AI Consent Futures: A Case Study on Voice Data Collection with Clinicians

LAUREN WILCOX, Google Research, USA
ROBIN BREWER*, Google Research, USA
FERNANDO DIAZ, Google Research, USA

As new forms of data capture emerge to power new AI applications, questions abound about the ethical implications of these data collection practices. In this paper, we present clinicians' perspectives on the prospective benefits and harms of *voice data collection* during health consultations. Such data collection is being proposed as a means to power models to assist clinicians with medical data entry, administrative tasks, and consultation analysis. Yet, clinicians' attitudes and concerns are largely absent from the AI narratives surrounding these use cases, and the academic literature investigating them. Our qualitative interview study used the concept of an *informed consent process* as a type of *design fiction*, to support elicitation of clinicians' perspectives on voice data collection and use associated with a fictional, near-term AI assistant. Through reflexive thematic analysis of in-depth sessions with physicians, we distilled eight classes of potential risks that clinicians are concerned about, including workflow disruptions, self-censorship, and errors that could impact patient eligibility for services. We conclude with an in-depth discussion of these prospective risks, reflect on the use of the speculative processes that illuminated them, and reconsider evaluation criteria for AI-assisted clinical documentation technologies in light of our findings.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: ethics; AI; speculative design; design fiction; health; voice; audio; data

ACM Reference Format:
Lauren Wilcox, Robin Brewer, and Fernando Diaz. 2023. AI Consent Futures: A Case Study on Voice Data Collection with Clinicians. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 316 (October 2023), 30 pages. https://doi.org/10.1145/3610107

1 INTRODUCTION

Artificial intelligence (AI) applications, including those based on machine learning (ML) models, are increasingly developed for high-stakes domains[1] such as digital health and health care. New and existing forms of data are proposed to play a role in model development in these domains, yet how we treat the collection of data for these applications can have consequential impacts (e.g., the selection of one kind of health data: health care costs, as a proxy to allocate resources

*The work was done while the author was a visiting researcher at Google Research.

[1]Our definition of *high-stakes* draws from Sambasivan et al. [2021] who characterize high-stakes AI as having safety impacts on living beings, due to the complex systems within which they are deployed, the involvement of populations at risk, and the combination of two or more disciplines leading to "greater collaboration challenges among stakeholders across organizations and domains"[132]. In our case, we are also concerned with complex informatics systems such as electronic health record systems, which mediate communication and decision-making among members of a care team, and manage administrative functions such as billing.

Authors' addresses: Lauren Wilcox, Google Research, Mountain View, CA, USA; Robin Brewer, Google Research, Ann Arbor, MI, USA; Fernando Diaz, Google Research, Pittsburgh, PA, USA.

2573-0142/2023/10-ART316
https://doi.org/10.1145/3610107

to patients, which in practice led to inequitable distribution of health resources [116]). Recent discourse in the research community focuses on data for ML model development as a critical source of differential performance across subgroups in various ML tasks [e.g., 28, 117, 127, 130]. These conversations center on characteristics of collected data and how they affect downstream model performance, which is often also evaluated using collected data. In response, mitigation strategies for differential performance focus on documenting data collection, or adjusting data in a dataset [e.g., 3, 67, 111, 130, 131].

Yet, questions abound about the implications of data collection and dataset construction practices themselves, including the impacts of the act of data capture in specific conversational and social contexts: recent reflections on ML research demonstrate that treatment of humans in dataset construction has generally been undervalued and understudied [36, 81, 136, 154], and values between those who contribute data and those who make use of data may be misaligned [52, 66, 110]. While there are calls to bring human values more prominently to the forefront of data collection and model development, work in this area is still early-stage, with few case studies [e.g., 82, 94].

In this paper, we explore the role of speculative design genres such as *design fiction* [79], to enable the exploration, critique, and better planning of possible AI futures in high-stakes collaborative scenarios. Building on work identifying prospective harms of data collection relating to privacy and algorithmic experimentation [17], we are interested here in methods that enable us to better anticipate risks related to new forms of data collection in collaborative, high-stakes contexts. We focus on the context of clinical health, in which collection of new multi-modal forms of data—particularly conversational voice data collected during a health encounter—is beginning to broadly emerge in research and product applications of generative models (such as large language models (LLMs) [98, 157]) that assist with clinical tasks, from data entry and documentation to visit analyses [e.g., 50, 63, 95, 115, 129, 134].

Clinicians' attitudes and concerns are largely absent from the academic literature and industry narratives surrounding the use of AI for these clinical tasks. As such, we are concerned here with clinicians' perspectives on new forms of data collection being proposed—in both academic and industry settings—for training models to enable AI-based assistance with clinical documentation tasks. We scoped this study to medical doctors in outpatient primary care, urgent care, and emergency settings, who bring extensive lived experience related to the natural setting and sociotechnical systems that surround AI deployments.

Our research questions include:

(1) How can design fiction be used with clinical caregivers to elicit values around AI-related voice data collection in consultations?
(2) As conversational voice data collection is proposed as a means to train ML models to assist clinicians, what potential benefits and harms of such data collection do clinicians foresee?
(3) How do the prospective risks of harm surfaced by clinicians compare to those documented in the literature identifying ML and AI harms? What new areas of inquiry are suggested by differences?

This paper contributes the following to the CSCW literature:

(1) A case study in which we used the creation of a conceptual *informed consent process* as a type of design fiction, to support qualitative elicitation of the prospective benefits and risks of data collection associated with a plausible, near-term AI technology.
(2) Findings from a study with 16 medical doctors, through which we distilled eight classes of potential risks of conversational voice data collection for AI assistance that clinicians are concerned about, including clinical workflow disruptions, self-censorship, and errors that could impact patient eligibility for services.

(3) Reflections on the prospective risks we identified—including considerations for similar, future studies and system evaluation approaches, and expansions upon the types of AI-related harms identified in prior work.

2 BACKGROUND AND RELATED WORK

Participatory forms of research, long-acknowledged and celebrated in CSCW and HCI, are becoming increasingly applied in AI/ML research. Recent work points to benefits of stakeholder participation in AI/ML development that range from improvements in problem formulation [107], to methods for algorithmic critique and accountability [76, 82], to better benchmarks for machine translation [113]. Researchers have also argued for the need to bring more human participatory methods to topics related to AI ethics [18, 44, 94]. Lee et al. [2019], in order to address algorithmic fairness issues, use participatory design to understand participants' values and build models reflecting them [97].

On the other hand, critiques of the use of participatory methods in AI/ML have cited the potential for co-optation [18], as well as ephemeral engagement and insincerity to justice [18, 143]. So, while the literature suggests an opportunity for integrating participatory design methods into AI ethics research, case studies demonstrating precisely *how* to use specific methods for this purpose are limited [e.g., 9, 53]. In this work, we adopted elements of speculative design research to explore its role in surfacing prospective benefits and harms associated with a fictional, yet plausible near-term sociotechnical AI scenario.

2.1 Speculative design research and design fiction

Speculative design is a type of systematic inquiry through which designers envision, reason about, and offer for debate aspects of alternate futures [54]. While speculative design has always sought to prompt consideration and conversation, much of the early work was limited to specific institutions, such as museums. Now, speculative methods are increasingly applied in computing research [e.g., 25, 53, 62, 72, 73]. Speculative design activities in research often invite participation from envisioned 'users' and other stakeholders, and can enable imagination of "future configurations of technology and society within the world" by adopting creative and critical lenses [85] as well as critique of current practices [6]. This approach has been widely applied to uncover novel perspectives on complex problems (e.g., in global education [85], human rights [4], public safety [68], food supply chain optimization [47], and health care [74, 75]). By considering alternate futures, speculative design "interrogates questions about values" [92, 161], explores ethical boundaries of existing systems [54], and employs creative activities and probes that invite critical reflection and consideration of different people's lived experiences [24, 35, 68, 72], with the aim to avoid futures that are maladapted to the realities of everyday life [7]. Speculative approaches explore many stages and types of technologies, with recent work calling for more emphasis on plausible, foreseeable futures [119, 150].

Design fiction is an approach that is often used in speculative design, to engage people with "technological futures" and artifacts that make discussions and debates more tangible [54]. It offers ways to consider alternative appealing and unappealing possibilities with others, by speculating to immediate, near-term, or more distant futures [79]. A key aspect of fictions is their focus on future-driven narratives about a particular context, and use of material representations of technological futures as props and provocations for reflection [27, 79, 102, 151]. As such, they have started to emerge in combination with Value-Sensitive Design (VSD) [64] as a means of surfacing ethics concerns in AI and encouraging reflection about the potential social impacts including downsides of technology design [e.g., 9, 119, 123].

Speculative fictions can take different forms and explore discussions of tangible and less tangible contexts [13]. Ethics explorations include design fictions in the form of games to engage product

designers in discussions about ethics and AI futures [9], media to engage people in discussion about data collection and harms of the data economy [106], toolkits to explore how AI ethics intersect with other areas such as policy [93], and fictional papers about data and consent in online platforms [60]. Researchers have also used design fiction to speculate about futures within specific AI-driven use cases such as personal assistants, robots, smart homes, and social agents [e.g., 7, 41, 105, 128, 138, 145].

In this paper, we use design fiction as a means to explore physicians' perspectives on the collection of conversational voice data for an AI documentation assistant, in the context of their consultations. We scoped the current study to physicians who have experience caring for a wide variety of patients, and who are responsible for managing a range of patient consent processes as part of their care practice (Section 3.1.3). The clinicians' perspectives we report on are vital to understand and abundantly rich. Yet, as we stress in Section 5.5, subsequent studies with patients, communities of informal and formal caregivers, and other stakeholders in the larger care ecosystem, are essential to conduct, to complement this work.

2.2 Use of informed consent as a design fiction

This study is not about the design, development, or study of a specific consent protocol (different forms of consent are obtained for different purposes in health care, and include permission to perform procedures, and health information collection and permission to coordinate care across care providers) [142]. Rather, we use the *concept* of informed consent, and particularly its emphasis on *disclosure of information* and distillation of *benefits, risks, and alternatives* in research, as a *frame* for a generative study activity in which we elicited physicians' perspectives on the benefits and risks of consenting to voice data collection (i.e., capture of patient–clinician dialogue unfolding throughout a consultation) associated with a class of near-future AI technology.

The use of informed consent as a design fiction enabled us to scaffold study activities with values and bioethics principles considered to be fundamental. These principles, *respect for persons*, *beneficence and non-maleficence*, *justice*, and *autonomy*, oblige physicians to disclose truthful medical information and options to the patient to support their self-determination, informed voluntary agreement, and confidentiality [51, 61]. For example, *respect for persons* is a core principle demonstrated through the informed consent process, as the construct of *informed* encompasses disclosure and comprehension, while that of *consent* encompasses voluntariness, competence, and permission of the consenting party [64].

While the concept of informed consent enabled us to reference principles familiar to participants and use these principles as *lenses* to elicit qualitative feedback, we also note the limitations of this framing. There are many critiques of the current positioning of informed consent in ethics and its underlying assumptions. Informed consent has been characterized as hard to truly define, and more of an ideal for which to strive, given the complexity of knowledge that "informedness" requires, and the situated nature of consent [86]. Though bioethical foundations of consent require that benefits and harms be explained and weighed to a consenting party, people can have differing views of what constitutes harm [65]. Furthermore, a voluntary "choice" to consent is not a single event, but a complex, ambiguous process [2].

Previous critiques of informed consent in health research also argue that it is principle-led, and such an approach can abstract the process of consent away from its clinical and social settings [45]. Take, for example, the principle of autonomy. Distinct conceptions of what individual autonomy means, and the ethics associated with these conceptions, can vary cross-culturally [118]. A rights-based notion of autonomy can ignore the actual context in which consent takes place, which, in many global health settings, is shaped by the absence of vital medical treatment, desperation for health care, and important cultural differences. Regardless of researchers' or practitioners'

intentions, their power and status shape biomedical encounters, including how those choosing to consent to activities actually make decisions [16]. Refusal to consent to health activities can be met with stigma and social penalties [16]. All of these findings belie the supposition that assent or consent can be freely (voluntarily) granted [139]. Thus, the implementation of the principles undergirding informed consent and their operationalization in context are active areas of research.

2.3 Speculation context: language modeling in health consultations

In the previous section, we outlined our rationale for the use of an informed consent process as a design fiction. In this section, we discuss the specific class of technology for which participants were asked to design a consent process, the clinical context within which it would be expected to fit, and how the speculative activity was planned with this context in mind.

