“It’s a conversation, not a quiz”: A Risk Taxonomy and Reflection Tool for LLM Adoption in Public Health

JIAWEI ZHOU, Georgia Institute of Technology, USA
AMY Z. CHEN, Georgia Institute of Technology, USA
DARSHI SHAH, Georgia Institute of Technology, USA
LAURA SCHWAB REESE, Purdue University, USA
MUNMUN DE CHOUDHURY, Georgia Institute of Technology, USA

Recent breakthroughs in large language models (LLMs) have generated both interest and concern about their potential adoption as accessible information sources or communication tools across different domains. In public health—where stakes are high and impacts extend across populations—adopting LLMs poses unique challenges that require thorough evaluation. However, structured approaches for assessing potential risks in public health remain under-explored. To address this gap, we conducted focus groups with health professionals and health issue experiencers to unpack their concerns, situated across three distinct and critical public health issues that demand high-quality information: vaccines, opioid use disorder, and intimate partner violence. We synthesize participants’ perspectives into a risk taxonomy, distinguishing and contextualizing the potential harms LLMs may introduce when positioned alongside traditional health communication. This taxonomy highlights four dimensions of risk in individual behaviors, human-centered care, information ecosystem, and technology accountability. For each dimension, we discuss specific risks and example reflection questions to help practitioners adopt a risk-reflexive approach. This work offers a shared vocabulary and reflection tool for experts in both computing and public health to collaboratively anticipate, evaluate, and mitigate risks in deciding when to employ LLM capabilities (or not) and how to mitigate harm when they are used.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **Natural language interfaces**; • **Applied computing → Consumer health**.

Additional Key Words and Phrases: large language models, AI, public health, risk taxonomy, information seeking and support

ACM Reference Format:
Jiawei Zhou, Amy Z. Chen, Darshi Shah, Laura Schwab Reese, and Munmun De Choudhury. 2024. “It’s a conversation, not a quiz”: A Risk Taxonomy and Reflection Tool for LLM Adoption in Public Health. 1, 1 (November 2024), 24 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

---

Authors’ addresses: Jiawei Zhou, Georgia Institute of Technology, Atlanta, GA, USA, j.zhou@gatech.edu; Amy Z. Chen, Georgia Institute of Technology, Atlanta, GA, USA, amychen@gatech.edu; Darshi Shah, Georgia Institute of Technology, Atlanta, GA, USA, dshah435@gatech.edu; Laura Schwab Reese, Purdue University, West Lafayette, IN, USA, lschwabr@purdue.edu; Munmun De Choudhury, Georgia Institute of Technology, Atlanta, GA, USA, munmund@gatech.edu.

Manuscript submitted to ACM

1 Introduction

> *"I love new technologies but this one has me scared. I myself wouldn't feel like I know enough to ask the right questions and to evaluate the impact of it."* — P5 (public health director)

Recent breakthroughs in large language models (LLMs) have spurred widespread attention and rapid adoption across different fields. With their abilities to generate convincing human-like language on the basis of large sets of human-written content [12, 70], LLMs hold the potential to influence how we interact with information. LLMs have quickly gathered hundreds of millions of active users [55] and made their way into everyday products, sometimes even without users' full awareness. Accordingly, public discourse has shown excitement about LLM assistance in various information-seeking and support tasks and even hype about the framing of LLMs as a "next-generation search engine." [1]

This technology enthusiasm extends into the health sector, where studies have presented evidence of LLMs' competence in completing tasks such as clinical documentation [65, 79], decision support [65, 68], therapeutic conversations [35, 61, 71], and public health intervention [30–32]. As a matter of fact, LLMs have already been incorporated into large-scale health systems. For instance, Epic has partnered with Microsoft to start integrating LLM tools into its electronic health record (EHR) software, which owns the largest market share of acute care hospitals in the U.S., to automatically draft clinical notes [38] and patient message responses [39]. While this proliferation seems promising, experts and scholars caution against exaggerating LLM capabilities [18, 77]. Specifically, uncertainty about information quality is problematic, particularly regarding inaccurate information, biases, and harmful content [23, 41, 82]. Given the high-stakes nature and data sensitivity of population-level health, it warrants careful deliberations of when it is appropriate to employ LLM capacities and how we can mitigate potential harm when we do.

