




Stakeholder perspectives on ethical and trustworthy voice AI in health care

Jean-Christophe Bélisle-Pipon¹ , Maria Powell², Renee English¹, Marie-Françoise Malo¹, Vardit Ravitsky^{3,4}, Bridge2AI-Voice Consortium and Yael Bensoussan⁵

Abstract

Objective: Voice as a health biomarker using artificial intelligence (AI) is gaining momentum in research. The noninvasiveness of voice data collection through accessible technology (such as smartphones, telehealth, and ambient recordings) or within clinical contexts means voice AI may help address health disparities and promote the inclusion of marginalized communities. However, the development of AI-ready voice datasets free from bias and discrimination is a complex task. The objective of this study is to better understand the perspectives of engaged and interested stakeholders regarding ethical and trustworthy voice AI, to inform both further ethical inquiry and technology innovation.

Methods: A questionnaire was administered to voice AI experts, clinicians, scholars, patients, trainees, and policy-makers who participated at the 2023 Voice AI Symposium organized by the Bridge2AI-Voice AI Consortium. The survey used a mix of Likert scale, ranking and open-ended questions. A total of 27 stakeholders participated in the study.

Results: The main results of the study are the identification of priorities in terms of ethical issues, an initial definition of ethically sourced data for voice AI, insights into the use of synthetic voice data, and proposals for acting on the trustworthiness of voice AI. The study shows a diversity of perspectives and adds nuance to the planning and development of ethical and trustworthy voice AI.

Conclusions: This study represents the first stakeholder survey related to voice as a biomarker of health published to date. This study sheds light on the critical importance of ethics and trustworthiness in the development of voice AI technologies for health applications.

Keywords

AI in medicine, biomarkers, data collection, ethical considerations, health data, health disparities, inclusivity, telehealth, trustworthiness, voice AI

Submission date: 19 August 2023; Acceptance date: 21 May 2024

Introduction

Voice as a health biomarker is gaining momentum in both academic and industry-led research, driven by increased availability of voice samples, high-quality microphones accessible through smartphones and tablets, novel voice analysis technologies, and a growing demand for telehealth services.^{1–8} The generation of the human voice is a complex process involving the coordination of various biological systems including respiration, phonation, resonance articulation and

¹Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

²Vanderbilt University Medical Center, Department of Otolaryngology-Head & Neck Surgery, Nashville, TN, Canada

³Hastings Center, Garrison, NY, USA

⁴Department of Global Health and Social Medicine, Harvard University, Cambridge, MA, USA

⁵Department of Otolaryngology-Head & Neck Surgery, University of South Florida, Tampa, FL, USA

Corresponding author:

Jean-Christophe Bélisle-Pipon, Faculty of Health Sciences, Simon Fraser University, BLU 10006, 8888 University Dr., Burnaby, BC, Canada.
Email: jean-christophe_belisle-pipon@sfu.ca



prosody. Any disruptions in these systems can lead to acoustic biomarkers linked to various diseases, from Parkinson's disease to dementia, mood disorders, and some cancers. AI-powered acoustic analysis technology is shedding new light on these diagnostic potentials, positioning voice as a promising diagnostic tool in digital health.^{9–14}

Beyond the direct health outcomes and benefits, the concept of voice holds metaphorical relevance for inclusion and representation. Literature on equity, diversity, and inclusion often uses “voice” to signify self-advocacy and agency.^{15–17} This symbolism underscores the crucial need to address representation and bias concerns in data collection and utilization. Voice data collection, being non-invasive and feasible even in resource-limited settings through common technology like mobile phones, could potentially address some health disparities. Given its low collection cost and noninvasive nature, compared to sources like blood or DNA,^{18–21} the voice may foster wider participation from marginalized communities who may have historical mistrust toward the medical establishment. Therefore, voice not only stands as a potentially powerful health biomarker from a clinical perspective but also has the potential to serve as an inclusive and equalizing tool in healthcare.

However, the development of new AI-ready datasets and the aspiration to create them free from bias and potential discrimination, and to ensure fair benefits distribution, is a complex task. Health-related data collection and utilization do not operate in a vacuum but mirror social dynamics, institutional structures, and sadly, also existing biases and health disparities.^{22–24} Moreover, health data are subject to specific regulations and standards designed to manage their use.^{25–27} Thus, before conducting large-scale voice data acquisition, analysis, and integrating them into clinical applications, there is a need to reflect on technical and practical aspects, as well as voice AI's ethical, legal, and social implications (ELSI). This comprehensive examination is crucial to ascertain their therapeutic relevance and equitable impact on clinical decision-making processes, and ultimately, on the healthcare system at large.^{28–30}

Although some authors and organizations are focusing their efforts on developing best practices for voice in the setting of conversational AI platforms,³¹ voice linked to other health data poses far greater challenges than the use of voice alone. For voice to emerge as a biomarker of health, voice data collection must be linked to extensive health information, including specific diagnoses and health confounders, and therefore requires a more in-depth analysis of ethical implications linked to patient protection and trust.³² Meanwhile, much work is focusing on the governance of AI in health and medicine,^{33–35} notably by focusing on the exceptional dimensions of AI,^{36,37} democratic participation in AI oversight,³⁸ and the development of practical guidelines. However, this has not yet been inquired, developed, adapted, or made

relevant to voice AI. Referring to the term “voice AI” in the context of this article implies voice AI for health purposes and always involves health information.

Although the need is poignant, no studies have queried stakeholders' perspectives on ethical and trustworthy voice AI for health research. To address this situation, we conducted a study to explore the perspectives of voice AI experts, clinicians, scholars, patients, trainees, and relevant policymakers. A questionnaire allowed us to investigate four main themes: (1) perspectives on voice AI's usefulness and impacts, (2) ethically sourced data for voice AI, (3) diversity and inclusivity in voice data and voice AI, (3) synthetic voice data, and (4) trustworthiness of voice AI technologies. The goal of this work is to better understand the perspectives of engaged and interested stakeholders regarding ethical and trustworthy voice AI to inform both further ethical inquiry and technology innovation.

Methods

Overview

This study was performed as the third part of a three-survey series conducted before, during, and after the 2023 Voice AI Symposium held in Washington DC on April 19th, 2023.³⁹ This symposium was organized by the Bridge2AI-Voice AI Consortium,⁴⁰ which is a funded National Institutes of Health (NIH) entity with the mission of building an ethically sourced, large-scale, hypothesis-agnostic, human voice database linked to health information, to help diagnose diseases. This inaugural Voice AI Symposium hosted a variety of stakeholders invested in the voice as a health biomarker through interactive sessions centered around four aspects of voice AI: purpose, evidence, ethics, and trust. All attendees could be characterized as highly committed to, invested in, interested in, and/or knowledgeable about voice AI, from a technical, medical, ethical, legal, experiential, or commercial point of view.