2.3.1 Technology focus. Our study used a *fictional* demonstration of a *realistic* class of technology that is being actively developed and advertised by vendors, concerned with the *interface between natural language and health record data tasks*. Electronic Health Records (EHR) system vendors and clinical AI labs have recently developed natural language models to automate medical chart reviews by evaluating EHR data and identifying opportunities for improvement and validation of documentation for patient encounters [39, 89, 146]. Attention is now shifting from enabling data reviews to automating data entry: clinical documentation assistance would take as input semi-structured and ambient multi-modal or audio data from the patient consultation to generate EHR data entries. Here, speech recognition is envisioned to play a role in eliminating the burden of data entry by clinicians, by building on medical dictation technology and language modeling of natural speech during the consultation. In these scenarios, AI is envisioned as an "auto-scribe", performing speech recognition and natural language processing and modeling, by "listening in" on the consultation (e.g., an AI assistant extracts as input speech data from the conversation between patients and the clinician, and returns clinical note content, prescription orders and referrals, or narrative data to directly populate the EHR for that patient) [e.g., 43, 63, 88, 89, 101, 115, 134, 146]. Other anticipated uses of conversational data (currently textual but with implications for voice data) include training models to assist with evaluation of effectiveness of the delivery of therapeutic messages and suggesting improvements, in mental health care conversations [140].

2.3.2 Clinical context. The technologies described above provide real-time assistance to reduce the need for manual data entry or human transcription, and they are being proposed both for in-person encounters and as added capabilities in telehealth contexts, to document visits [63, 95, 115], with assurances of Health Insurance Portability and Accountability Act (HIPAA) compliance. In the U.S., HIPAA sets legal standards for the protection of personal health information and applies to health care providers, health plans, and health care clearinghouses. Thus, AI documentation systems in clinical settings belong in a category of regulated medical devices or systems. However, according to Parasidis et al. [2019], existing regulatory frameworks that could cover health data do not fully protect use of health datasets for machine learning. HIPAA does not necessarily mandate ethics review for data collection and downstream use, and data that is de-identified is no longer considered protected health information, regardless of who controls the information. As Parasidis et al.[2019] argue, the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act focus on notification, consent, and deletion rights, but this doesn't necessarily address specific issues concerning ethical collection *in situ*, nor specific information for patients about use of their data for ML models [121, 131].

Other potential concerns related to data collection, such as patient comfort with multi-modal data capture during the encounter, social risks associated with recognition or interpretation errors, or possible privacy violations due to unintentional capture of data with ambient recording methods

in the clinical environment, have neither been considered in research nor accounted for in standards to-date. We explored these concerns with near-term speculative research, using consent as a design fiction to explore potential benefits and risks.

2.4 Current practices, privacy norms, and contextual appropriateness of voice data capture

Of course, patients also have interest in digital recordings of clinical encounters, motivated by the wish to replay and share the recordings with others, to own a personal record, and to potentially verify substandard care [57, 58]. In the UK, almost 70% of patients surveyed by Elwyn et al. [2018] desired recordings of their doctor visits, yet over a third of those surveyed indicated they would ask permission first, so as to prioritize a trusting relationship with the health professional. Those who recorded covertly cited a pre-existing lack of trust or fear of denial of services [57]. In fact, patient recordings of health consultations are generally seen as a threat by health professionals [153]. Patients' *covert* recording of clinical encounters is seen as emblematic of an erosion of trust between patients and clinicians, and evokes strong negative reactions from health care stakeholders [153]. As such, policies surrounding use of such data are still in progress [56]. Yet, AI technology concepts—including those advanced by EHR companies—seem to take for granted the ability of an EHR system to capture ambient, multi-modal data during consultations [e.g., 50, 63, 89, 95, 115, 134, 146].

Thus far, capture and use of such data is not discussed in the context of patient ownership, nor are its implications for trust or power relationships between clinicians and patients (nor clinicians and institutions). Health professionals' responses to both covert and consented patient recordings of health consultations—taken together with the mixed experiences and behaviors of patients surveyed by Elwyn et al.—suggests that the health consultation setting embeds contextual expectations of privacy and confidentiality that are interacting with modern technology in new ways.

Looking at these findings through the lens of Nissenbaum's [114] concept of *contextual integrity* suggests that clinicians' discomfort is a response to a perceived deviation of the act of audio recording from acceptable norms of a health consultation. Contextual integrity ties *"adequate protection for privacy to norms of specific contexts, demanding that information gathering and dissemination be appropriate to that context and obey the governing norms of distribution within it"* [114]. There is a need for contextual standards and benchmarks for upholding privacy in this setting; however, in the age of EHR systems and telehealth, norms become ambiguous: digital capture of a variety of patient data prior to a visit (e.g., questionnaires and instruments connected to mobile devices) is now indeed a norm—but ambient recordings of the visit on a patient's mobile device are not. To understand how to develop standards and policies for multi-modal forms of data capture such as voice data, we need to understand attitudes and values that help shape norms (e.g., is there an appropriate form of voice data capture in a health consultation? What about voice data capture to train ML models? If appropriate, under what circumstances of consent, ownership of that data, and ongoing use?). These overarching questions directly motivate the specific research questions and study design we report in this paper.

2.5 Additional risks in clinical documentation automation

There are potential consequences of automating clinical documentation beyond the risks associated with conversational voice data collection. These consequences include automation bias [70], which researchers have argued could lead to a diminishing ability of clinicians to apply their background and skills, and fewer opportunities to think, problem-solve, and reflect critically through manual documentation activities [159].

Additionally, researchers have warned of risks to data privacy and implementation challenges when adding new technologies to the existing informatics ecosystem in general (e.g., upfront

costs, integration with existing technology, time-intensive training, accommodating linguistic differences) [70, 159]. These concerns are due, in part, to complex work practices, with institutional, social, and situated dimensions [11], and issues with the accuracy, completeness, relevance, and comprehensibility of the documentation [99]. In response to these concerns, Willis and Jarrahi [2019] suggest a model of human–assistant interaction in which clinicians make sense of patient context, while an AI assistant provides decision support and text entry into the EHR.

These studies highlight application-specific considerations; yet, in high-stakes domains like health, not paying enough attention to the ethics of *data collection* and *data quality* can have outsize impact on vulnerable communities (of which both clinicians and patients can be a part) [132].

Our field is still grappling with fundamental questions around the ethics of collecting data across various audio-visual and sensing modalities, including voice data to train models for new kinds of AI assistant technology in consultations. Notably, no studies to date examine human concerns related to contributing conversational voice data to an AI-relevant, clinical dataset. Nor do we understand what challenges clinicians expect to encounter when they and their patients weigh the decision to consent to such new forms of data collection and use. Focusing in this study on clinical caregivers, we sought to understand how elements of design fiction could be used to elicit their perspectives on prospective benefits and risks of using conversational voice data for AI assistant technology.

3 STUDY

In the context of informed consent, beneficence and non-maleficence require that benefits and risks associated with data collection and use of a near-term technology be identified and weighed. This requires the conceptualization of all potential harms the participants could foresee. In our study, participants conceptualized such benefits and harms based on their personal and professional experiences treating patients in primary care, urgent care, and emergency settings, and overseeing the patient consent process for treatment and research in these settings.

3.1 Methods and participants

3.1.1 Participant recruitment. Through outreach to a pool of MD-holding contractors assisting with projects at our institution, we recruited a purposeful sample [120] of participants who currently practice primary care, urgent care, or emergency medicine. These care settings have been referred to in the literature as being able to benefit from voice-enabled AI assistance [43, 88, 95, 101], and require management of a broad range of health topics, conditions, and social situations. Physician participants had a diverse range of residency training, worked in a range of institution types in the U.S. (Tables 1 & 2), and covered a diverse set of patient populations (e.g., pediatric, geriatric, patients with chronic conditions, patients with disabilities, unhoused patients). Half of the participants work in public care settings, and half provide care in a large city.[2] Participants had 3–37 (mean=10.4) years of practice experience (including their residency). We did not collect age, race, or ethnicity data, nor did we ask participants about gender beyond preferred pronouns (leaving this open-ended for their own specification rather than selection from a list). Eight participants preferred “she" pronouns, while the other eight preferred “he".

3.1.2 Study design. Our study sessions lasted approximately 60 minutes each, and took place both in-person (n=12) and through video calls (n=4) to accommodate geographic differences of participants. Most sessions (n=13) were conducted with one researcher and one participant, with

[2]We use “large city" to refer to a principal city of a metropolitan area, population $\geq$ 500,000 and pop. density $\geq$ 5000 people per square mile. “Small city” refers to a city with a smaller population $\leq$ 250,000 and pop. density $\geq$ 5000 people per square mile. Suburbs are outlying districts of a principal city.

one group session including three participants and one researcher. All sessions were recorded and transcribed, and consisted of semi-structured interviews and generative narrative activities. Prior to conducting study activities, we received written consent from all participants. All participants completed the study, and each received the fair market rate for one hour of their time.

3.1.3 Approach to design fiction. We mapped our study process to Lindley and Coulton's [102] three descriptive steps in design fiction. To contextualize this process, we first establish that "design fiction uses narrative elements to envision and explain possible futures for design" [148]. The narrative elements we employed included: (1) visual concepts and explanations of the AI documentation assistant, to first "create a story world"[102]. We then (2) began to envision and design something situated in that story world (a conceptual informed consent process). Finally, through freelisting activities and narrative explanations of considerations for consent design, we (3) "created discursive space" [102], within which to explore prospective benefits and risks of the AI assistant.[3] We distilled these three elements into two stages, as follows.

3.1.4 Stage 1: Introducing the AI assistant concept. After a series of questions about the physician's background, practice setting, and experiences with technology, we presented narrative concepts to introduce a fictional, near-future AI documentation assistant, explaining to each participant a definition of AI (similar to [122]) as *"a computing system or application that can learn from existing data to perceive information and perform tasks."* We explained that this AI assistant is trained to recognize elements of dialogue in a health consultation that are clinically-relevant. We then provided participants with examples of phrases that the AI could recognize and use as input for assistance with tasks such as generating clinical note suggestions, walking through a technique used by health care providers for eliciting a medical history from a patient, in order to outline the phases of assistance in the concept: voice recognition, speaker identification (i.e., doctor or patient), analysis of natural language for medical concept identification, and suggestions for EHR data and documentation input. These elements created the "story world" for a design fiction, within and about which we next envisioned possible futures [102].

3.1.5 Stage 2: Introducing informed consent as a design fiction. After showing and explaining the concepts, we engaged participants in a generative speculative activity. We asked each physician to consider a hypothetical, conceptual informed consent process such as a study consent protocol, in which disclosure of information and distillation of benefits and risks would be part of the consent process. We asked them to contribute to designing this process, by elaborating on all of the considerations they would bring to the consent design, including information for the patient, and their perspectives on benefits and risks.

To produce these elaborations, we asked them to *freelist* [37] thoughts and narrative scenarios as a generative activity (freelisting is a technique used to elicit all the items that come to mind in response to a probing prompt, and enables the discursive aspects of design fiction referred to in 3.1.3). We then discussed their freelist responses and narratives explaining them, and each participant elaborated on them orally in discussion with the researcher.

3.2 Data analysis

All data were analyzed inductively, using reflexive thematic analysis [22, 23]. Being a post-positivist approach, reflexive thematic analysis acknowledges researcher influence on data interpretation and the expertise they bring to interpretation and analysis. Rather than focusing on inter-rater

[3]Selected study materials are included in the Supplementary Materials

Table 1. Study participants' medical training (N=16)

Training (Residency, Additional Programs)	*n*	(%)
Internal Medicine	7	(44%)
Family Medicine	2	(12.5%)
Internal Medicine and Pediatrics	1	(6.25%)
Internal Medicine and Pediatrics, Public Health	1	(6.25%)
Family Medicine, Public Health	2	(12.5%)
Internal Medicine and Endocrinology	1	(6.25%)
Emergency Medicine	1	(6%)
Public Health	1	(6%)

Table 2. Institutions (U.S.) in which participants currently practice, grouped by specialty. Participants practice outpatient primary care, urgent care, and emergency medicine in these settings. Speciality (Training) is used to group medical center types. Some participants provided care in multiple types of settings.

Training	Medical Institution Type(s) (institution setting)
Internal Medicine	Veteran's Administration (VA) medical center (suburban) Public hospital and level I trauma center (large city) Public hospital and level II trauma center (large city) County public health department (suburban) Academic medical center (suburban) Small medical group (suburban) Large integrated managed care system (large city, suburban) Not-for-profit acute care hospital (large city, suburban)
Internal Medicine and Pediatrics	Not-for-profit medical center (large city, suburban) Not-for-profit hospital network (large city)
Internal Medicine and Endocrinology	Academic hospital and level I trauma center (small city)
Family Medicine	Not-for-profit primary care clinic (suburban)
Family Medicine and Public Health	Not-for-profit primary care clinic (suburban)
Emergency Medicine	Public hospital and level I trauma center (large city)
Public Health	County public health department (suburban)

reliability metrics and consensus-based agreement, researcher engagement with—and in-depth reflexive discussions about—the data are encouraged [22, 23].