However, in practice, evaluating the risks of potential LLM adoption in public health remains challenging. The first gap lies in the lack of domain understanding, as the majority of risk taxonomies lack granularity to be applied for specific uses [42] or are created within the computer science community and tend to leave out domain experts and real users [6, 63, 72]. Second, when considering population-level impact, potential risks come in ecological "layers" from individuals to society, and conventional categorizations based on content types (e.g., misinformation, hate speech, and biases) are not sufficient to evaluate real-world cases [57] while different types often intertwine with each other [24, 34]. Identifying the presence of generated harmful content is only the first step; we must also contextualize low-quality information within specific health issues and relevant populations in order to evaluate the consequences and severity and mitigate potential harm. In response to these gaps, this paper is situated in three distinct and critical public health issues and grounded in the perspectives of both health professionals and members of the general public who might use such tools to seek health information. Specifically, we ask: what negative consequences might arise in adopting LLMs in public health for informational needs and support?

We conducted focus group sessions with ten health professionals and ten health issue experiencers to uncover potential negative influences of using LLMs as a communication tool or information source in public health. We selected vaccines, opioid use disorder (OUD), and intimate partner violence (IPV) as topics for different sessions based on their significance across different dimensions of public health—infectious disease prevention, chronic and well-being care, and community health and safety—all demanding high-quality information with existing prevalent issues such as misinformation, biases, and sensitivity. The result is a risk taxonomy of potential adoption of LLM for public health. Our taxonomy consists of four dimensions of harm: individual behaviors, human-centered care, information ecosystem, and

[1] https://ico.org.uk/about-the-ico/research-reports-impact-and-evaluation/research-and-reports/technology-and-innovation/tech-horizons-report/next-generation-search/

technology accountability (Fig 2). Within each area, we list specific risks and associated example reflection questions to help assist practitioners in both computing and health fields in becoming reflexive and risk-aware.

This risk taxonomy makes several contributions. First, it gives a comprehensive and grounded list of possible risks in implementing LLMs for public health, and differs from pre-existing generic taxonomies by being grounded in public health issues and learning from domain experts and real users. To our knowledge, this is the first work to comprehensively explore potential risks of LLMs for public health. Second, by offering a shared vocabulary, it paves the path for future collaborations between experts in computing and health fields. As public interest starts to gather in LLMs as an emerging technology, it allows for careful and thorough reflections on potential negative consequences and prevents reckless adoption that could result in unbearable disruptions and real harms to individuals and communities and further erode trust in public health responses.

Content Warning: We caution the readers that this paper discusses sensitive topics, including intimate partner violence and opioid use disorder. Some readers may find certain quotes and descriptions to be emotionally triggering.

2 Related Work

2.1 LLMs for Health

A growing body of work has explored the potential of utilizing large language models (LLMs) in health. Studies have presented evidence of their potential in a variety of health-related tasks such as clinical documentation [65, 79], question answering and medical reasoning [3, 48, 79], decision support [65, 68], therapeutic conversations [35, 61, 71], and public health intervention [30–32]. In practice, LLMs have already been implemented in existing systems or new technologies, including EHR systems [39], virtual agents for dealing with domestic violence [45], and public health intervention [30].