Recruitment and sampling

Approximately six weeks after the Symposium, we distributed a web-based questionnaire. This was developed by a literature review that explored the ethical, legal, and social implications of voice AI, voice data, and voice as a biomarker (article forthcoming). This postsymposium questionnaire was distributed to the 108 attendees between 31st May and 19th June 2023. Participants were recruited through postsymposium thank-you mailings, a private LinkedIn group for all Voice AI Symposium participants, and a private WhatsApp group for Voice AI Symposium panelists and organizers. The questionnaire was administered through SurveyMonkey, which is hosted on Simon Fraser University's (SFU) servers. A link to the questionnaire was sent directly to the participants.

Survey design

The questionnaire was designed to collect the perspectives of stakeholders on each of the four main themes. The questionnaire was mostly composed of qualitative questions (short-answer and open-ended questions) as well as some quantitative questions (100-point Likert scale and ranking questions). Since comprehensive demographic data was collected during a presymposium survey, for this postsymposium survey, respondents were simply asked to identify as one of three personas depending on how they interacted with voice AI and voice data: generator, consumer, or ancillary support. The descriptions provided for each persona were the following: (1) *Generator*: As a generator, I (and/or those I advocate for) are most likely to collect/generate/acquire data from patients/participants in research (e.g. patients, patient advocates, underrepresented populations advocates, diversity, and inclusion advocates, etc.). (2) *Consumer*: As a consumer, I am most likely to analyze patient/participant data and make AI applications or clinical interventions based on findings. (3) *Ancillary*: As an ancillary, I am less likely to utilize patient/participant data directly; however, I support the collection and dissemination of data.

Data analysis

We used thematic analysis to identify topics and arguments common to the respondents and more marginal ones that call for a broader discussion of our research object. For quantitative responses, we used Excel (Microsoft) for basic statistical analysis (means, medians, and standard deviation). We abstained from employing statistical hypothesis testing, thereby rendering the quantitative results incapable of generalization owing to the modest cohort size.

Ethics approval

The protocol and the questionnaire were approved by the SFU Research Ethics Board (#30001567). The study was delegated and considered to represent a minimal risk for the participants.

Results

Overview

A total of 27 participants completed the questionnaire, while five others provided partial answers (response rate: 30%). Only fully answered questionnaires were kept for analysis. Figures 1 and 2 present a portrait of the participants from the 2023 Voice AI Symposium.

Perspectives on voice AI's usefulness and impacts

Stakeholders hold varying perspectives on the impact of voice AI technologies on precision public health (see

Figure 3). While some perceive a significant impact, others are more cautious, resulting in a lack of consensus. The wide range of responses, from 20 to 100 on the 100-point Likert scale, indicates a diversity of viewpoints. The mean score of 83.22 suggests an overall positive perception, while the standard deviation of 19.96 demonstrates the variability in opinions. These diverse perspectives underscore the need for inclusive dialogue and collaboration among stakeholders to navigate the potential impact of voice AI technologies on precision public health effectively.

Then, stakeholders were asked to rank the perceived usefulness of voice data collection methods for voice AI development (see Table 1). Overall, the method considered to be the most useful is a clinical setting, through clinical protocols for data collection. The more common methods of communication, over the phone or through consumer communication technologies, are widely considered to be less useful. Interestingly, consumers and generators found ambient recording in a person's environment/home quite useful (Rank 3). Ancillaries considered it to be the least useful. Wearable technologies/devices and dedicated mobile applications seemed relatively useful (somewhat in the middle of the pack), except for consumers who indicated that this was the most useful data collection method.

Stakeholders were then asked to priority rank common issues related to AI design initiatives, but specifically in the context of voice AI (see Table 2). Stakeholders' reported roles seemed to be at play as there were marked differences in priorities concerning certain issues. There were only two priority rankings that were common to all respondents. Privacy protection consistently appeared in the top three issues to tackle for all roles, indicating its significance in maintaining trust and ethical standards in Voice AI systems. Conversely, explainability consistently appeared as one of the lowest priorities, which may seem surprising, especially given the importance of current debates on the black box problem.

Both Generators and Consumers had the same top three, therefore they also agreed on prioritizing "accuracy and reliability" and "fairness and nondiscrimination." Accuracy and reliability denote the importance for voice AI developers or data collectors as well as end-users of having accurate and reliable tools and data, while fairness and nondiscrimination highlight the importance of addressing biases and ensuring equity considerations in voice-based interventions. Only ancillaries prioritized causal insights (AI being based on causal mechanisms, rather than correlations) and informed consent, suggesting a focus on understanding the underlying mechanisms and implications of voice AI systems and better capacitating users in decision-making processes. Overall, the trends indicate a collective recognition of the significance of privacy, fairness, and accuracy in voice AI, with some variations in emphasis based on the roles' specific perspectives and responsibilities.

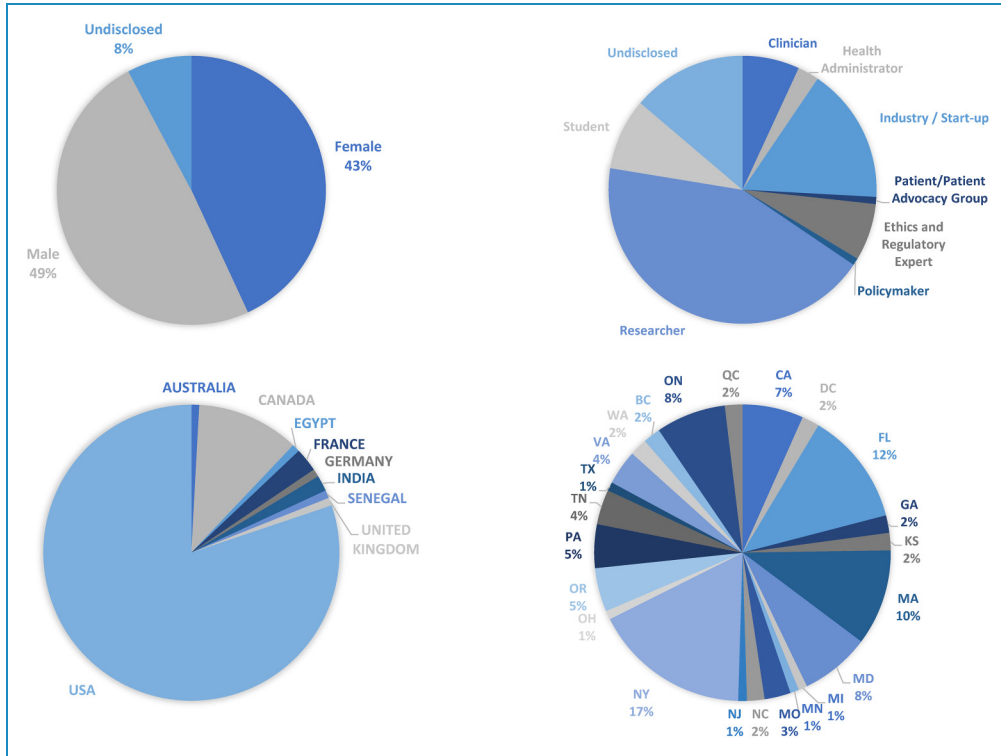


Figure 1. Sociodemographics of the Voice AI Symposium participants. (a) Gender distribution. (b) Stakeholder Group. (c) Country. (d) U.S. state or Canadian provinces (BC, ON, QC).