To arrive at topic categories and overarching themes, two researchers first engaged in data familiarisation by independently reviewing transcripts and all generated text from the study sessions. They engaged in deep and prolonged data immersion and discussion, independent open coding with written notes and spreadsheets, and then discussion to compare interpretations of the meaning of the data to generate initial domain categories (which represent the diversity of meaning covered by topic area [23]) and themes from coded data. They met in multiple rounds to collaboratively revise and refine open codes for these data. All disagreements were resolved through collaborative rounds of review among researchers, conducted in synchronous discussions. Finally, the researchers met in multiple rounds to generate and review domain categories and themes based on reflections on data and relationships among codes, reflexively [22, 23].

4 FINDINGS

Our analysis distilled a set of prospective benefits, as well as eight categories of risks of voice data collection to enable AI assistance with clinical documentation tasks. After summarizing prospective benefits here, we describe findings relating to risks in detail. We reflect on the benefits and limitations of our study approach and the specific risks we surfaced in Section 5, and further consider them in the context of related work at the intersection of CSCW, health informatics, and AI ethics. We conclude by discussing future directions for research that our findings suggest.

4.1 Benefits include a focus on the patient, and depend on contingencies

Study participants surfaced potential benefits for themselves and patients, all of which centered the patient, including: the use of conversational voice data to supplement physician memory about the visit (n=2) and easier data entry if the AI assistant is highly accurate (n=12). Interestingly, a few physicians mentioned the possibility that audio capture during the consultation could allow the "patient's own words" to be included in clinical documentation. Many physicians (n=8) also discussed the possibility of sharing segments of the captured audio data with the patient (for their own record-keeping, potential for translation, and education) as a potential benefit. As P8 explained, *"[Patient could] get to re-listen to MD's thoughts on assessment or plan on their own time, at their own rate, in a setting other than one in which they have an abundance of data and info thrown at them (i.e., clinic). Also helpful when patient forgets instructions (e.g., dosing of meds)."*

However, physicians discussed the need to mitigate risks of sharing such data. As P1 told us, *"What might the patient do with this data that I wouldn't like, e.g., put [it] on YouTube…"*. In fact, almost all physicians who discussed benefits also discussed their contingent nature (e.g., dependency on "perfect" voice recognition for the net benefit to be realized, availability of control capabilities to select specific audio segments for an AI assistant to use). Below, we elaborate further on the classes of risks our participants surfaced.

4.2 Prospective risks highlight physicians' concern for both patients and clinicians

4.2.1 Erosion of trust in clinical caregivers and the health care system. Study participants talked about the general lack of comfort they expected patients, and themselves, to experience with an AI assistant that captures conversational voice data in a consultation. Patient and clinician comfort is an important part of good communication, which is key to patient satisfaction and other outcomes that health providers measure [14, 147]. Many clinicians had concerns about patient discomfort with being asked to consent to collection of their voice data during the encounter.

Specifically, many study participants raised concerns about impacts of voice data collection during the clinical consultation on trust. As P8 told us, *"Especially [for] some patients, they might wonder: the intended purpose is to decrease my doctor's burden, but is there another purpose? Am I being put into a [situation] that actually has some perverse incentives?"* This view was echoed by P9, a public health expert who cares for patients with complex conditions, who anticipated patients' *"concerns over trust regarding perceptions of medical research—especially in historically marginalized and mis-used populations or communities, particularly communities of color."*

Several physicians (n=9) brought up the possibility of changing the way they approach the conversation if voice data capture is introduced. These changes could make for less natural interactions with the patient and foreground data collection and institutional protection over patient concerns. For example, P15 wondered, *"Will I feel the need to say different things like a bunch of legal disclaimers at the end of the conversation?"*

4.2.2 Self-censorship. Physicians also raised concerns about expected changes in communication and behavior in the encounter, by both the physician and the patient. P7 was concerned that

patients may alter on a surface level what they say: *"Will the presence [of the AI] alter what the patient says—now that I write this I think this is HUGE."* Specifically, physicians were concerned that patients would withhold information or "self-censor" during the consultation. This concern was one of the largest shared by physicians in our study. As P3 explained, *"I'm concerned that patients will withhold information [after consenting]."*

This change in communication can affect their relationship with the patient, as P6 described:

> *"The patient knowing they're being listened to [by an AI assistant] is going to affect what they say, and that's going to affect the content of what happens in the encounter. And it's going to affect our relationship. You know the patients ... more likely than not they're going to have a filter or they're not going to disclose, you know, necessarily all the emotions that they have, which can affect not just the the topic or the content of the interaction but the ongoing relationship between me and them."*

Specifically, this self-censorship can impact a physician's response to the patient, as P8 noted:

> *"Even if you don't have conscious concern about whether that data will be get into others' hands—even at the subconscious level, will it affect the patient's willingness to share certain pieces of sensitive information like sexual information, recreational drug use? Much of that affects decision-making and diagnoses. [...] Will this then affect what MD chooses to discuss with patient (especially if not backed by full medical evidence)?"*

We heard from many physicians that they expect to take certain concerns "off-the-record". Physicians described their experiences with this request, noting that some patients would request this even if asked to consent to voice data collection, so that they can speak freely, but others might *not* feel comfortable doing so. As P5, explained:

> *"You know, I'm sure there are certain cases where a patient would say, 'Hey Doc, can you stop [voice] recording right now so I can tell you about this other thing.' And that could be from some domestic dispute or custody issues, drug use. 'I don't want my boss to find out. I don't want my spouse or my partner to find out about this. I don't wanna talk about depression or erectile dysfunction in front of...' I could see a situation where the patient would either not talk about it at all and pretend it wasn't a problem or just say 'can you turn it off right now.' Which would be fine of course. But some people might feel like, 'Oh is that rude or I don't want to do that.' Or it might make some people uncomfortable."*

These findings suggest risks not only of nondisclosure, but also potential exacerbation of existing disclosure challenges (e.g., the documented difficulties that LGBQT+ patients experience with disclosure of aspects of their identity to primary care providers) [96, 158].

4.2.3 Obstructions to care services. Physicians expressed concerns about the impacts that capturing certain verbatim statements from patients could have on those patients. Certain statements, or even summaries resulting from them, can work against patients if misinterpreted by third parties. For example, P11 explained unintended consequences of voice data collection leading to ineligibility for services based on misinterpretation of a verbatim statement captured during the consultation.

> *"Physicians who are well versed in benefits and eligibility [in the U.S.] will pay close attention because patients can say things off-the-cuff that would make them ineligible*

for certain services or programs. For example, [this could happen] if you are a transplant candidate and you're trying to receive an organ or donate an organ, if you say certain things that indicate that your home life is not that stable, even if that's not true, like 'my kids don't like me, I don't think they will come to visit.' But a lot of us know our patients really well. We know that Mr. Jones (pseudonym) is always cantankerous when talking about his son, but his son is actually a caring person who would be able to give him his meds. But if we were to document [Jones' comments] when they did his review [for an organ] that could count against him. You don't want to include things like that!

"For people that are eligible for certain services like hospice or SNF [Skilled Nursing Facility] Care they need to meet certain requirements. I had a patient who used to joke around that he smoked around his oxygen tank. Which we knew he actually did not. But it was a personality thing, he used to rib us. But if a scribe included that in a note, a lot of facilities might not take [the patient]. They might see him as a fire risk. Even though I know as a clinician that the chances of that happening are '0' because he's just being funny.

"That's the hardest thing [about an assistant] that's not a human being. You can see the patient's facial expression and tone of voice. It would be so much harder if you were just seeing or hearing words in a stream."

This risk can, in turn, result in the self-censorship discussed in Section 4.2.2, as P10 described:

"The patient will tell me, 'this is personal, don't write it.' And yes, the things they say can affect application for insurance and disability."

As we discuss more below, physicians described the importance of maintaining control over decisions to document certain statements from patients, rather than a documentation model in which patients' and physicians' comments are automatically captured by default.

4.2.4 Legal risks for patients and clinicians. Some study participants raised potential legal implications if conversational voice data collection produced a "record" of verbatim statements. Currently, clinicians make sure that unwarranted patient statements that could harm them are not documented. These experiences factored into their generation of risks of ramifications of an AI system that collected voice data. As they explained:

"Patients can go back on their word—with a documented record they might be cautious sharing information—especially surrounding their legal matters. This doesn't happen often, but enough with the [unhoused] patient population to be notable."–P10

"It's often off-hand, humorous comments that patients will make with you as a clinician because they know you. You know it's a joke, in person. But to just look at the transcript, [you] might not know from just reading that. Very serious things, like if they make any sort of threats to another person. As a provider you are mandated to report all of those things. That said, most providers don't report things they know aren't true. You hear kids tell parents that 'if you don't get out of here, I'll kill you.' I know they don't mean that. Most physicians will only document things they believe to be true or there is some plausibility of truth to them. If that kid had access to a weapon for example. But in most cases, you know the family well enough to know that they don't mean it literally." –P11

Physicians in our study also brought up questions about whether voice data would be discoverable, or would ever need to be surfaced as part of medicolegal proceedings. For example, P3 was concerned about secondary uses of voice data, wondering, *"[From the patient's standpoint] can the data be used medical-legally for either the physician or myself?"* These concerns call to mind the potential impacts of "data externalities" [48], the phenomenon that some of the data collected about people can reveal information about others. These externalities could impact those directly involved in a consultation, but *unintentional data capture of others* was also mentioned as a risk.

> *"[What about] audio from outside the exam room (registration / med[ical] assistant)—[there] could be unintentional capture of other people/patients/families etc."* –P9

> *"[I have] medicolegal concerns. Can this audio be used as evidence in court in case of dispute of the nature of [the] conversation, after bad patient outcome?"* –P8

Physicians also expected that many patients would want access to their data, and had questions about the patient's legitimate claims to voice data. As P8 pondered, *"[From the patient's standpoint] if my data is being used to build something, can I [the patient] have access to it later?"*

4.2.5 Patients' inability to consent to data collection. It can be difficult to gauge a patient's ability to give consent and study participants considered risks related to potential inability to consent, how they might assess such inability, and the complexities of proxy involvement. P4 told us: *"Many of my patients are severely geriatric or demented (and/or hearing impaired) and this may be difficult to understand, although they may still give consent."* Stating similar concerns, P6 mentioned, *"Patients may not understand what the data will be used for."*

Participants expected that proxy involvement will be necessary for many patients, yet processes for their inclusion will be non-trivial. As P16 explained, *"What if the patient isn't 'consentable'? Will family consent for them? Will that require a witness?"* P4 described how inclusion challenges can be exacerbated when proxies are unavailable or the need for proxies goes unrecognized:

> *"Sometimes you have people show up who just have some mild to moderate cognitive impairment and they don't actually have family with them because either, you know, it's not recognized, or it is—to some extent—but the patient is still functional. So you could see getting into some discussions with them if they didn't totally understand it."*

They further explained that even when proxies are involved for patients with limited ability to provide consent themselves, following a legal process alone is not enough to ensure a patient-centered consent process:

> *"There are issues around consent for children or for adults with developmental delay/cognitive impairment. (Legally I imagine [one] would follow the same process as any other consent, but the perception from the caregivers / surrogate / guardian may be different in this case [of voice data capture].)"*–P9

4.2.6 Workflow disruptions and additional clinician labor. Some of the prospective risks of data collection included impacts on the care team's workflow, which could, in turn, negatively impact both their and the patient's experience. In particular, physicians discussed the patient consent journey and what portions would be handled by medical staff. This was important because they thought patients would turn to them (their care provider) to answer questions about consent and the AI technology, leading to concerns that time would be taken away from the health consultation in order to address these questions. P1 summarized this concern: *"Will the patient have a lot of*

follow-up questions (i.e., will this disrupt the visit)?"

> *"Who answers all the technical details of what should happen outside of the visit (e.g., how/where [data] will be stored, used, discarded, accessed?) How does partial consent work (e.g., patient mentions something but wants it stricken from the record)?"* –P2
>
> *"Reading and signing the consent may slow down the office workflow (check-in process) even more. As it is, the insurance paperwork takes up a lot of time."* –P6
>
> *"Communication with [the] MA [Medical Assistant]—what if the patient has questions and the MA needs to follow up with [the] physician first in order to answer or address concerns? Workflow and added time leads to less time with patient."*-P4

4.2.7 Privacy. Physicians thought patients would be concerned about insurance companies gaining access to the data. They saw both the status of third-party access to data and information about any data controls available as vital types of information to disclose. Information about data access was the first concern P8 mentioned in the study: *"Privacy. Will audio [data] get into unexpected hands (e.g., insurance, employer) and negatively affect my patients?"* Others discussed the complexity of explaining access to voice data to patients who would be asked to consent to its collection, as P4 described:

> *"[Patients] are going to want to know...for the consent, who actually gets to be able to listen and look at that [data]. Because there's so many people who look at a patient chart, from billing, to panel coordinator, to the referral person, to the front desk, to the physician, MA, other people's MAs. You know, if my doc's out (or if I'm out) and the MAs are just trying to help look in my inbox, then other MAs will touch the chart too."*

Other participants noted that insurance company access in particular may be of concern to patients.

