Rogers, W., Mackenzie, C. & Dodds, S. Why bioethics needs a concept of vulnerability. *IJFAB: International Journal of Feminist Approaches to Bioethics* **5**, 11–38 (2012).

23. Emanuel, E. J. & Wachter, R. M. Artificial intelligence in healthcare: will the value match the hype? *JAMA*, May 20, E1-E2 (2019).

24. Vayena, E., & Blasimme, A. Biomedical big data: new models of control over access, use and governance. *J Bioethical Inq,* **14**, 501-513 (2017).

25. McDougall, R. J. Computer knows best? The need for value-flexibility in medical AI. *J. Med. Ethics* **45**, 156–160 (2019).

SUPPLEMENTARY 1: FULL STUDY PROTOCOL

Generating the vignette

Narrative review of the literature

Guided by the larger question of "what are the ethical issues associated with AI methods in health information research?", a non-systematic literature search was performed using PubMed, Medline, JSTOR, and PsycInfo, as well as using Google for publication of conference proceedings and other reports. Search terms were chosen to maximize the number of papers identifying ethical concepts relevant to our proposed work, while remaining focused on biomedical issues (versus a larger concept of AI, for example). The title and abstract of each article was reviewed to determine relevance. Articles were included if they were in English and involved explicit discussion of ethical issues surrounding both AI and Big Data within a biomedical context. We did not place limitations on the length or quality of included articles nor did we exclude specific types (i.e., gray literature, commentaries, opinion articles), but did aim to review only papers published within the past 10 years.

Articles were read and summarized, with key ethical messages highlighted and reviewed by members of the research team (SA, MDM). These were collected and refined to represent bioethical concepts that were relevant to our AI approach that could be operationalized to elicit participant responses. The final set of ethical concepts were determined when no new concepts were introduced after five consecutive articles. Initially, three scenarios were developed to target attitudes toward big data/data-driven approaches to research, unintended consequences and discrimination, and secondary use of data, as these were determined by members of our group (AS, MDC, MDM) to be the most relevant to the immediate context of AI in healthcare research.

The initial vignettes were constructed to reflect hypothetical but realistic situations that warrant public input regarding AI in healthcare. The bioethics concepts were kept opaque to avoid directly querying participants on these constructs as this was felt to be likely to elicit a social desirability bias. Hypothetical scenarios are well recognized to impose a distance between the participant and the vignette, which can minimize social desirability bias [22–24].

After initial construction of the vignettes, the research team presented them to a diverse group of undergraduate and graduate students, postdoctoral fellows, data scientists, and healthcare practitioners, each having expertise or experience in one of the following areas: computer science/AI, healthcare, ethics. The group provided specific feedback regarding language use, balancing of the vignettes for neutrality (avoiding guiding the participant toward a particular response), accuracy (but low specificity), and whether they felt the questions had construct validity. The initial vignettes were revised and presented again to the group, with further revisions to arrive at the final script.

Extended interview protocol

The interviews were initially conducted by a postdoctoral fellow (MDM) and a research assistant (AB; both female) together for 6 interviews to establish consistency in interviewing style before diving the remaining interviews between them. Both interviewers have degrees in psychology, and extensive experience with both patient contact and counselling.

No prior relationship existed between the participants and the interviewers, with the exception of two healthcare providers who had collaborative relationships with the PI of this project. Members of our research team (AS, on this paper, and others) are currently involved in AI and as such the interviewers hold a generally positive view of the technology, which can be construed as a form of bias. That the interviewers were part of a lab conducting AI work was disclosed to participants as the rationale for the project they were participating in. To mitigate such bias, the interview protocol underwent several reviews by many individuals prior to the first interview. Both interviewers were cognizant of potential framing effects, and took care to refrain from expressing a view on AI as either positive or negative, and, as mentioned, elected not to provide any feedback to participants on their responses.