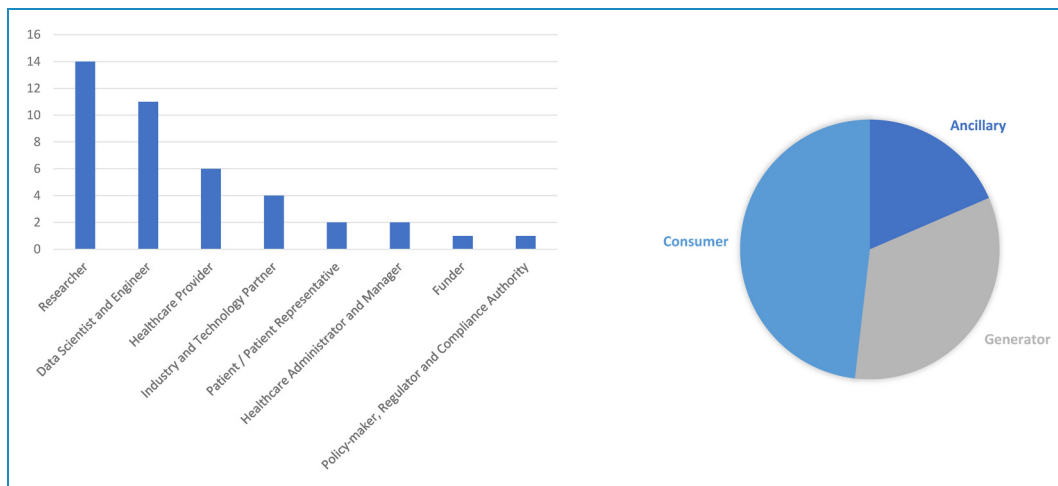


Figure 2. Survey's respondents' roles and relationship to voice AI. (a) Respondents' roles (multiple choice, so people may have more than one role). (b) Relationship to Voice AI (limited to 1 per respondent).

The stakeholders' perspectives on the implementation of voice AI in clinical settings and its impact on health inequities vary widely (see Figure 3). The mean score of 55.52 suggests a mixed perception, while the standard deviation of 31.26 indicates a significant degree of variability in opinions. The median score of 50 reflects a relatively balanced distribution of responses. While some stakeholders believe

that the implementation of voice AI can reduce health inequities, others express concerns about exacerbating them. The wide range of scores, from 1 to 100, highlights the complexity and nuance surrounding this issue. These divergent perspectives underscore the need for careful consideration, ethical frameworks, and inclusive discussions to ensure that the implementation of voice AI in clinical

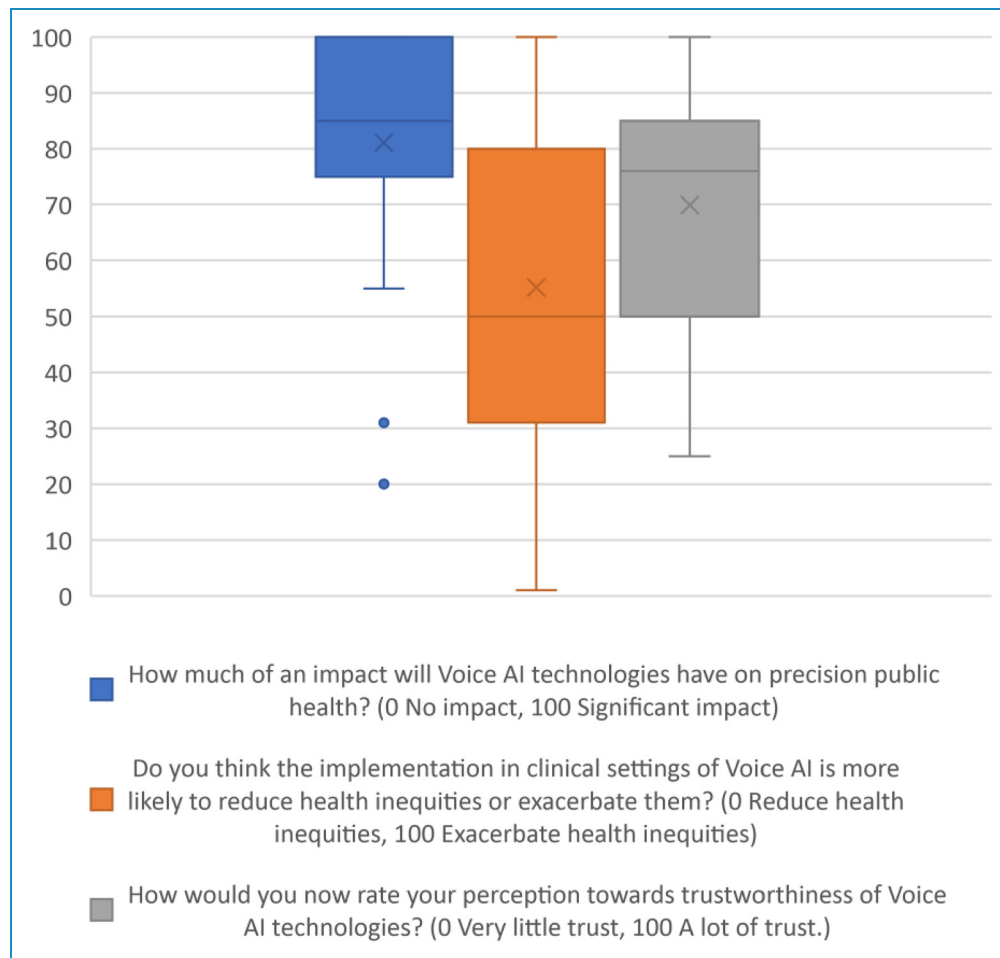


Figure 3. Stakeholders' perceptions on Voice AI's usefulness and impacts, its impact on health inequities, and its trustworthiness. *Note:* X: the cross is the mean of the dataset; \square : the box represents the interquartile range (IQR), indicating the range between the first quartile (Q1) and the third quartile (Q3). It captures the central 50% of the data; $-$: the line within the box represents the median, which indicates the middle value of the dataset; \pm : the lines extending from the box, referred to as whiskers, represent the minimum and maximum values within a specified range. Any data point beyond the whiskers is considered an outlier; \bullet : the dots are individual data points falling outside the whiskers are depicted as individual data markers, representing potential anomalies or extreme values.

settings prioritizes equity and avoids unintended negative consequences.

Ethically sourced data for voice AI

We also asked in an open question what the notion of “ethically sourced data” for voice AI development implied for participants. Ethically sourced data is a concept fundamental to the Bridge2AI program (which funds the voice AI consortium), but it has not been defined by the NIH, nor does it have a commonly accepted definition in the literature. Table 3 presents the key aspects raised by respondents and a description using their language.