> *"One thing in terms of security I think a lot of the patients would be very concerned about insurance companies getting access to this, and it affecting their coverage, which is, you know, an unfortunate reality—or an unfortunate, realistic concern."* –P7

Meanwhile, P9 mentioned that the consequences of sharing information with third parties would be especially concerning to those involved in partner violence:

> *"[There is] concern over things patients want to discuss but do not want documented in the [EHR]. Sometimes this might be a safety issue, such as intimate partner violence."*

4.2.8 Inaccuracies in speech recognition and documentation. Most participants (n=13) discussed the importance of accuracy of voice recognition and language models, and foresaw potential ways in which these technologies could fail for both clinicians and patients. For example, P7 stressed that sub-optimal speech recognition accuracy could impact not just the resulting data, but other aspects of care:

> *"The [speech recognition] needs to be VERY close to 100% accurate—that is, if I am spending time editing for inaccuracies, why consent at all?"*

From the perspective of the patient, differential accuracy—due to using an interpreter or being a non-native language speaker for example—risks quality of service harms.

> *"I'm concerned about issues around interpretation ([i.e.,] how would this [technology] work for people using a telephonic interpreter, or an iPad interpreter, ESL [English as a Second Language])?"* -P9

Physicians also expected that language models using recognized speech might incorrectly segment speech, identifying patients' suggestions to be physicians' diagnoses, and might not be able to differentiate between *possibilities* offered at the beginning of a consultation and final *assessments*.

> *"We could be [discussing] the history section [of the visit]. They could say, 'my asthma' and that could get sort of flagged as "problem: Asthma" whereas the physician would know OK, actually this is a constellation of symptoms that sort of mimics the effects that you think is asthma."* -P16

> *"If my working diagnosis changes as new information comes in, my recommendation might change. So what I said at the beginning of the encounter would not be accurate anymore."* –P9

5 DISCUSSION

Our discussion returns to the research questions that guided our study. In particular, we reflect on our choice to position an informed consent process as a design fiction, as a means to integrate ethics frameworks with speculative design research. We discuss the ways in which positioning an informed consent process as a fiction enabled us to elicit context and narrative that embedded clinical caregivers' values around AI-related voice data collection in consultations, and situated engagement with ethical questions related to the potential benefits and harms of using this data to train ML models to assist clinicians. We discuss opacity and transparency for AI in clinical health scenarios in light of our findings. We conclude by considering the prospective risks of harm clinicians surfaced, how they extend those documented in the literature identifying AI/ML harms, and how evaluation criteria for AI-powered documentation can better reflect them.

5.1 The informed consent process as a design fiction

Research conversations at the intersection of AI ethics and CSCW have highlighted the growing importance of incorporating perspectives of impacted stakeholders in AI research [1, 53, 156]. In this paper, we extend these discussions to consider how speculative design research, using design fiction, offers a lens to understand values around AI-related data collection for dataset construction, model training, and real-time use. Although many prior speculative design inquiries have engaged with issues of ethics and bioethics [54, 84], such engagement has often taken place in the context of informal learning environments and research institutions. Methodologically, our use of design fiction contributes to a growing body of work that situates speculative approaches in specific *communities* and *institutions* of practice, directly engaging decision-makers and those whose daily lives and work will be affected by those technologies [91, 160]. While a speculative design approach can be used to imagine futures of entirely new technology landscapes [25], it is often used to intentionally foreground a mundane cultural artifact [79]. We chose to foreground the mundane artifacts of informed consent in order to build upon familiarity with clinicians' lived experience of patient consent processes. Informed consent, as both a speculative process and principle, supported critical consideration of near-term future voice data collection scenarios. Grounding speculative activities in the context of consent associated with this near-term scenario helped situate clinicians' generative thinking about risks and benefits in their real-world context, to base expected possibilities on their experiences. A distinctive aspect of our work is using design fiction as a prompt to elicit insights from practicing clinicians, and the subsequent interpretation of those insights into design considerations for their practical context.

Using the informed consent process as design fiction was also valuable because: 1) voice data collection in clinical settings is still an emerging practice, neither widely available nor accepted as common practice, and 2) the components of design fictions (world building, creating discursive

space [102]) afforded more contextualized feedback than we would expect to elicit through traditional scenario-based methods. Other methods such as interviews and surveys could be limited in capturing perceptions rather than engaging stakeholders in productive, nuanced discussions that allowed free-form, generative conceptions of the numerous risks associated with multi-modal data collection in our domain.

As demonstrated in our study, the types of information surfaced through this approach can equip teams with perspectives of stakeholders, as they pursue answers to vital questions such as: is the current direction of our research and development ethically viable? What other efforts might be needed to gain directional guidance? To address any prospective risks surfaced in early studies, how might we think through the design and testing of safeguards? Next, we detail selected findings that demonstrate the importance of these questions.

5.2 Implications for human decision-making with and about AI in clinical settings

Prior studies have found that a lack of representative data, AI-relevant literacy, and consistent standards to guide the use of AI in health scenarios can lead to notable gaps in how risks and benefits of AI technology are calculated [40, 112]. However, many of these problems are connected to open challenges in ML and health and medical informatics. In this section, we discuss these challenges, grounding them in recent research and discussing the known, open areas of research still needed.

5.2.1 Factors affecting clinicians' decisions to use AI assistance. Clinicians in our study echoed the importance of gaining knowledge at the outset of an AI development project to better identify potential harms related to uses of data (Sections 4.2.3, 4.2.4, and 4.2.7), as well as risks associated with the processes and context surrounding data collection (Sections 4.2.1, 4.2.2, 4.2.5, and 4.2.8) including the impacts of changes to workflow (Section 4.2.6). Yet, AI literacy about these and other aspects of new forms of AI could depend, in part, on how we enable transparency about AI systems in context, and thus depends on progress of an active area of current research (we elaborate further on this research area below in 5.2.2).

Turning to participants' concerns about how patients may be negatively affected by decision-making based on inaccurate or unintentional voice data collection (e.g., risks of service ineligibility (4.2.3), legal implications (4.2.4), privacy (4.2.7)), it will be important for systems to foster transparency and explanatory techniques regarding how decisions that stem from any voice data input were reached, and allow patients or clinicians to correct misinterpretations. In the remainder of this section, we highlight the importance and nuanced complexities of obtaining better AI explanations for decision-making in clinical settings.

Jacobs et al.'s [2021] work with primary care physicians to co-design decision support tools found that ML model explanations were vital in clinicians' decision to use the tools, but that current trends in explainable AI do not necessarily suit clinical environments and use cases. This finding is not surprising in light of Cai et al.'s [2019] discovery that pathologists using a deep-learning-based AI cancer grading assistant desired more *global* system and model attributes (above explanations of local model decisions). These attributes included the strengths and weaknesses of model performance, both overall and on specific cases, relative to human performance, and whether and how the model inputs differ from information a pathologist might learn from and consult [30].

These findings reinforce the importance of the ability of AI systems to foster such understanding and awareness of model decisions and capabilities. However, prior approaches to producing AI explanations for clinical decision support, attempting to respond to clinicians' needs, have struggled or failed in practice, due in part to their overemphasis on explaining how an outcome was reached and prior system performance (see Yang et al. [2023] for a review). Thus, an active area

of research now seeks to identify how AI-powered decision-support designs can more effectively calibrate clinicians' trust in AI suggestions, on a case-by-case basis [162]. The next section discusses challenges researchers face in achieving these abilities.

5.2.2 AI explanations, user control, and transparency in clinical contexts. The inherent difficulties in understanding and explaining how AI systems work is an established issue, referred to more generally as *opacity* of the system (e.g., "interpretation" itself is currently under-specified) [103, 137]. Ferretti et al. [2018] found three different semantic dimensions of opacity particularly relevant in health contexts: *lack of disclosure* (i.e., data contributors and subjects of AI decisions are unaware of automated decision-making activities, undermining the relationship between patients and clinical caregivers), *epistemic opacity*, or inaccessible information or insufficient understanding, possibly across multiple parties in a sociotechnical AI ecosystem, and *explanatory opacity* (i.e., the rationale behind specific outcomes) [59].

Our study suggests that alongside the availability of system outcome rationale and disclosure of automated decision-making as forms of transparency, *data transparency* is a fundamental need: explanatory transparency of implications of voice data collection and—in individual collection contexts—which data is recognized and how. Yet, as Brewer and Kameswaran [2018] found, if a system is solely voice- or audio-based, it is difficult to confirm that it is doing what it purports to, especially for people with disabilities and older adults. This finding suggests the importance of future research concerned with data and system transparency in multiple forms: from the composition of a data set [67, 131] to methods for interactive, user-controllable, and multi-modal forms of system output. As a starting point for advancing these methods, researchers could explore similarities with prior work on multi-modal design for clinical information applications, including studies of human trust in different types of system-vocalized explanations [55], interaction techniques for visual analysis of data associated with diagnostic screening decision-making [31, 32, 34], and cross-cultural design considerations [100].

Turning now to epistemic opacity and its concern with accessible information, our findings also highlight the potential for "too much information" associated with an AI explanation, which clinicians have historically found can hinder care, given the time it takes them to reason about model output explanations [77]. This issue of having "too little time" in practice to reason about explanations is important, as it reinforces that decision-making in clinical settings is not based on access to information alone. Our findings on workflow considerations (Section 4.2.6) extend these and other findings from Burgess et al. [2023], who studied a conceptual prototype to demonstrate plausible AI-generated treatment insights for diabetes medications, eliciting attitudes toward the AI prototype from U.S.-based clinicians. They found that introducing AI-generated recommendations for care that require time for clinician review during the patient encounter was unacceptable. Physician participants reported being unable to validate AI-generated insights in the context of their patient consultations [29]. Our study surfaced physician concerns that the work required to adequately manage decisions around use of an AI assistant would, in fact, take time away from the health consultation and add to already-burdensome information review.

The lessons from these studies heighten the importance of attending to the ways in which model explanations or transparency artifacts are in fact situated in complex sociotechnical contexts, and suggest that a focus on producing information alone is insufficient for decision-making about, and with, AI systems in clinical settings. These reflections echo those of researchers of automated decision systems (ADS) (a class of AI systems used for business operations) who have recently argued for contextual approaches to transparency for these systems) [144]. Next, we discuss the ways in which knowledge of the clinical sociotechnical context enabled us to surface new

dimensions of risk of harm that can inform further research directions, as well as approaches to transparency and explanatory strategies.

5.3 Use of design fiction to discover risks of potential AI harms

While AI/ML approaches present opportunities to assist health care professionals and *potentially* improve health outcomes for patients, they bring a range ethical concerns [10, 71, 108]. These concerns include, but are not limited to, racial and gender disparities, accessibility of clinical trials, and exacerbation of bias in medical practices[40]. Our use of informed consent as a design fiction distilled findings that expand the notion of explanatory opacity challenges for this AI application space, which are exacerbated for conversational voice data. Voice recognition models themselves mediate downstream tasks, making physicians' concerns about accuracy and reliability even more pronounced [69]. For example, problems could be pronounced for people with speech differences such dysarthria, deaf accent, etc., since training data does not typically include samples from such populations [80, 152]. Additionally, gender and dialect biases have been found in commercial systems [109, 149].

Speech recognition and conversation modeling currently also misunderstand sarcasm and humor [42, 46, 104], lack access to the longer-term conversation context [104] and fail to discern accents, tonalities, rhythmic variations, and speech patterns across subgroups, which impacts performance [87]. Thus, transparency around these concerns, as well as how data is maintained [132], and transformed to clinical data entries—are all candidates for data transparency that go beyond data transparency artifacts in use today and call for future research in this area to avoid harm.

However, as Boyarskaya et al. [2020], Katell et al.[2020], and Shelby et al.[2023] note, to understand possible negative impacts of particular pieces of technical work, our field needs to look beyond currently identified classes of harms: *allocational* (unfairly assigned opportunities or resources) or *representational* (discriminatory or incomplete data or algorithmic depictions), and *quality of service* disparities, to the open, context dependent, and unobservable nature of harms, some of which are not suited to algorithmic "fixes". While these classes of harms are amenable to evaluation with existing machine learning epistemologies and methods (e.g., disaggregated data analyses, adversarial testing and model experimentation), they do not capture the breadth of contextual and material downstream harms that must be mitigated or prevented.

Our study surfaced risks of potential allocational and representational harms (e.g., interpretations, intended or not, of data leading to ineligibility for services or other obstructions to care) and quality-of-service harms (e.g., accuracy concerns due to language differences), yet our findings around medicolegal implications, impacts of data collection on patient care, the patient–clinician relationship, and added work for clinicians, suggest types of risks that extend prior conceptions of computational harms.