Additional methodological details

There was one additional patient who was approached, but who elected not to participate in the study due to wanting to move on from anything associated with her disease. Some patients were interviewed with their caregivers present for a portion

Previous research has demonstrated that views on issues such as use of health data for research are highly diverse, and even divisive on certain issues, thus we were unsure of whether saturation would be reached for this project. The endpoint for the interviews was decided mutually by the two interviewers when it became clear after a few interviews that no new perspectives were coming about. The range of perspectives was discussed with the PI and other team members, and it was agreed that an acceptable level of saturation had been reached.

SUPPLEMENTARY 2: VIGNETTE GUIDE

Scenarios were presented uniformly and questioning proceeded based on the participant's initial responses. Interviewers gave no feedback on participant's responses as to the 'correct' response.

Scenario 1: *Jaimie is a researcher who wants to treat disease Z, which affects a lot of people in the country of Atlantis. Jaimie knows that patients' data (like medical records, X-ray or MRI pictures, and medications) is stored in big databases where they can get a lot of information that can help them answer the question of how to treat disease Z. Jaimie is thinking about whether or not to ask the people of Atlantis for permission to see their information.*

- *Do you think that Jaimie should have to ask every resident of Atlantis for permission to use their information?* [Prompt: *what if there was no identifying information included in the database?*]
- *What if there were so many people in Atlantis that it was nearly impossible for them to do so without seriously slowing down their ability to research disease Z?*
- *Do you think children can consent to the use of their health information? How do you feel about parents consenting to the use of their children's health information?* [Prompt for a suggested age]

Disease X is extremely lethal and there was just an outbreak in Atlantis. Researchers want immediate access to the health information of disease-sufferers in the hopes of finding useful information, but it is unclear what they might find. Normally, consent would be required to access this information.

- *Do you think the consent rule should change under circumstances like these? Why or why not?*

Jaimie wants to make sure that the people of Atlantis are aware of the study and wants to provide them with the opportunity to have a say in the research.

- *Do you think people in general want to be involved in research?*
- *How do you think Jaimie should involve them?*
- *Do you think that the population of Atlantis has a duty to participate in research as research subjects or use of their information?*
- *Do you think that Jaimie has the same or a different opinion about whether people have a duty to participate?*

Scenario 2: *Jaimie has started the study on disease Z. The computer*[1] *goes through all of the data and uncovers that there is a particular group of people (Profile X) who will not be able to recover from the disease no matter what treatments they try. People with profiles D and F* are *able to recover with treatment.*

Atlantis does not have enough treatment programs to treat everyone with disease Z. The waitlists are months long and many patients who urgently need help. The treatment programs consider using the profile information to decide who gets admitted for treatment and who does not.

- *Do you think treatment programs should be allowed to use the profile information to decide who gets admitted for treatment? Why or why not?*

[1] The term 'computer' was chosen to avoid using technical jargon but to connote that a machine (and not a human clinician or health expert) that would essentially divide people in treated and untreated groups.

It is later discovered that there was a mistake in the way the data was collected, and it turns out that a lot of Profile X patients actually do respond to a particular treatment regime.

- *What should happen now?* [Prompts: for Jaimie, Atlantans, the government, the untreated patients]
- *Whose responsibility is the mistake?*
- *What would Atlantans think?*

Scenario 3: *Ali is the manager of Knytes – a health information company that stores data for researchers to use to understand diseases. The people whose information is stored with Knytes agreed to let the company give their data to researchers who had received funding to do research through government granting agencies. Another company, Marcotier Health Corporation, approaches Ali about purchasing some of the data so that they can design a medical product to address the needs of a group of patients with disease A. Ali is not sure what they will design or whether it will work. The money that Knytes will receive can help fund more research and would be beneficial to the company. People who agreed to allowing their data to be used in research were never asked if their data could be used for other purposes as well.*

- *What should Ali should do in this situation?*

Suppose Ali sells the data and Marcotier developed a product for disease A. Now Marcotier is asking Ali if they can sell the identifiable data so that Marcotier can contact these patients and try to sell them the new device.

- *What do you think Ali should do?*
- *Do you think Marcotier should be allowed to do this?*

Let's say that all companies found out that they could purchase this information – some are selling life-changing products with huge benefits to patients, while others are selling ineffective products and are looking for profit.

- *What problems do you think there might be with companies accessing this sort of information?*
- *What do you think the public could do about it?*