According to our participants, ethically sourced voice data refers to the collection and use of data in a manner that upholds key ethical principles (consent, transparency, fairness, and diversity), respects the rights and privacy of

individuals, and follows regulatory requirements and industry best practices. It involves obtaining informed consent and being transparent about the data collection process, including who is collecting the data, how it will be used, and for what duration. Further, ethically sourced data minimizes biases, treats data subjects fairly, and promotes diverse representation. It ensures that data is stored securely, protecting individuals' privacy and personal information. Ethically sourced data empowers individuals by giving them control over their data and educates them about the responsible use of data. It also considers the environmental impact of data storage and management, promoting energy efficiency and sustainability.

Ultimately, ethically sourced data collection focuses on the value, purpose, and potential benefits of the data while prioritizing the well-being and rights of the individuals involved. The definition highlights the fundamental

Table 1. Ranking the usefulness of Various voice data collection methods.

Voice data collection methods\role	Generator	Consumer	Ancillary
Ambient recordings in clinic settings	2	6	4
Ambient recording in a person's environment/home	3	3	7
Dedicated mobile applications	4	4	2
In a clinical setting, through clinical protocols for data collection	1	2	1
Over phone (cellphones and landlines)	6	5	6
Through communication technologies available to consumers (Zoom, Teams, Whatsapp, etc.)	7	7	5
Wearable technologies/devices	5	1	3

Note: The lower the number, the higher the usefulness ranking.

Table 2. Priority ranking of voice AI issues to address.

Role	Top 3 issues	Middle 3 issues	Lowest 3 issues
Generator	<ul style="list-style-type: none"> • Accuracy and reliability • Fairness and nondiscrimination • Privacy protection 	<ul style="list-style-type: none"> • Informed consent • Regulatory/legal compliance • Diverse data set 	<ul style="list-style-type: none"> • Patient control over their data • Explainability • Causal insights
Consumer	<ul style="list-style-type: none"> • Fairness and nondiscrimination • Privacy protection • Accuracy and reliability 	<ul style="list-style-type: none"> • Patient control over their data • Informed consent • Regulatory/legal compliance 	<ul style="list-style-type: none"> • Causal insights • Explainability • Diverse data set
Ancillary	<ul style="list-style-type: none"> • Causal insights • Informed consent • Privacy protection 	<ul style="list-style-type: none"> • Diverse data set • Accuracy and reliability • Fairness and nondiscrimination 	<ul style="list-style-type: none"> • Regulatory/legal compliance • Patient control over their data • Explainability

considerations in ensuring ethical practices throughout the data lifecycle, or as one of the respondents puts it through the “data chain, from data collection to AI usage.”

Ensuring diversity and inclusivity in voice data and voice AI

In response to the question of how to ensure diverse and inclusive voice data collection (see Table 4), most respondents emphasized the importance of targeted recruitment and oversampling from diverse groups to ensure proper representation. They proposed approaches such as matching the current population (i.e. ensuring that the sample distribution is equivalent to the actual population), including marginalized and orphan patient groups, and understanding vocal parameters within different populations. Additionally, suggestions were made to actively recruit participants, utilize mobile applications for random sampling, and establish easy access data collection points like

clinics and handheld devices. These strategies aim to expand the sources of voice data and capture a wide range of diverse voices, thus promoting inclusivity.

Respondents also highlighted the need for transparency, accountability, and community engagement in achieving diverse and inclusive voice data. They suggested developing tools to assess the inclusiveness of speech databases and involving diverse partners in data collection efforts. Ongoing reporting of data set composition and continuous feedback collection from communities were seen as essential for tailoring data collection approaches to reach a broader audience. Guidelines for data collectors, including bias training and cultural sensitivity, were proposed to ensure respectful engagement with participants. These recommendations emphasize the importance of actively involving diverse populations and creating an inclusive environment throughout the data collection process.

To properly assess the diversity and inclusivity of voice data, the survey respondents proposed various metrics. Demographic analysis of data sets and language diversity

Table 3. Key aspects of “ethically sourced data.”

Key aspects	Description	Percent of respondents raising the aspect
Consent and transparency	Obtaining informed consent from individuals and being transparent about the data collection process, including usage and duration.	74%
Privacy and security	Securely collecting and storing data, protecting the privacy and personal information of individuals.	63
Fairness and avoiding bias	Treating data subjects fairly, avoiding biases, and ensuring diverse representation in data collection.	52
Compliance and standards	Following regulatory requirements and adhering to industry best practices for data protection and privacy.	37
Empowerment and control	Giving individuals control over their data and educating them about responsible use.	26
Environmental responsibility	Minimizing the environmental impact of data storage and management, promoting energy efficiency and sustainability.	22
Value and purpose	Collecting data with a clear purpose and considering the value and potential benefits of the data for ethical AI development and research.	19

were suggested as ways to evaluate representation. The idea of tracking the inclusion of marginalized and underserved groups, along with different languages, dialects, accents, genders, and age groups, was prominent among the suggestions. Participants also emphasized the need for ongoing evaluation, user empowerment, and involvement of communities affected by medical conditions in shaping the metrics. By prioritizing diversity and inclusivity in voice data collection and continuously refining the evaluation criteria, stakeholders can work collectively to ensure that voice data reflects the rich tapestry of voices and perspectives in a diverse society.

Perceptions on synthetic voice data

The respondents were asked to provide their insights on the usefulness and ethicality surrounding synthetic voice data for voice AI development (see Table 5). Regarding the usefulness of synthetic voice data, opinions were divided. Some respondents recognized the potential of synthetic data, particularly in terms of generating voice data that may not exist (and be difficult to obtain or access) in real-world recordings. They indicated that it also offers scalability and versatility in creating accessible voice samples. Proponents highlighted benefits such as privacy protection (as no patient data are needed or may be exposed). Others emphasized the value of synthetic data to improve treatment outcomes by facilitating the generation of behavioral voice therapy targets and the development of physiologically appropriate speech-generative augmentative and alternative communication devices.

However, others expressed skepticism, raising concerns about biases in synthetic datasets and the risk of shifting from people-centric to data-centric approaches. Questions were raised about the credibility of generated evidence and potential biases introduced by researchers. Biases in synthetic voice data were identified as a significant ethical challenge, as they may perpetuate existing biases or introduce new ones. Privacy and data protection emerged as another critical issue, with the risk of deepfake technology and potential harm to individuals and their families. Transparency and user knowledge were highlighted as essential for building trust, while verification of assumptions and pathologies in data generation were seen as important for ethical use. Accountability, regulations, and guidelines for researchers were proposed as measures to address the ethical issues associated with synthetic voice data.