Our use of design fiction enabled us to identify a wider range of potential AI harms, each of which represent important areas for further systematic investigation. This is especially important for the health context, which requires research methods that enable computing researchers to see relationships between data and models through multiple lenses. For example, patients and clinicians already inhabit an asymmetrical power relationship [78], and physicians in our study were concerned that the introduction of additional data collection for AI systems could create conditions for these asymmetries to expand (prior research has found that power dynamics can more generally impact the relationship between those whose data is collected and those who make use of that data) [52, 110]. Study participants stressed the importance of attending to these power differentials, as a vital part of supporting patient decision-making and trust. As P9 raised, trust in the health care system has been eroded through previous mistreatment of people in research and

technological innovation, and experiences of racism and other forms of discrimination by Western institutions [135].

These factors highlight historical contexts that could impact how patients and their informal caregivers perceive the introduction of AI-mediated care. Our study surfaced physicians' concerns that conversational voice data collected to train and interact with AI assistants, might impact their communication with patients. Physicians mentioned experiences with patients requesting to speak "off the record" in their consultations; thus, an important area of future work should examine, with patients, their comfort level with ambient voice data capture to understand whether and how information they would otherwise be forthcoming about might be withheld. This is important to study further, because the risk of an implicit silencing of concerns in a setting meant to provide confidential care is especially deserving of attention and safeguards, in particular because a decision by a patient to withhold information could be invisible to their clinical caregivers.

Physicians also had concerns about the consent conversation itself consuming too much of the visit. These concerns are consistent with prior research that when data collection and use of AI are foregrounded in the consultation, people seeking care could be left with a sense that they are "bearers of medical data" rather than seen holistically [5].

As we described in depth in 5.1, our use of design fiction enabled us to identify a wider range of potential AI harms that are harder to distill through harm mitigation strategies currently emphasized in the AI literature. Next, we discuss how the concerns we surfaced contrast with the types of concerns and evaluative metrics that ML-driven health informatics have attended to regarding voice-enable clinical documentation.

5.4 Reconsidering evaluation criteria for AI-assisted documentation

To consider our findings with respect to research on voice data collection to automate clinical documentation more broadly, we turn to early work in medical informatics calling for better automated documentation practices, noting voice recognition as an example [49]. This and related work often cites *clinician burden* (e.g., [124]) and more *comprehensive notes* [12] as motivations for voice data collection, without acknowledging the tensions that participants in our study acknowledged (e.g., self-censorship (4.2.2), accuracy concerns (4.2.8)) that could burden patients and clinicians, nor the additional work that can be introduced, negating the intended benefits of automation (4.2.6).

Other work describes improvements in provider *satisfaction, documentation quality* and *efficiency* [133] or *reduced time* [83] when using speech recognition to support documentation and transcription. However, these narratives prioritize efficiency and cost with limited or no critique of the impacts of language model accuracy (4.2.8), patient comfort (4.2.2), or ethics surrounding consent to allow ambient voice data capture (4.2.5) in the first place. Some research does describe ways in which ASR-generated notes are *less organized* and *more verbose* than traditional note-taking [83], have *high error rates* and are *less accurate* with health concepts [90]. However, future work calls for improved accuracy in terms of word error rates without acknowledging differences between lexical dictation accuracy and semantic intent (e.g., sarcasm), as participants mentioned in our study (4.2.3).

Quiroz et al. [2019] critique the "digital scribe" in clinical settings, noting challenges, *"recording high-quality audio, converting audio to transcripts using speech recognition, inducing topic structure from conversation data, extracting medical concepts, generating clinically meaningful summaries of conversations, and obtaining clinical data for AI and ML algorithms."* [126]. They call for more attention to how these scribes will impact clinician–patient communication and relationships in health care settings. A literature review on speech-assisted documentation also highlights that there is a focus on limited model accuracy metrics, with less research on document quality, patient safety, usability, and the need for better types of error detection [19]). In Bossen and Pine's [2023]

studies of Clinical Documentation Integrity Specialists (CDIS), who review patient charts in near real-time to improve clinicians' documentation, specialists found that AI-generated medical coding suggestions (part of clinical documentation) were often inaccurate, and they identified very specific conditions to be met for effective "human–AI collaboration" in clinical documentation, including CDIS' awareness of and knowledge about the (un)reliability of the AI-generated suggestions. Our study findings related to accuracy (see 4.2.8) echo and extend these concerns.

Concerns about voice data collection could also vary cross-culturally and globally, yet AI systems for health—developed in the Western region of the world—have a troubled history considering these contexts [15]. Since AI performance for health-care-related tasks has been shown to generalize poorly to populations outside of those whose data was used to train and validate the algorithms or models [38], we expect that populations in regions with more infrastructure, human expertise, and audio data available for training (with the storage, privacy, and security capabilities that entails) would experience performance gains, while those without could experience further entrenchment of limited health care access and health disparities. Thus, advancing equity-oriented evaluations of AI/ML for clinical use cases is a vital area of future work.

5.5 Limitations and future work

Risks associated with the introduction of new technologies (and processes surrounding them) in high-stakes domains like health are characterized by a confluence of factors that cut across ecosystems of technology development and use [8]. We do not expect that our findings, nor the methods we used here alone, are sufficient to account for the entirety of possible risks associated with new forms of data collection or new types of AI technologies. First, general use of speculative approaches do not necessarily provide a basis for ethical design. Speculative design research—though seen as a critical methodological intervention today for its ability to challenge dominant practices in technology—was introduced to computing research by corporate design research initiatives to showcase more conventional technology concepts [161]. The "criticality" is not inherent in the method itself, but in the orientation and praxis around its application. Researcher acknowledgement of positionality and commitment to reflexivity requires us to ask: who has the opportunity to speculate? Whose lives are we imagining? Whose challenges are we representing? Who is excluded? [155]. Participatory approaches to speculative research can surface a diversity of hopes and concerns for people, lest they lack equitable implementation, treat participants as a "resource" or burden them in the research process itself without mutual benefit [125].

We found that including subject matter experts, who bring lived experience to the natural setting and sociotechnical system in which an AI technology would be embedded, was necessary to provide more context awareness and a wider view of potential harms, which can serve as foundations for designing subsequent studies with patients, formal and informal caregivers, and other affected stakeholders. As Katell et al. [2020] note, the current framing of harms of algorithmic bias suggests solutions focused on data accuracy, when in fact deeper questions—that are highly *situated*—concern whether some technologies should be developed at all. Thus, participatory design methods that specifically enable participants' sharing of situated knowledge is paramount to elucidating a broader set of prospective harms.

While our study uncovered risks of voice-enabled documentation assistance that clinicians envision, further exploration is needed to connect issues of risk with evidence of patient and clinician harm, trust in AI decisions, and adoption of AI assistance for clinical documentation. These efforts could begin by considering work linking perceived risk, trust, and adoption, for other types of intelligent agents [33]. Similar studies that cover a variety of stakeholders in the health information and AI ecosystem—from patients and their informal caregivers, to those responsible

for working with data, to a range of clinical caregivers that participate in the patient's care—are a necessary complement to our work.

6 CONCLUSION

We described how we engaged clinicians in a speculative exercise in which we used elements of an *informed consent process* as a *design fiction*, to elicit their perspectives on prospective benefits and risks of a new form of data collection—conversational voice data—for use by a plausible, near-term AI documentation assistant. This study adds a vital complement to the academic literature—and larger surrounding narratives—about the expected benefits to clinicians of such technologies. Our study surfaced both prospective benefits of collecting voice data in health consultations, along with eight classes of risks. Risks that clinicians were concerned about included, in part, disruptions to workflow, self-censorship that can impact care, and errors that could obstruct patient access to care services. Some of the potential risks clinicians identified have thus far been invisible to AI researchers and developers, and could be difficult to identify directly in the context of AI assistant usage in practice, pointing to the importance of the use of our approach to identify risks of potential harms of emerging AI technologies. Our work suggests new dimensions of opacity and transparency for AI in clinical health scenarios, and ways that we can reconsider evaluation criteria for AI-assistance for clinical tasks in light of our findings. We hope that findings from this study encourage adoption of similar studies in other domains, as early as possible in an AI project, to gain insights into prospective benefits and risks that can guide ethics thinking and responsiveness to issues and concerns.

ACKNOWLEDGMENTS

We thank all study participants. We thank Carl DiSalvo, Renee Shelby, and Khai N. Truong for valuable feedback. We thank Marissa J. Smith for research support.

REFERENCES

[1] NIST AI. 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* Retrieved April 3, 2023 from https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook

[2] Priscilla Alderson and C.F. Goodey. 1998. Theories of Consent. *BMJ (Clinical research ed.)* 317 (12 1998), 1313–5.

[3] M. Arnold, R. K. E. Bellamy, M. Hind, S. Houde, S. Mehta, A. Mojsilović, R. Nair, K. Natesan Ramamurthy, A. Olteanu, D. Piorkowski, D. Reimer, J. Richards, J. Tsay, and K. R. Varshney. 2019. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development* 63, 4/5 (2019), 6:1–6:13. https://doi.org/10.1147/JRD.2019.2942288

[4] Paula Astrid Mendez Gonzalez, Sofía Castañeda Mosquera, María Paula Bernal Tinjaca, Ricardo Mejía Sarmiento, Roberto Alejandro Morales Rubio, Juan Camilo Giraldo Manrique, and Santiago Baquero Lozano. 2020. Participatory Construction of Futures for the Defense of Human Rights. In *Proceedings of the 16th Participatory Design Conference 2020 - Participation(s) Otherwise - Volume 2* (Manizales, Colombia) *(PDC '20)*. Association for Computing Machinery, New York, NY, USA, 10–16. https://doi.org/10.1145/3384772.3385155

[5] Kristina Astromskė, Eimantas Peičius, and Paulius Astromskis. 2020. Ethical and legal challenges of informed consent applying artificial intelligence in medical diagnostic consultations. *AI & SOCIETY* (2020), 1–12. https://doi.org/10.1007/s00146-020-01008-9

[6] James Auger. 2013. Speculative design: crafting the speculation. *Digital Creativity* 24, 1 (2013), 11–35. https://doi.org/10.1080/14626268.2013.767276

[7] James Auger. 2014. Living with robots: A speculative design approach. *Journal of Human-Robot Interaction* 3, 1 (2014), 20–42. https://doi.org/10.5898/JHRI.3.1.Auger

[8] Thomas Christian Bächle and Alina Wernick. 2019. *The futures of eHealth.: Social, ethical and legal challenges.* Alexander von Humboldt Institute for Internet and Society.

[9] Stephanie Ballard, Karen M Chappell, and Kristen Kennedy. 2019. Judgment call the game: Using value sensitive design and design fiction to surface ethical concerns related to technology. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 421–433. https://doi.org/10.1145/3322276.3323697

[10] Imon Banerjee, Ananth Reddy Bhimireddy, John L Burns, Leo Anthony Celi, Li-Ching Chen, Ramon Correa, Natalie Dullerud, Marzyeh Ghassemi, Shih-Cheng Huang, Po-Chih Kuo, et al. 2021. Reading Race: AI Recognises Patient's Racial Identity In Medical Images. *arXiv preprint arXiv:2107.10356* (2021).

[11] Jørgen P Bansler, Erling C Havn, Kjeld Schmidt, Troels Mønsted, Helen Høgh Petersen, and Jesper Hastrup Svendsen. 2016. Cooperative epistemic work in medical practice: an analysis of physicians' clinical notes. *Computer Supported Cooperative Work (CSCW)* 25, 6 (2016), 503–546. https://doi.org/10.1007/s10606-016-9261-x

[12] Paul J Barr, James Ryan, and Nicholas C Jacobson. 2021. Precision Assessment of COVID-19 Phenotypes Using Large-Scale Clinic Visit Audio Recordings: Harnessing the Power of Patient Voice. *J Med Internet Res* 23, 2 (19 Feb 2021), e20545. https://doi.org/10.2196/20545

[13] Eric PS Baumer, Mark Blythe, and Theresa Jean Tanenbaum. 2020. Evaluating design fiction: The right tool for the job. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 1901–1913. https://doi.org/10.1145/3357236.3395464

[14] Rainer S Beck, Rebecca Daughtridge, and Philip D Sloane. 2002. Physician-patient communication in the primary care office: a systematic review. *The Journal of the American Board of Family Practice* 15, 1 (2002), 25–38.

[15] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M Vardoulakis. 2020. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12. https://doi.org/10.1145/3313831.3376718

[16] Ruha Benjamin. 2016. Informed refusal: Toward a justice-based bioethics. *Science, Technology, & Human Values* 41, 6 (2016), 967–990.