To address these identified challenges, the respondents provided several solutions. Informed consent and privacy protection were deemed crucial for ethical data collection. Guidelines for researchers and data collectors were suggested to ensure respectful engagement, bias training, and cultural sensitivity. Transparency in the data generation process, disclosure to users, and verification of assumptions were proposed to build trust and address biases. The importance of diversity in synthetic datasets to mitigate biases was emphasized, along with the need for proper definitions, regulations, and accountability measures. Respondents highlighted the significance of regulatory guidance and responsible practices to ensure the ethical use of synthetic voice data and prevent misuse or deceptive practices.

Table 4. Suggestions, ideas, and metrics for ensuring diversity and inclusivity.

Suggestions for ensuring diversity and inclusivity	Best practices for ensuring diversity and inclusivity	Metrics for ensuring diversity and inclusivity
<ul style="list-style-type: none"> • Precise targeting and oversampling from diverse groups. • Matching the current population and overrepresenting certain communities or orphan patients. • Understanding vocal parameters and characteristics within each population. • Active, not passive, recruitment. • Effective mobile applications and protocols for random sampling. • Using easy access data collection points such as clinics, handheld devices, and nearables. • Conducting data collection at different sites with diverse subjects. • Deliberate efforts to pursue data collection from diverse populations. • Making data collection accessible through cell phones or other accessible methods. • Providing guidelines to data collectors for respectful engagement and bias training. • Broad outreach and continuous feedback collection from diverse communities. 	<ul style="list-style-type: none"> • Developing automatic tools to assess the inclusiveness of a speech database. • Collaborating with diverse partners and expanding data collection globally. • Making data collection protocols publicly available for improvement. • Ongoing reporting of the composition of data sets. • Involving a variety of healthcare centers and locations. • Including different languages, dialects, and accents from multiple jurisdictions. • Ensuring representation of all genders, age groups, and disease cohorts. 	<ul style="list-style-type: none"> • Sociodemographic analysis of the datasets. • Language diversity (across languages, accents, and dialects) • Regular evaluation and user empowerment. • Tracking in the datasets the representation of marginalized and underserved groups.

Fostering or strengthening trustworthy voice AI

Respondents were asked to provide approaches and resources for fostering or strengthening trustworthy voice AI; they emphasized the importance of transparency as a key factor in building trust. First, they suggested that open and inclusive discussions should be prioritized before developing general voice AI, allowing people to express their concerns and contribute to the decision-making process. By demonstrating trust and transparency in the implementation of voice AI projects, users can feel more confident in the technology and its applications. Second, education and awareness were also highlighted as crucial resources for fostering trust in voice AI technologies. Respondents suggested the need to share information with the general public, across all age groups, to dispel myths, provide an understanding of the basics, and address the risks and limitations associated with voice AI. They emphasized the importance of educating users about the rationale behind the technology, its future uses, and ways to overcome its potential risks. By empowering individuals with knowledge, it becomes easier to build trust and alleviate concerns related to voice AI.

In terms of resources (see Table 6 for their description, benefits and limitations), respondents mentioned the value of symposiums and events as platforms for addressing concerns and engaging in open discussions. These gatherings

provide opportunities to foster dialog and gather input from various stakeholders. Additionally, they emphasized the significance of guidelines and standards for voice AI development. By establishing common frameworks, best practices, and ethical guidelines, it becomes possible to ensure consistency and promote the responsible and trustworthy development of voice AI systems. They also acknowledged the importance of leveraging resources from reputable organizations such as the Institute of Electrical and Electronics Engineers, which offers ethical guidelines and insights to inform the development process.

Overall, respondents highlighted the significance of transparency, education, open dialogue, and adherence to ethical guidelines as key approaches and resources for fostering and strengthening trustworthy voice AI. By embracing these principles, developers can build confidence among users and the public, while ensuring responsible and ethical practices in the field of voice AI technology.

Trustworthiness of voice AI technologies

When asked about the three words that they associate with trustworthy voice AI (see Figure 4), most responses conveyed positive sentiments, emphasizing attributes such as “responsible,” “secure,” “reliable,” “ethical,” “helpful,” “inclusive,” and “transparent” that carry positive

Table 5. Usefulness, risks, and ethical issues surrounding synthetic voice data.

Usefulness of synthetic voice data	Risks of voice synthetic data	Ethical issues with voice synthetic data	Solutions to Ethical Issues with Voice Synthetic Data
<ul style="list-style-type: none"> • Scalability and versatility in generating voice samples • Ability to generate voice data that may not be available in real recordings • Can assist in baseline data comparison • Enables precise measurement of particular parameters • Promising approach, especially with diverse underlying data • Can be used for training new AI models • Can be helpful for generative AAC devices to match user’s particularities and needs • Correct articulatory errors in the patient’s own voice for treatment outcomes 	<ul style="list-style-type: none"> • Shifting from people-centric to data-centric approaches, potentially overlooking qualitative assessments and focusing solely on quantitative metrics. • Biases introduced by dataset builders, including potential biases of researchers and AI algorithms. • Inherent biases built into synthetic voice data, which may misrepresent desired representations and hinder development outcomes. • Potential misuse of synthetic voice data and life-threatening implications of deepfake technology, where synthetic voices can be used to create deceptive and harmful content. • Lack of transparency and accountability in the generation and use of synthetic voice data. • Skepticism and trust issues among clinicians and patients regarding the credibility and reliability of synthetic voice data. • Challenges in capturing the full variability of the human voice using synthetic data. • Limited diversity and representativeness of synthetic voice datasets, leading to potential biases and inadequate generalization. 	<ul style="list-style-type: none"> • Potential biases in synthetic datasets • Risk of shifting from people-centric to data-centric approaches • Privacy and data protection concerns • Deceptive practices and the risk of deepfakes • Lack of transparency and user knowledge • Verification of assumptions and effects of pathologies in data generation • Trustworthiness and credibility of synthetic voice data • Potential harm caused by misuse or misrepresentation • Accountability for the data and responsible practices • Regulatory guidance and proper definitions to prevent misuse and deception 	<ul style="list-style-type: none"> • Education on the benefits, risks, and limitations with synthetics data. • Guidelines for researchers and data collectors on respectful engagement and bias training • Transparency in the data generation process and disclosure to users • Verification of assumptions and effects of pathologies in synthetic data generation • Purposeful diversity in synthetic datasets to mitigate biases • Proper definitions and regulations to prevent misuse and deceptive practices • Accountability for data and responsible practices • Regulatory guidance to ensure ethical use of synthetic voice data • Building trust through transparency and disclosure • Cultural sensitivity in synthetic data generation and implementation

AAC: augmentative and alternative communication; AI: artificial intelligence.

connotations, indicating a favorable perception of trustworthy voice AI. Some responses also include words like “hopeful,” “exciting,” and “innovative” which further suggest a positive sentiment. There are a few indications of uncertainty or caution in some responses using words like “probabilistic,” “ambiguous,” “immature,” “developing-field,” “much-to-learn,” “limits-of-validity,” and “not-quite-there-yet,” suggesting a level of hesitation or reservation regarding the trustworthiness of voice AI in its current state. These responses indicate a more neutral or slightly negative sentiment compared to the positive associations observed in other responses. The limited presence of clearly negative sentiment suggests that the overall perceptions of respondents toward trustworthy voice AI are predominantly positive or neutral, with concerns or uncertainties being less common.