[17] Sarah Bird, Solon Barocas, Kate Crawford, Fernando Diaz, and Hanna Wallach. 2016. Exploring or Exploiting? Social and Ethical Implications of Autonomous Experimentation in AI. In *Workshop on Fairness, Accountability, and Transparency in Machine Learning*.

[18] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the People? Opportunities and Challenges for Participatory AI. In *Equity and Access in Algorithms, Mechanisms, and Optimization* (Arlington, VA, USA) *(EAAMO '22)*. Association for Computing Machinery, New York, NY, USA, Article 6, 8 pages. https://doi.org/10.1145/3551624.3555290

[19] Suzanne V Blackley, Jessica Huynh, Liqin Wang, Zfania Korach, and Li Zhou. 2019. Speech recognition for clinical documentation from 1990 to 2018: a systematic review. *Journal of the American Medical Informatics Association* 26, 4 (2019), 324–338. https://doi.org/10.1093/jamia/ocy179

[20] Claus Bossen and Kathleen H. Pine. 2023. Batman and Robin in Healthcare Knowledge Work: Human-AI Collaboration by Clinical Documentation Integrity Specialists. *ACM Trans. Comput.-Hum. Interact.* 30, 2, Article 26 (mar 2023), 29 pages. https://doi.org/10.1145/3569892

[21] Margarita Boyarskaya, Alexandra Olteanu, and Kate Crawford. 2020. Overcoming Failures of Imagination in AI Infused System Development and Deployment. *arXiv preprint arXiv:2011.13416* (2020).

[22] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

[23] Virginia Braun and Victoria Clarke. 2020. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology* (2020), 1–25. https://doi.org/10.1080/14780887.2020.1769238

[24] Kirsten Bray and Christina Harrington. 2021. *Speculative Blackness: Considering Afrofuturism in the Creation of Inclusive Speculative Design Probes.* Association for Computing Machinery, New York, NY, USA, 1793–1806. https://doi.org/10.1145/3461778.3462002

[25] Kirsten E Bray, Christina Harrington, Andrea G Parker, N'Deye Diakhate, and Jennifer Roberts. 2022. Radical Futures: Supporting Community-Led Design Engagements through an Afrofuturist Speculative Design Toolkit. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 452, 13 pages. https://doi.org/10.1145/3491102.3501945

[26] Robin N. Brewer and Vaishnav Kameswaran. 2018. Understanding the Power of Control in Autonomous Vehicles for People with Vision Impairment. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 185–197. https://doi.org/10.1145/3234695.3236347

[27] Barry Brown, Julian Bleecker, Marco D'Adamo, Pedro Ferreira, Joakim Formo, Mareike Glöss, Maria Holm, Kristina Höök, Eva-Carin Banka Johnson, Emil Kaburuan, Anna Karlsson, Elsa Vaara, Jarmo Laaksolahti, Airi Lampinen, Lucian Leahu, Vincent Lewandowski, Donald McMillan, Anders Mellbratt, Johanna Mercurio, Cristian Norlin, Nicolas Nova, Stefania Pizza, Asreen Rostami, Mårten Sundquist, Konrad Tollmar, Vasiliki Tsaknaki, Jinyi Wang, Charles Windlin, and Mikael Ydholm. 2016. The IKEA Catalogue: Design Fiction in Academic and Industrial Collaborations. In *Proceedings of the 19th International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '16)*. Association for Computing Machinery, New York, NY, USA, 335–344. https://doi.org/10.1145/2957276.2957298

[28] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

[29] Eleanor R. Burgess, Ivana Jankovic, Melissa Austin, Nancy Cai, Adela Kapuundefinedcińska, Suzanne Currie, J. Marc Overhage, Erika S Poole, and Jofish Kaye. 2023. Healthcare AI Treatment Decision Support: Design Principles to Enhance Clinician Adoption and Trust. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 15, 19 pages. https://doi.org/10.1145/3544548.3581251

[30] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 104 (nov 2019), 24 pages. https://doi.org/10.1145/3359206

[31] Francisco M. Calisto, Alfredo Ferreira, Jacinto C. Nascimento, and Daniel Gonçalves. 2017. Towards Touch-Based Medical Image Diagnosis Annotation. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces* (Brighton, United Kingdom) *(ISS '17)*. Association for Computing Machinery, New York, NY, USA, 390–395. https://doi.org/10.1145/3132272.3134111

[32] Francisco Maria Calisto, Nuno Nunes, and Jacinto C. Nascimento. 2020. BreastScreening: On the Use of Multi-Modality in Medical Imaging Diagnosis. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Salerno, Italy) *(AVI '20)*. Association for Computing Machinery, New York, NY, USA, Article 49, 5 pages. https://doi.org/10.1145/3399715.3399744

[33] Francisco Maria Calisto, Nuno Nunes, and Jacinto C. Nascimento. 2022. Modeling adoption of intelligent agents in medical imaging. *International Journal of Human-Computer Studies* 168 (2022), 102922. https://doi.org/10.1016/j.ijhcs.2022.102922

[34] Francisco Maria Calisto, Carlos Santiago, Nuno Nunes, and Jacinto C. Nascimento. 2021. Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification. *International Journal of Human-Computer Studies* 150 (2021), 102607. https://doi.org/10.1016/j.ijhcs.2021.102607

[35] Hillary Carey, Alexandra To, Jessica Hammer, and Geoff Kaufman. 2020. Fictional, Interactive Narrative as a Foundation to Talk about Racism. In *Companion Publication of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS' 20 Companion)*. Association for Computing Machinery, New York, NY, USA, 171–177. https://doi.org/10.1145/3393914.3395885

[36] Benedetta Catanzariti, Srravya Chandhiramowuli, Suha Mohamed, Sarayu Natarajan, Shantanu Prabhat, Noopur Raval, Alex S Taylor, and Ding Wang. 2021. The Global Labours of AI and Data Intensive Systems. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. 319–322.

[37] CDC. 2021. *Practical Guidance for Studies Using Freelisting Interviews.* Retrieved April 3, 2023 from https://www.cdc.gov/pcd/issues/2021/20_0355.htm

[38] Leo Anthony Celi, Jacqueline Cellini, Marie-Laure Charpignon, Edward Christopher Dee, Franck Dernoncourt, Rene Eber, William Greig Mitchell, Lama Moukheiber, Julian Schirmer, Julia Situ, et al. 2022. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLOS Digital Health* 1, 3 (2022), e0000022. https://doi.org/10.1371/journal.pdig.0000022

[39] Cerner. 2021. *Harnessing AI to Enhance Clinical Documentation Improvement.* Retrieved April 3, 2023 from https://www.cerner.com/-/media/cerner-media-united-states/landing-pages/harnessing-ai/20200326himsscernerwep-final.aspx

[40] Irene Y. Chen, Emma Pierson, Sherri Rose, Shalmali Joshi, Kadija Ferryman, and Marzyeh Ghassemi. 2021. Ethical Machine Learning in Healthcare. *Annual Review of Biomedical Data Science* 4, 1 (2021), 123–144. https://doi.org/10.1146/annurev-biodatasci-092820-114757 arXiv:https://doi.org/10.1146/annurev-biodatasci-092820-114757

[41] EunJeong Cheon and Norman Makoto Su. 2018. Futuristic autobiographies: Weaving participant narratives to elicit values around robots. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 388–397. https://doi.org/10.1145/3171221.3171244

[42] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. *What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300705

[43] Enrico Coiera, Baki Kocaballi, John Halamka, and Liliana Laranjo. 2018. The digital scribe. *NPJ digital medicine* 1, 1 (2018), 1–5. https://doi.org/10.1038/s41746-018-0066-9

[44] Ned Cooper, Tiffanie Horne, Gillian R Hayes, Courtney Heldreth, Michal Lahav, Jess Holbrook, and Lauren Wilcox. 2022. A Systematic Review and Thematic Analysis of Community-Collaborative Approaches to Computing Research. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 73, 18 pages. https://doi.org/10.1145/3491102.

3517716
[45] Oonagh Corrigan. 2003. Empty ethics: the problem with informed consent. *Sociology of Health & Illness* 25, 7 (2003), 768–792. https://doi.org/10.1046/j.1467-9566.2003.00369.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1467-9566.2003.00369.x
[46] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can i Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 43, 12 pages. https://doi.org/10.1145/3098279.3098539
[47] Peter J. Craigon, Justin Sacks, Steve Brewer, Jeremy Frey, Anabel Gutierrez, Naomi Jacobs, Samantha Kanza, Louise Manning, Samuel Munday, Alexsis Wintour, and Simon Pearson. 2023. Ethics by design: Responsible research innovation for AI in the food sector. *Journal of Responsible Technology* 13 (2023), 100051. https://doi.org/10.1016/j.jrt.2022.100051
[48] Charles Cui, Andreas Haupt, Mihaela Curmei, Yixin Wang, and Rediet Abebe. 2021. *The Sign of Data Externalities*. Retrieved April 3, 2023 from https://www.md4sg.com/workshop/faact21/the_sign_of_data_externalities.pdf
[49] Caitlin M Cusack, George Hripcsak, Meryl Bloomrosen, S Trent Rosenbloom, Charlotte A Weaver, Adam Wright, David K Vawdrey, Jim Walker, and Lena Mamykina. 2012. The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. *Journal of the American Medical Informatics Association* 20, 1 (09 2012), 134–140. https://doi.org/10.1136/amiajnl-2012-001093
[50] deepscribe.ai. 2021. *Demo*. Retrieved April 3, 2023 from https://www.deepscribe.ai/demo
[51] Marcela G Del Carmen and Steven Joffe. 2005. Informed consent for medical treatment and research: a review. *The Oncologist* 10, 8 (2005), 636–641. https://doi.org/10.1634/theoncologist.10-8-636
[52] Catherine D'ignazio and Lauren F Klein. 2020. *Data Feminism*. MIT press.
[53] Joseph Donia and James A. Shaw. 2021. Co-design and ethical artificial intelligence for health: An agenda for critical research and practice. *Big Data & Society* 8, 2 (2021), 20539517211065248. https://doi.org/10.1177/20539517211065248
[54] Anthony Dunne and Fiona Raby. 2013. *Speculative everything: design, fiction, and social dreaming*. MIT press.
[55] Chad Edwards, Autumn Edwards, Brett Stoll, Xialing Lin, and Noelle Massey. 2019. Evaluations of an artificial intelligence instructor's voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior* 90 (2019), 357–362. https://doi.org/10.1016/j.chb.2018.08.027
[56] Glyn Elwyn, Paul James Barr, and Mary Castaldo. 2017. Can patients make recordings of medical encounters?: what does the law say? *JAMA* 318, 6 (2017), 513–514. https://doi.org/10.1001/jama.2017.7511
[57] Glyn Elwyn, Paul James Barr, and Stuart W Grande. 2015. Patients recording clinical encounters: a path to empowerment? Assessment by mixed methods. *BMJ open* 5, 8 (2015), e008566. https://doi.org/10.1136/bmjopen-2015-008566
[58] Glyn Elwyn, Paul J Barr, and Sheri Piper. 2018. Digital clinical encounters. *BMJ* 361 (2018). https://doi.org/10.1136/bmj.k2061
[59] Agata Ferretti, Manuel Schneider, and Alessandro Blasimme. 2018. Machine learning in medicine: opening the new data protection black box. *Eur. Data Prot. L. Rev.* 4 (2018), 320. https://doi.org/10.21552/edpl/2018/3/10
[60] Casey Fiesler. 2019. Ethical Considerations for Research Involving (Speculative) Public Data. *Proc. ACM Hum.-Comput. Interact.* 3, GROUP, Article 249 (dec 2019), 13 pages. https://doi.org/10.1145/3370271
[61] National Commission for the Protection of Human Subjects of Biomedical and Md Behavioral Research, Bethesda. 1978. *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Superintendent of Documents.
[62] Laura Forlano and Anijo Mathew. 2017. From design fiction to design friction: Speculative and participatory design of values-embedded urban technology. (2017), 7–24.
[63] Freed. 2023. *getfreed.ai*. Retrieved April 3, 2023 from https://www.getfreed.ai/?utm_campaign=Competitors&gclid=Cj0KCQjw8qmhBhClARIsANAtbodta3SnUxmMZaXvM0AXs0So7Tdmop3hQ-PSLpTgxDLR_StaD2NzF_UaAkZjEALw_wcB
[64] Batya Friedman and David G Hendry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press.
[65] Maya Indira Ganesh and Emanuel Moss. 2022. Resistance and refusal to algorithmic harms: Varieties of 'knowledge projects'. *Media International Australia* 183, 1 (2022), 90–106. https://doi.org/10.1177/1329878x221076288
[66] Patricia Garcia, Tonia Sutherland, Marika Cifor, Anita Say Chan, Lauren Klein, Catherine D'Ignazio, and Niloufar Salehi. 2020. No: Critical Refusal as Feminist Data Practice. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '20 Companion)*. Association for Computing Machinery, New York, NY, USA, 199–202. https://doi.org/10.1145/3406865.3419014
[67] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for Datasets. *Commun. ACM* 64, 12 (nov 2021), 86–92. https://doi.org/10.1145/3458723

[68] Alix Gerber. 2018. Participatory Speculation: Futures of Public Safety. In *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2* (Hasselt and Genk, Belgium) *(PDC '18)*. Association for Computing Machinery, New York, NY, USA, Article 23, 4 pages. https://doi.org/10.1145/3210604.3210640

[69] Sara Gerke, Timo Minssen, and Glenn Cohen. 2020. Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial Intelligence in Healthcare*. Elsevier, 295–336. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

[70] Shilpa Ghatnekar, Adam Faletsky, and Vinod E Nambudiri. 2021. Digital scribe utility and barriers to implementation in clinical practice: a scoping review. *Health and Technology* (2021), 1–7. https://doi.org/10.1007/s12553-021-00568-0

[71] Judy Wawira Gichoya, Liam G McCoy, Leo Anthony Celi, and Marzyeh Ghassemi. 2021. Equity in essence: a call for operationalising fairness in machine learning for healthcare. *BMJ Health & Care Informatics* 28, 1 (2021), e100289. https://doi.org/10.1136/bmjhci-2020-100289

[72] Christina Harrington and Tawanna R Dillahunt. 2021. Eliciting Tech Futures Among Black Young Adults: A Case Study of Remote Speculative Co-Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 397, 15 pages. https://doi.org/10.1145/3411764.3445723

[73] Christina N. Harrington, Shamika Klassen, and Yolanda A. Rankin. 2022. "All That You Touch, You Change": Expanding the Canon of Speculative Design Towards Black Futuring. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 450, 10 pages. https://doi.org/10.1145/3491102.3502118

[74] Ti Hoang, Rohit Ashok Khot, Noel Waite, and Florian 'Floyd' Mueller. 2018. What Can Speculative Design Teach Us about Designing for Healthcare Services?. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (Melbourne, Australia) *(OzCHI '18)*. Association for Computing Machinery, New York, NY, USA, 463–472. https://doi.org/10.1145/3292147.3292160

[75] Matthew K. Hong, Udaya Lakshmi, Thomas A. Olson, and Lauren Wilcox. 2018. Visual ODLs: Co-Designing Patient-Generated Observations of Daily Living to Support Data-Driven Conversations in Pediatric Care. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174050

[76] Organizers Of Queer in AI, Anaelia Ovalle, Arjun Subramonian, Ashwin Singh, Claas Voelcker, Danica J. Sutherland, Davide Locatelli, Eva Breznik, Filip Klubička, Hang Yuan, Hetvi J, Huan Zhang, Jaidev Shriram, Kruno Lehman, Luca Soldaini, Maarten Sap, Marc Peter Deisenroth, Maria Leonor Pacheco, Maria Ryskina, Martin Mundt, Melvin Selim Atay, Milind Agarwal, Nyx McLean, Pan Xu, A Pranav, Raj Korpan, Ruchira Ray, Sarah Mathew, Sarthak Arora, St John, Tanvi Anand, Vishakha Agrawal, William Agnew, Yanan Long, Zijie J. Wang, Zeerak Talat, Avijit Ghosh, Nathaniel Dennler, Michael Noseworthy, Sharvani Jha, Emi Baylor, Aditya Joshi, Natalia Y. Bilenko, Andrew McNamara, Raphael Gontijo-Lopes, Alex Markham, Evyn Dǒng, Jackie Kay, Manu Saraswat, Nikhil Vytla, and Luke Stark. 2023. Queer In AI: A Case Study in Community-Led Participatory AI. arXiv:2303.16972 [cs.CY]

[77] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 659, 14 pages. https://doi.org/10.1145/3411764.3445385

[78] Natalie Joseph-Williams, Glyn Elwyn, and Adrian Edwards. 2014. Knowledge is not power for patients: a systematic review and thematic synthesis of patient-reported barriers and facilitators to shared decision making. *Patient education and counseling* 94, 3 (2014), 291–309.

[79] Near Future Laboratory Julian Bleecker. 2009. *Design Fiction: A Short Essay on Design, Science, Fact and Fiction*. Retrieved April 3, 2023 from https://blog.nearfuturelaboratory.com/2009/03/17/design-fiction-a-short-essay-on-design-science-fact-and-fiction/

[80] Shaun K. Kane, Meredith Ringel Morris, Ann Paradiso, and Jon Campbell. 2017. "At Times Avuncular and Cantankerous, with the Reflexes of a Mongoose": Understanding Self-Expression through Augmentative and Alternative Communication Devices. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1166–1179. https://doi.org/10.1145/2998181.2998284

[81] Shivani Kapania, Alex S Taylor, and Ding Wang. 2023. A hunt for the Snark: Annotator Diversity in Data Practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3544548.3580645

[82] Michael Katell, Meg Young, Dharma Dailey, Bernease Herman, Vivian Guetler, Aaron Tam, Corinne Bintz, Daniella Raz, and PM Krafft. 2020. Toward situated interventions for algorithmic equity: lessons from the field. In *Proceedings*

of the 2020 conference on fairness, accountability, and transparency. 45–55.

[83] David R Kaufman, Barbara Sheehan, Peter Stetson, Ashish R Bhatt, Adele I Field, Chirag Patel, and James Mark Maisel. 2016. Natural Language Processing–Enabled and Conventional Data Capture Methods for Input to Electronic Health Records: A Comparative Usability Study. *JMIR Med Inform* 4, 4 (28 Oct 2016), e35. https://doi.org/10.2196/medinform.5544

[84] Tobie Kerridge. 2015. *Designing Debate: The entanglement of speculative design and upstream engagement.* Ph. D. Dissertation. Goldsmiths, University of London. https://doi.org/10.21606/drs.2016.22

[85] Awais Hameed Khan, Neelam Ejaz, Sarah Matthews, Stephen Snow, and Ben Matthews. 2021. *Speculative Design for Education: Using Participatory Methods to Map Design Challenges and Opportunities in Pakistan.* Association for Computing Machinery, New York, NY, USA, 1748–1764. https://doi.org/10.1145/3461778.3462117

[86] MD Kirby. 1983. Informed consent: what does it mean? *Journal of Medical Ethics* 9, 2 (1983), 69–75. https://doi.org/10.1136/jme.9.2.69.

[87] Keith Kirkpatrick. 2020. Natural Language Misunderstanding. *Commun. ACM* 63, 11 (oct 2020), 17–18. https://doi.org/10.1145/3422600

[88] Jeffrey G Klann and Peter Szolovits. 2009. An intelligent listening framework for capturing encounter notes from a doctor-patient dialog. *BMC medical informatics and decision making* 9, 1 (2009), 1–10. https://doi.org/10.1186/1472-6947-9-S1-S3

[89] KLAS. 2023. *Mayo Clinic 3M: The Path to Automation and Ambient Documentation.* Retrieved April 3, 2023 from https://klasresearch.com/archcollaborative/casestudy/mayo-clinic-and-3m-the-path-to-automation-and-ambient-documentation/468

[90] Jodi Kodish-Wachs, Emin Agassi, Patrick Kenny III, and J Marc Overhage. 2018. A systematic comparison of contemporary automatic speech recognition engines for conversational clinical speech. In *AMIA Annual Symposium Proceedings*, Vol. 2018. American Medical Informatics Association, 683.

[91] Sandjar Kozubaev and Carl DiSalvo. 2020. The Future of Public Libraries as Convivial Spaces: A Design Fiction. In *Companion Proceedings of the 2020 ACM International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '20)*. Association for Computing Machinery, New York, NY, USA, 83–90. https://doi.org/10.1145/3323994.3369901

[92] Sandjar Kozubaev, Chris Elsden, Noura Howell, Marie Louise Juul Søndergaard, Nick Merrill, Britta Schulte, and Richmond Y. Wong. 2020. Expanding Modes of Reflection in Design Futuring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376526

[93] P. M. Krafft, Meg Young, Michael Katell, Jennifer E. Lee, Shankar Narayan, Micah Epstein, Dharma Dailey, Bernease Herman, Aaron Tam, Vivian Guetler, Corinne Bintz, Daniella Raz, Pa Ousman Jobe, Franziska Putz, Brian Robick, and Bissan Barghouti. 2021. An Action-Oriented AI Policy Toolkit for Technology Audits by Community Advocates and Activists. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 772–781. https://doi.org/10.1145/3442188.3445938

[94] Bogdan Kulynych, David Madras, Smitha Milli, Inioluwa Deborah Raji, Angela Zhou, and Richard Zemel. 2020. Participatory Approaches to Machine Learning. International Conference on Machine Learning Workshop.

[95] Sindhu Kutty. 2021. *The Rise Of AI Voice Assistants In Clinical Documentation.* Retrieved April 3, 2023 from https://www.forbes.com/sites/forbesbusinesscouncil/2021/03/03/the-rise-of-ai-voice-assistants-in-clinical-documentation/

[96] Marcus Law, Anila Mathai, Paula Veinot, Fiona Webster, and Maria Mylopoulos. 2015. Exploring lesbian, gay, bisexual, and queer (LGBQ) people's experiences with disclosure of sexual identity to primary care physicians: a qualitative study. *BMC family practice* 16, 1 (2015), 1–8. https://doi.org/10.1186/s12875-015-0389-4

[97] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, and Ariel D. Procaccia. 2019. WeBuildAI: Participatory Framework for Algorithmic Governance. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 181 (nov 2019), 35 pages. https://doi.org/10.1145/3359283

[98] Peter Lee, Carey Goldberg, and Isaac Kohane. 2023. *The AI Revolution in Medicine: GPT-4 and Beyond.* Pearson Education, Limited.

[99] Brenna Li, Noah Crampton, Thomas Yeates, Yu Xia, Xirong Tian, and Khai Truong. 2021. *Automating Clinical Documentation with Digital Scribes: Understanding the Impact on Physicians.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445172

[100] Wanyue Li, Xinyue Zhou, and Qian Yang. 2021. Designing medical artificial intelligence for in- and out-groups. *Computers in Human Behavior* 124 (2021), 106929. https://doi.org/10.1016/j.chb.2021.106929

[101] Steven Y Lin, Tait D Shanafelt, and Steven M Asch. 2018. Reimagining clinical documentation with artificial intelligence. In *Mayo Clinic Proceedings*, Vol. 93. Mayo Foundation for Medical Education and Research, 563–565.

https://doi.org/10.1016/j.mayocp.2018.02.016

[102] Joseph Lindley and Paul Coulton. 2015. Back to the Future: 10 Years of Design Fiction. In *Proceedings of the 2015 British HCI Conference* (Lincoln, Lincolnshire, United Kingdom) *(British HCI '15)*. Association for Computing Machinery, New York, NY, USA, 210–211. https://doi.org/10.1145/2783446.2783592

[103] Zachary C Lipton. 2018. The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16, 3 (2018), 31–57. https://doi.org/doi={10.1145/3236386.3241340}

[104] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[105] Michal Luria and Stuart Candy. 2022. Letters from the Future: Exploring Ethical Dilemmas in the Design of Social Agents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 419, 13 pages. https://doi.org/10.1145/3491102.3517536

[106] Trieuvy Luu, Martijn van den Broeck, and Marie Louise Juul Søndergaard. 2018. Data Economy: Interweaving Storytelling and World Building in Design Fiction. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction* (Oslo, Norway) *(NordiCHI '18)*. Association for Computing Machinery, New York, NY, USA, 771–786. https://doi.org/10.1145/3240167.3240270

[107] Donald Martin Jr, Vinodkumar Prabhakaran, Jill Kuhlberg, Andrew Smart, and William S Isaac. 2020. Participatory problem formulation for fairer machine learning through community-based system dynamics. *arXiv preprint arXiv:2005.07572* (2020).