A final question sought to ponder perceptions on trustworthiness towards voice AI technologies (see Figure 3),

as measured on a scale of 1–100, shows a mean score of 70.15, which can be perceived as quite high. The standard deviation of 26.51 suggests a moderate degree of variability in perceptions, while the median score of 75 reflects a relatively balanced distribution. This indicates that participants’ trust in voice AI technologies varies, ranging from low levels to high levels. While some respondents expressed significantly high levels of trust (e.g. scores close to 100), others reported levels of distrust (e.g. scores below 50). These variations highlight the complexity of trust perceptions and the need for continued efforts to enhance the trustworthiness of voice AI technologies through ethical practices, transparency, and accountability.

Discussion

This study represents the first study exploring stakeholders’ perspectives on voice as a biomarker of health published to

biases, and potential misuse of voice data. Participants emphasized the significance of obtaining informed consent, ensuring individuals have a comprehensive understanding of how their voice data will be used and empowering them to make informed decisions. Privacy emerged as a critical concern, with participants highlighting the need for robust measures to protect personal information and ensure secure storage and handling of voice data. Moreover, the potential biases inherent in voice data collection and analysis were underscored, emphasizing the necessity for transparency and fairness in the development of AI systems reliant on voice data.

Currently, in the literature, there is no definition of what “ethically sourced data” means for health data and medical AI development, including for voice data and voice AI. While the NIH’s Bridge2AI program is structured on the importance of responsible ethically sourced data acquisition, the program does not provide a definition for it, although it is at the core and underlying Bridge2AI’s three pillars: data, people and ethics.⁴¹ A notable finding pertains to a preliminary stakeholder definition of ethically sourced data. While the recognition of ethical considerations was unanimous, the lack of a cohesive framework for ethically sourced voice data indicates that this still has different meanings for respondents. This highlights the urgent need for comprehensive ethical guidelines and frameworks tailored specifically to voice data collection and AI development. These guidelines should address critical issues such as informed consent procedures, privacy protection, robust data anonymization techniques, and strategies for mitigating biases. Establishing a shared understanding and agreement on the ethical requirements for voice data collection will be pivotal in fostering trust and ensuring responsible utilization of voice data in AI applications.

The perceptions of respondents regarding synthetic voice data shed light on both its potential benefits and ethical concerns in the context of voice AI development. Participants acknowledged the usefulness of synthetic voice data for various applications, including protecting privacy and greatly increasing data volume and diversity. However, concerns regarding biases in synthetic datasets and the potential deviation of research and medical development toward addressing synthetic datasets instead of the real-world needs of patient populations were expressed. These concerns emphasize the importance of rigorous verification and validation procedures to ensure the accuracy, fairness, and inclusivity of synthetic voice data. Participants also raised ethical issues pertaining to informed consent, privacy, and the risk of deepfake voice manipulation. These considerations underscore the imperative of responsible practices and transparency when employing synthetic voice data in AI systems.^{42,43} Interestingly, however, given that the lowest priority aspect to address is causal insights (Table 1), synthetic data, if properly generated, sufficiently validated, and appropriately used, could be of great interest.

Ideas and approaches for fostering diversity and inclusion in voice AI development were prominent in the participants’ responses. They emphasized the necessity of collecting data from diverse populations, including marginalized and underserved groups, to ensure representative datasets and mitigate biases. Strategies such as oversampling from diverse groups and conducting data collection at various sites with diverse subjects were proposed as means to achieve diversity. Furthermore, participants stressed the importance of engaging diverse linguistic experts and ethnic communities to better understand their linguistic nuances, cultural contexts, and voice variations. The development of inclusive algorithms and open access to data were highlighted as additional strategies. Metrics such as demographic analysis and language diversity were recommended for assessing the diversity and inclusivity of voice AI systems. These suggestions align with the principles of fairness, inclusivity, and bias avoidance, thereby promoting the development of more robust and equitable voice AI technologies.

Analysis of the sentiment expressed by respondents toward trustworthy voice AI revealed a mix of positive, negative, and uncertain sentiments. While some participants conveyed optimism and enthusiasm regarding the potential of voice AI, others raised concerns pertaining to privacy, biases, and the need for further development and testing. This variation in sentiment underscores the significance of addressing ethical considerations, promoting transparency, and establishing user trust in the development of trustworthy voice AI systems. It also underscores the necessity for ongoing dialog and collaboration among researchers, developers, users, and regulatory bodies to address these concerns and ensure the responsible and reliable development of voice AI technologies.

Furthermore, participants shared their ideas for fostering or strengthening trustworthy voice AI. Recommendations included proactive regulation, peer-reviewed research, continuous updates informed by field input, transparency in data, models, and evaluation processes, and collaboration with diverse stakeholders. Participants emphasized the importance of adhering to ethical guidelines, detecting, and addressing biases, protecting user privacy, and continuously improving and learning about responsible AI practices. The availability of resources such as ethics guidelines from reputable organizations was highlighted as valuable support for developers and researchers in navigating the ethical dimensions of voice AI development. However, to date, no relevant ethical guidelines exist for voice AI.

In summary, the analysis of participant responses provides valuable insights into the ethical considerations, perceptions of voice synthetic data, approaches for diversity and inclusion, sentiment toward trustworthy voice AI, and ideas for fostering its development. These findings underscore the urgency of establishing comprehensive ethics

frameworks, guidelines, and regulatory measures specific to voice data collection and voice AI development. They highlight the critical importance of transparency, informed consent, privacy protection, fairness, inclusivity, and user trust in the responsible development of voice AI technologies. By addressing these considerations and incorporating diverse perspectives, we can foster the development of trustworthy voice AI that upholds ethical principles, enhances inclusivity, and ensures the positive impact of AI on society.

Going forward

In moving forward, the considerable variation in perceptions noted in our survey points to the importance of early stakeholder involvement in the process. Engaging stakeholders at the outset can encourage future utilization and acceptance of voice AI technologies. However, for stakeholders to form informed opinions and decisions, they must first have an understanding of the potential risks and benefits associated with sharing their voice data. Currently, this essential information is not widely known, and voices might be undervalued compared to other types of data, such as genetic information. As such, it becomes critical to listen to as well as to educate the various stakeholders, primarily clinicians, patients, and patient advocates, who will be the ultimate users of these technologies. By fostering knowledge and understanding, we can build trust in voice AI technologies and facilitate their responsible and ethical use.