[108] Maciej A Mazurowski. 2020. Artificial intelligence in radiology: some ethical considerations for radiologists and algorithm developers. *Academic Radiology* 27, 1 (2020), 127–129. https://doi.org/10.1016/j.acra.2019.04.024

[109] Zion Mengesha, Courtney Heldreth, Michal Lahav, Juliana Sublewski, and Elyse Tuennerman. 2021. "I don't Think These Devices are Very Culturally Sensitive."—Impact of Automated Speech Recognition Errors on African Americans. *Frontiers in Artificial Intelligence* 4 (2021). https://doi.org/10.3389/frai.2021.725911

[110] Milagros Miceli, Tianling Yang, Laurens Naudts, Martin Schuessler, Diana Serbanescu, and Alex Hanna. 2021. Documenting Computer Vision Datasets: An Invitation to Reflexive Data Practices. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 161–172. https://doi.org/10.1145/3442188.3445880

[111] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3287560.3287596

[112] Camille Nebeker, John Torous, and Rebecca J Bartlett Ellis. 2019. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC medicine* 17, 1 (2019), 1–7. https://doi.org/10.1186/s12916-019-1377-7

[113] Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Tajudeen Kolawole, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Hassan Muhammad, Salomon Kabongo, Salomey Osei, Sackey Freshia, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa, Mofe Adeyemi, Masabata Mokgesi-Selinga, Lawrence Okegbemi, Laura Jane Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreutzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroro Orife, Ignatius Ezeani, Idris Abdulkabir Dangana, Herman Kamper, Hady Elsahar, Goodness Duru, Ghollah Kioko, Espoir Murhabazi, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Emezue, Bonaventure Dossou, Blessing Sibanda, Blessing Itoro Bassey, Ayodele Olabiyi, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. 2020. Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages. arXiv:2010.02353 [cs.CL]

[114] Helen Nissenbaum. 2004. Privacy as Contextual Integrity. *Wash. L. Rev.* 79 (2004), 119.

[115] Nuance. 2023. *Announcing DAX Express.* Retrieved April 3, 2023 from https://www.nuance.com/healthcare/campaign/learn-more/express.html?cid=7016T000002543rQAA&utm_campaign=NHA-IQ-20230320-Express&utm_medium=Display&utm_source=terminus-web

[116] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 6464 (2019), 447–453. https://doi.org/10.1126/science.aax2342

[117] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. 2019. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data* 2 (2019), 13. https://doi.org/10.3389/fdata.2019.00013

[118] Onora O'Neill. 2003. Some limits of informed consent. *Journal of Medical Ethics* 29, 1 (2003), 4–7.

[119] Mimi Onuoha and Mother Cyborg (Diana Nucera). 2018. *A People's Guide to AI.* Retrieved April 3, 2023 from https://alliedmedia.org/wp-content/uploads/2020/09/peoples-guide-ai.pdf

[120] Anthony J Onwuegbuzie and Kathleen MT Collins. 2007. A typology of mixed methods sampling designs in social science research. *Qualitative Report* 12, 2 (2007), 281–316.

[121] Efthimios Parasidis, Elizabeth Pike, and Deven McGraw. 2019. A Belmont report for health data. *The New England Journal of Medicine* 380, 16 (2019), 1493–1495. https://doi.org/10.1056/NEJMp1816373

[122] Joon Sung Park, Danielle Bragg, Ece Kamar, and Meredith Ringel Morris. 2021. Designing an Online Infrastructure for Collecting AI Data From People With Disabilities. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 52–63. https://doi.org/10.1145/3442188.3445870

[123] Jessica Pater, Casey Fiesler, and Michael Zimmer. 2022. No Humans Here: Ethical Speculation on Public Data, Unintended Consequences, and the Limits of Institutional Review. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP, Article 38 (jan 2022), 13 pages. https://doi.org/10.1145/3492857

[124] Thomas H Payne, Sarah Corley, Theresa A Cullen, Tejal K Gandhi, Linda Harrington, Gilad J Kuperman, John E Mattison, David P McCallie, Clement J McDonald, Paul C Tang, William M Tierney, Charlotte Weaver, Charlene R Weir, and Michael H Zaroukian. 2015. Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. *Journal of the American Medical Informatics Association* 22, 5 (05 2015), 1102–1110. https://doi.org/10.1093/jamia/ocv066

[125] Jennifer Pierre, Roderic Crooks, Morgan Currie, Britt Paris, and Irene Pasquetto. 2021. Getting Ourselves Together: Data-Centered Participatory Design Research & Epistemic Burden. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3411764.3445103

[126] Juan C Quiroz, Liliana Laranjo, Ahmet Baki Kocaballi, Shlomo Berkovsky, Dana Rezazadegan, and Enrico Coiera. 2019. Challenges of developing a digital scribe to reduce clinical documentation burden. *NPJ digital medicine* 2, 1 (2019), 114. https://doi.org/10.1038/s41746-019-0190-1

[127] Inioluwa Deborah Raji and Joy Buolamwini. 2019. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Honolulu, HI, USA) *(AIES '19)*. Association for Computing Machinery, New York, NY, USA, 429–435. https://doi.org/10.1145/3306618.3314244

[128] Anuradha Reddy, A. Baki Kocaballi, Iohanna Nicenboim, Marie Louise Juul Søndergaard, Maria Luce Lupetti, Cayla Key, Chris Speed, Dan Lockton, Elisa Giaccardi, Francisca Grommé, Holly Robbins, Namrata Primlani, Paulina Yurman, Shanti Sumartojo, Thao Phan, Viktor Bedö, and Yolande Strengers. 2021. Making Everyday Things Talk: Speculative Conversations into the Future of Voice Interfaces at Home. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 23, 16 pages. https://doi.org/10.1145/3411763.3450390

[129] Microsoft Research. 2021. *EmpowerMD Demo*. Retrieved April 3, 2023 from https://www.microsoft.com/en-us/research/video/empowermd-medical-conversations-to-medical-intelligence/

[130] Esther Rolf, Theodora T Worledge, Benjamin Recht, and Michael Jordan. 2021. Representation Matters: Assessing the Importance of Subgroup Allocations in Training Data. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 9040–9051. https://proceedings.mlr.press/v139/rolf21a.html

[131] Negar Rostamzadeh, Diana Mincu, Subhrajit Roy, Andrew Smart, Lauren Wilcox, Mahima Pushkarna, Jessica Schrouff, Razvan Amironesei, Nyalleng Moorosi, and Katherine Heller. 2022. Healthsheet: Development of a Transparency Artifact for Health Datasets. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1943–1961. https://doi.org/10.1145/3531146.3533239

[132] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone Wants to Do the Model Work, Not the Data Work": Data Cascades in High-Stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 39, 15 pages. https://doi.org/10.1145/3411764.3445518

[133] Kshitij Saxena, Robert Diamond, Reid F Conant, Terri H Mitchell, Guido Gallopyn, and Kristin E Yakimow. 2018. Provider adoption of speech recognition and its impact on satisfaction, documentation quality, efficiency, and cost in an inpatient EHR. *AMIA Summits on Translational Science Proceedings* 2018 (2018), 186.

[134] Saykara. 2021. *Website*. Retrieved April 3, 2023 from https://hitconsultant.net/2021/02/08/nuances-acquires-saykara/#.YcuUWBPMI4s

[135] Darcell P Scharff, Katherine J Mathews, Pamela Jackson, Jonathan Hoffsuemmer, Emeobong Martin, and Dorothy Edwards. 2010. More than Tuskegee: understanding mistrust about research participation. *Journal of health care for the poor and underserved* 21, 3 (2010), 879. https://doi.org/10.1353%2Fhpu.0.0323

[136] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 317 (oct 2021), 37 pages. https://doi.org/10.1145/3476058

[137] Daniel Schönberger. 2019. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. *International Journal of Law and Information Technology* 27, 2 (2019), 171–203. https://doi.org/10.1093/ijlit/eaz004

[138] Britta F. Schulte. 2016. Using Design Fiction to Reflect on Autonomy in Smart Technology for People Living with Dementia. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 1110–1113. https://doi.org/10.1145/2968219.2972717

[139] Jacob Schuman. 2012. Beyond Nuremberg: A critique of informed consent in third world human subject research. *JL & Health* 25 (2012), 123.

[140] Ashish Sharma, Inna W. Lin, Adam S. Miner, David C. Atkins, and Tim Althoff. 2021. Towards Facilitating Empathic Conversations in Online Mental Health Support: A Reinforcement Learning Approach. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 194–205. https://doi.org/10.1145/3442381.3450097

[141] Renee Shelby, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla, Jess Gallegos, Andrew Smart, Emilio Garcia, and Gurleen Virk. 2023. Identifying Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. arXiv:2210.05791 [cs.HC]

[142] Peter Singleton and Michael Wadsworth. 2006. Consent for the use of personal medical data in research. *BMJ* 333, 7561 (2006), 255–258. https://doi.org/10.1136%2Fbmj.333.7561.255

[143] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2020. Participation is not a Design Fix for Machine Learning. *CoRR* abs/2007.02423 (2020). arXiv:2007.02423 https://arxiv.org/abs/2007.02423

[144] Mona Sloane, Ian René Solano-Kamaiko, Jun Yuan, Aritra Dasgupta, and Julia Stoyanovich. 2023. Introducing contextual transparency for automated decision systems. *Nature Machine Intelligence* (2023), 1–9. https://doi.org/10.1038/s42256-023-00623-7

[145] Marie Louise Juul Søndergaard and Lone Koefoed Hansen. 2018. Intimate Futures: Staying with the Trouble of Digital Personal Assistants through Design Fiction. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 869–880. https://doi.org/10.1145/3196709.3196766

[146] Stat. 2023. *'We're getting much more aggressive': Microsoft's Nuance adds GPT-4 AI to its medical note-taking tool.* Retrieved April 3, 2023 from https://www.statnews.com/2023/03/20/microsoft-nuance-gpt4-dax-chatgpt/

[147] Moira A Stewart. 1995. Effective physician-patient communication and health outcomes: a review. *CMAJ: Canadian Medical Association Journal* 152, 9 (1995), 1423.

[148] Theresa Jean Tanenbaum. 2014. *What is Design Fiction.* Retrieved April 3, 2023 from https://www.quora.com/What-is-design-fiction

[149] Rachael Tatman. 2017. Gender and dialect bias in YouTube's automatic captions. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. 53–59.

[150] Damon Taylor. 2019. Design Futures. *A Companion to Contemporary Design since 1945* (2019), 51–71.

[151] Alexandra To, Hillary Carey, Riya Shrivastava, Jessica Hammer, and Geoff Kaufman. 2022. Interactive Fiction Provotypes for Coping with Interpersonal Racism. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 453, 14 pages. https://doi.org/10.1145/3491102.3502044

[152] Shari Trewin. 2018. AI fairness for people with disabilities: Point of view. *arXiv preprint arXiv:1811.10670* (2018).

[153] Maka Tsulukidze, Stuart W Grande, Rachel Thompson, Kenneth Rudd, and Glyn Elwyn. 2015. Patients covertly recording clinical encounters: threat or opportunity? A qualitative analysis of online texts. *PloS one* 10, 5 (2015), e0125824. https://doi.org/10.1371/journal.pone.0125824

[154] Ding Wang, Shantanu Prabhat, and Nithya Sambasivan. 2022. Whose AI Dream? In Search of the Aspiration in Data Annotation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 582, 16 pages. https://doi.org/10.1145/3491102.3502121

[155] Matt Ward. 2019. *Critical about Critical and Speculative Design.* Retrieved April 3, 2023 from https://speculativeedu.eu/critical-about-critical-and-speculative-design/

[156] Meredith Whittaker, Meryl Alper, Cynthia L Bennett, Sara Hendren, Liz Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Rankin, Emily Rogers, Marcel Salas, et al. 2019. Disability, bias, and AI. *AI Now Institute* (2019). https://ainowinstitute.org/disabilitybiasai-2019.pdf

[157] Wikipedia. 2023. *Large Language Model.* Retrieved June 30, 2023 from https://en.wikipedia.org/wiki/Large_language_model

[158] Lauren Wilcox, Renee Shelby, Rajesh Veeraraghavan, Oliver L. Haimson, Gabriela Cruz Erickson, Michael Turken, and Rebecca Gulotta. 2023. Infrastructuring Care: How Trans and Non-Binary People Meet Health and Well-Being Needs through Technology. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 489, 17 pages. https://doi.org/10.1145/3544548.3581040

[159] Matt Willis and Mohammad Hossein Jarrahi. 2019. Automating Documentation: A Critical Perspective into the Role of Artificial Intelligence in Clinical Documentation. In *Information in Contemporary Society*, Natalie Greene Taylor, Caitlin Christian-Lamb, Michelle H. Martin, and Bonnie Nardi (Eds.). Springer International Publishing, Cham, 200–209. https://doi.org/10.1007/978-3-030-15742-5_19

[160] Richmond Y Wong. 2021. Using Design Fiction Memos to Analyze UX Professionals' Values Work Practices: A Case Study Bridging Ethnographic and Design Futuring Methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18. https://doi.org/10.1145/3411764.3445709

[161] Richmond Y Wong and Vera Khovanskaya. 2018. Speculative design in HCI: from corporate imaginations to critical orientations. In *New Directions in Third Wave Human-Computer Interaction: Volume 2-Methodologies*. Springer, 175–202. https://doi.org/10.1007/978-3-319-73374-6_10

[162] Qian Yang, Yuexing Hao, Kexin Quan, Stephen Yang, Yiran Zhao, Volodymyr Kuleshov, and Fei Wang. 2023. Harnessing biomedical literature to calibrate clinicians' trust in AI decision support systems. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3544548.3581393

Received January 2023; revised April 2023; accepted July 2023