Our findings also underscore the necessity to prioritize ethics in all aspects of data collection and AI-based work. Ethical considerations can provide a framework and language for people's concerns about voice data collection and can guide practices across the entire technology continuum—from data collection, through technology development, to eventual use. Moreover, we must delve into the nuances of ethical concerns related to proprietary versus open-access use of collected data. In particular, stakeholders' concerns for transparency related to the downstream use of voice data for AI applications have significant implications for large, hypothesis-agnostic, open-access data repositories, such as the Bridge2AI initiative. Moreover, there is an urgent need for a legal framework that companies can adhere to, promoting regulated practices and reducing the reliance on case-by-case decision-making. This would guide companies in their development and application of voice AI technologies and foster more uniform, ethical practices.

Finally, while our survey participants were diverse, the sample size was small. Hence, to ensure inclusivity and mitigate biases, we recommend seeking additional feedback, especially from those from underrepresented populations. Their insights will be invaluable in addressing concerns related to trust and fairness in voice AI, ensuring

a more robust and equitable development of these technologies. Therefore, the key steps moving forward are broadening stakeholder involvement, prioritizing education about voice AI technologies, developing regulatory frameworks, emphasizing ethical considerations, and engaging diverse perspectives. By addressing these areas, we can contribute to the development of trustworthy voice AI technologies that uphold ethical principles, enhance inclusivity, and have a positive impact on society. In light of the evolving nature of this research domain, it is imperative to conduct further exploration employing larger and more diverse populations, within a global context, to advance our understanding of ethical and trustworthy voice AI.

Limitations

The main limitation of this study is the small sample size. Only 27 participants completed the questionnaire out of 108 attendees at the 2023 Voice AI Symposium. The sample is composed of people already engaged, knowledgeable, involved, and/or interested in voice AI. However, despite this limitation, the participants had diverse profiles, and the study included both quantitative and qualitative questions covering a wide range of themes, which allowed for a thorough analysis of a topic that has not been explored to date from an empirical perspective in terms of experts' and stakeholders' viewpoints. Nevertheless, these preliminary findings serve as valuable contributions to the initial discourse and contemplation surrounding the ethical and trustworthy considerations pertaining to voice AI.

One limitation on the breadth of the results is that the voice AI community is emerging; the Voice AI Symposium represents one of the first gatherings of this embryonic community (including researchers, developers, clinicians, regulators, industry, patients, and others). Hopefully, more events and studies like this will allow the involvement of more people interested in voice AI, and help to bring the field to maturity and diversify voices and perspectives.

We also hope that over the following years, partnerships encouraged by future work from the Bridge2AI-Voice Consortium and the broader voice AI community will also translate into increased participation in discussions on ethical and trustworthy considerations pertaining to voice AI. This will allow for a deeper understanding and a representation of how ethical considerations regarding voice AI evolve in the coming years. Nevertheless, the groundwork presented in this research, done with individuals with significant interest and engagement in the nascent field of voice AI, was needed to gather a deeper understanding of key ethical and trust aspects. We do, however, recognize the need for broader representation. Subsequent Voice AI Symposia will aim to increase participation and diversity in order to ensure a comprehensive

perspective on ELSI of voice AI technologies. This will include people from underrepresented communities, patients with a diversity of conditions and diseases, as well as a diversity of clinicians in terms of professions, backgrounds, and areas of practice. This will, we hope, increase the size and diversity of the participants in future studies done during the Voice AI Symposium.

Additionally, future work from the Bridge2AI-Voice Consortium will aim to explore ethical considerations in voice AI across diverse cultural and geographical contexts. This will involve comparative studies to understand the unique ethical concerns and values of different regions, ensuring that the development of ethical frameworks for voice AI is globally informed and culturally sensitive. Collaboration with stakeholder groups including ethicists, legal experts, and technologists will be prioritized to adopt a multidisciplinary approach to voice AI research. This collaboration aims to integrate diverse expertise, addressing complex ethical dilemmas comprehensively and developing recommendations that are technologically feasible, ethically sound, and legally compliant and can help address the concerns about the lack of inclusion in the ethical governance of AI technologies.⁴⁴ We hope that this future work, will help mend a bridge between research and community, and that this bridge will help ensure that future iterations of the Voice AI Symposium will become an influential space for discussion in the Voice AI field.

Lastly, the practical application of ethical guidelines in voice AI will be a focus of future studies. This involves evaluating the implementation and impact of these guidelines through real-world applications, to identify challenges and areas for improvement. Such an approach will bridge the gap between theoretical ethical frameworks and their practical application, promoting responsible development and integration of voice AI technologies.

Conclusions

This study sheds light on the critical importance of ethics and trustworthiness in the development of voice AI technologies for health applications. The diversity of stakeholder perspectives underscores the necessity of early involvement and education to ensure responsible and inclusive utilization of voice data. Addressing ELSI is paramount to overcoming biases and disparities in voice data collection and utilization. Engaging stakeholders from diverse backgrounds and incorporating their perspectives will lead to more equitable and responsible voice AI development, whether using human or synthetic voices.

By fostering a shared understanding and agreement on ethical requirements, trust in voice AI and its potential as a health biomarker, trustworthy voice AI technologies have the capacity to transform healthcare, promoting inclusivity, equity, and positive impacts on society. As we move forward, ethical considerations must remain at the forefront

to ensure the responsible and beneficial integration of voice AI in healthcare settings.

Acknowledgments: Although the authors were responsible for the current study and article, our work stands on the shoulders of the giant that is Bridge2AI-Voice Consortium (<https://www.b2ai-voice.org/>). The authors would like to thank Kathleen Curp, MyVan John, Cindy Kostelnik, and Desiree McCutcheon for their combined efforts in coordinating and supporting the Voice AI Symposium 2023.

Authors' contributions: JCBP, MP, MFM, and VR conceptualized the research and created the protocol. JCBP, MP, RE, and MFM developed the data collection tools and contributed to participant recruitment and data collection. JCBP analyzed the data and wrote the first draft. All the authors participated in writing and editing the manuscript. All authors approved the content of the manuscript.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: The study was approved by the SFU Research Ethics Board (No. 30001567). The study was delegated and considered of representing a minimal risk for the participants.

Funding: The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: JCBP and VR are funded by the National Institutes of Health (NIH) under the program Bridge to Artificial Intelligence (Bridge2AI) awards: 1OT2OD032742-01 and 1OT2OD032720-01. YB and MP are funded by NIH Bridge2AI award 1OT2OD032720-01.

Guarantor: Jean-Christophe Bélisle-Pipon.

ORCID ID: Jean-Christophe Bélisle-Pipon  <https://orcid.org/0000-0002-8965-8153>

References

1. Fagherazzi G, Fischer A, Ismael M, et al. Voice for health: the use of vocal biomarkers from research to clinical practice. *Digit Biomark* 2021; 5: 78–88.
2. Zhang L, Duvvuri R, Chandra KKL, et al. Automated voice biomarkers for depression symptoms using an online cross-sectional data collection initiative. *Depress Anxiety* 2020; 37: 657–669.
3. Tracy JM, Özkanca Y, Atkins DC, et al. Investigating voice as a biomarker: deep phenotyping methods for early detection of Parkinson's disease. *J Biomed Inform* 2020; 104: 103362.
4. Sara JD S, Orbelo D, Maor E, et al. Guess what we can hear – novel voice biomarkers for the remote detection of disease. *Mayo Clin Proc* 2023; 98: 1353–1375. <https://www.sciencedirect.com/science/article/pii/S0025619623001301>. [cited 2023 Aug 3].

5. Shin D, Cho WI, Park CHK, et al. Detection of minor and major depression through voice as a biomarker using machine learning. *J Clin Med* 2021; 10: 3046.
6. Sara JDS, Maor E, Borlaug B, et al. Non-invasive vocal biomarker is associated with pulmonary hypertension. *PLoS One* 2020; 15: e0231441.
7. Lin H, Karjadi C, Ang TFA, et al. Identification of digital voice biomarkers for cognitive health. *Explor Med* 2020; 1: 406–417.
8. Iyer R, Nedeljkovic M and Meyer D. Using voice biomarkers to classify suicide risk in adult telehealth callers: retrospective observational study. *JMIR Ment Health* 2022; 9: e39807.
9. Suppa A, Costantini G, Ascì F, et al. Voice in Parkinson's disease: a machine learning study. *Front Neurol* 2022; 13: 831428.
10. Reid J, Parmar P, Lund T, et al. Development of a machine-learning based voice disorder screening tool. *Am J Otolaryngol* 2022; 43: 103327.
11. Bensoussan Y, Vanstrum EB, Johns MM, et al. Artificial intelligence and laryngeal cancer: from screening to prognosis: a state of the art review. *Otolaryngol Head Neck Surg* 2022; 168: 1945998221110839.
12. Low DM, Bentley KH and Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. *Laryngoscope Invest Otolaryngol* 2020; 5: 96–116.
13. Fraser KC, Meltzer JA and Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis* 2015; 49: 407–422.
14. Bensoussan Y, Elemento O and Rameau A. Voice as an AI biomarker of health—introducing audiomics. *JAMA Otolaryngol Head Neck Surg* 2024; 150: 283. [cited 2024 Apr 8]; Available from:..
15. Sylvester C. Voice, silence, agency, confusion. In: *Rethinking silence, voice and agency in contested gendered terrains*. Routledge, Abingdon, Oxon, 2018, p. 186.
16. Gammage S, Kabeer N and van der Meulen Rodgers Y. Voice and agency: where are we now? *Fem Econ* 2016; 22: 1–29.
17. Rodríguez LF and Brown TM. From voice to agency: guiding principles for participatory action research with youth. *New Dir Youth Dev* 2009; 2009: 19–34.
18. Jones NL. Making culture a verb: implications for health equity. *Am J Bioeth* 2021; 21: 11–13.
19. Kingsley J, Berkman ER and Derrington SF. The disruptive power of intersectionality. *Am J Bioeth* 2021; 21: 28–30.
20. Lapite FC, Morain SR and Fletcher FE. Grounding medical education in health equity: the time is now. *Am J Bioeth* 2021; 21: 23–25.
21. Ray R and Davis G. Cultural competence as new racism: working as intended? *Am J Bioeth* 2021; 21: 20–22.
22. Nuffield Council on Bioethics. *Artificial intelligence (AI) in health-care and research*. London, 2018, (Bioethics Briefing Note).
23. Klugman CM. Black boxes and bias in AI challenge autonomy. *Am J Bioeth* 2021; 21: 33–35.
24. Stahl BC and Wright D. Ethics and privacy in AI and big data: implementing responsible research and innovation. *IEEE Secur Priv* 2018; 16: 26–33.
25. D'Agostino M and Durante M. Introduction: the governance of algorithms. *Philos Technol* 2018; 31: 499–505.
26. Morley J, Machado CCV, Burr C, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med* 2020; 260: 113172.
27. Ahmad Z, Rahim S, Zubair M, et al. Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. *Diagn Pathol* 2021; 16: 24.
28. Maher NA, Senders JT, Hulsbergen AFC, et al. Passive data collection and use in healthcare: a systematic review of ethical issues. *Int J Med Inf* 2019; 129: 242–247.
29. Shaw J, Rudzicz F, Jamieson T, et al. Artificial intelligence and the implementation challenge. *J Med Internet Res* 2019; 21: e13659.
30. Alami H, Lehoux P, Denis JL, et al. Organizational readiness for artificial intelligence in health care: insights for decision-making and practice. *J Health Organ Manag* 2020; 35: 106–114.
31. Open Voice Network. *TrustMark initiative*. Open Voice Network, 2023 [cited 2023 Jun 26]. <https://openvoicenet.org/trustmark-initiative/>.
32. Gehrmann J, Herczog E, Decker S, et al. What prevents us from reusing medical real-world data in research. *Sci Data* 2023; 10: 459.
33. WHO. *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva, Switzerland, 2021.
34. van de Sande D, Van Genderen ME, Smit JM, et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. *BMJ Health Care Inform* 2022; 29: e100495.
35. Reddy S, Allan S, Coghlan S, et al. A governance model for the application of AI in health care. *J Am Med Inform Assoc* 2020; 27: 491–497.
36. Bouhouita-Guermech S, Gogognon P and Bélisle-Pipon J-C. Specific challenges posed by artificial intelligence in research ethics. *Front Artif Intell* 2023; 6: 1149082. doi:10.3389/frai.2023.1149082.
37. Bélisle-Pipon J-C, Couture V, Roy M-C, et al. What makes artificial intelligence exceptional in health technology assessment? *Front Artif Intell* 2021; 4: 736697.
38. Couture V, Roy M-C, Dez E, et al. Ethical implications of artificial intelligence in population health and the public's role in its governance: perspectives from a citizen and expert panel. *J Med Internet Res* 2023; 25: e44357.
39. Bridge2AI-Voice Consortium. *Voice AI Symposium [Internet]*. 2023, [cited 2023 Aug 3]. https://www.eventsquid.com/event.cfm?event_id=19526.
40. Bridge2AI-Voice Consortium. *Bridge2AI - Voice [Internet]*. 2023. <https://www.b2ai-voice.org/>.
41. Bridge2AI Consortium. *BRIDGE2AI [Internet]*. 2023. <https://bridge2ai.org/>.
42. Victor G, Bélisle-Pipon J-C and Ravitsky V. Generative AI, specific moral values: a closer look at ChatGPT's new ethical implications for medical AI. *Am J Bioeth* 2023; 23: 65–68.
43. Bélisle-Pipon J-C and Ravitsky V, Bridge2AI-Voice Consortium, and Bensoussan Y. Individuals and (synthetic) data points: using value-sensitive design to foster ethical deliberations on epistemic transitions. *Am J Bioeth* 2023; 23: 69–72.
44. Bélisle-Pipon JC, Monteferrante E, Roy MC, et al. Artificial intelligence ethics has a black box problem. *AI & Soc [Internet]* 2022; 38: 1507–1522. doi:10.1007/s00146-021-01380-0. [cited 2022 Mar 6].