Orthogonally, researchers have also raised concerns about the risks of adopting LLMs for health needs. General issues include data privacy [81], language and geographic disparities [29, 47], and the generation of inaccurate, biased, or toxic content [14, 41, 82]. When applied to health specifically, questions remain regarding LLMs' ability to meet clinical standards [25] and their tendency to propagate race-based medicine [50]. Harrer's study found that only half of LLM-generated messages meeting the clinical standards to be directly adopted for use. When used for communication assistance, LLMs are found to have difficulty in being customized or controlled [73] and lack of sensitivity in managing emotional needs and emergencies [71]. Personalizing language to suit diverse demographics is another challenge [30], particularly for minoritized and marginalized groups where cultural nuances and sensitive topics may be overlooked (such as the LGBTQIA+ [43, 51] and disability [22] community). This shortfall extends to well-being care, as studies have shown that LLMs often generate responses that are more insensitive compared to human therapists, especially for individuals experiencing high-intensity emotions [61]. In higher-risk cases, such as patients with borderline personality disorder or schizophrenia, LLMs may even amplify harmful behaviors or symptoms [35].

In reflecting on these issues, scholarship has been to critically examine how we approach LLM evaluation in terms of reporting transparency [18], assessment reliability [18, 77], practical deployment [77], and consideration of relationships in healthcare [75]. For instance, Drogt et al. found in their survey study that around half of clinical language models were not validated on clinical text, and most evaluations focused on traditional natural language processing (NLP) tasks that do not offer meaningful insights for healthcare applications. As a result, they called for frameworks that align more closely with values and factors crucial to health contexts. We build on prior scholarship by comprehensively examining risks posed by LLMs in public health through three critical yet distinctive health issues.

2.2 AI and LLMs Risk Taxonomies

In response to the literature on the ethics and limitations of LLMs, research community has proposed various taxonomies to address risks posed by these models. Some studies take a broader approach, examining AI or algorithmic systems by synthesizing privacy risks [40], sociotechnical harms [63], risks in algorithmic agency [13], and impacts in individual, social, and biospheric dimensions [17]. Others specifically focus on language models. For example, Bender et al. listed the risks of LM and suggested future mitigation efforts, including pre-development evaluations on how these models support stakeholder values. Weidinger et al. summarized six ethical and social risks, such as misinformation harms and human-computer interaction harms, while Kumar et al. looks into practical methods in mitigating harms. They summarized common intervention strategies across four stages of NLP model development: data, model training and design, inference and generation, and application. In the realm of health, De Choudhury et al. discussed the benefits and harms of LLMs in digital mental health across four application areas: supporting individual care-seeking, assisting caregiving, decision-making aid, and transforming telehealth paradigms. Antoniak et al. proposed principles for ethical use of NLP and LLMs for maternal healthcare in recognizing context, measurements, and values.

Despite these contributions, there is no established vocabulary to specifically frame the risks in public health contexts, a domain that is featured by its high-stakes nature and population-level impacts. This absence of a shared framework limits our ability to systematically explore, communicate, and mitigate potential harms in this specific domain.

From a methodological point, most existing taxonomies are position papers or review studies centered within the computer science field, which often exclude domain experts and end users in applied areas. Research has warned of the limitations of evaluating LLMs for clinical purposes using traditional NLP tasks and metrics [77]. Consequently, research in HCI and CSCW has called for human-centered approaches that closely involve stakeholders, as certain AI-related harms may be difficult to capture solely through existing AI literature [2, 33, 75]. Our work seeks to address this gap by proposing a shared vocabulary grounded in real public health issues, informed by perspectives from both health professionals and real users.

3 Method

3.1 Health Topic Selection and Context

This study aims to evaluate the risks of using large language models (LLMs) for informational needs in public health by focusing on three distinct and critical issues: vaccines, opioid use disorder (OUD), and intimate partner violence (IPV). The selection of these topics was driven by their significance across different dimensions of public health—infectious disease prevention, chronic and wellbeing care, and community health and safety—and through consultation with the public health expert coauthor. Each topic underscores the importance of high-quality information in public health communication. Specifically, vaccines play a crucial role in infectious disease management, while misinformation and pubic distrust significantly contribute to vaccine hesitancy [16, 53, 54]. Effective counter-speech is necessary to combat these misconceptions and enhance trust [80, 83]. OUD is a major issue within the realm of chronic and wellbeing care, which is exacerbated by stigma [19, 49] and vulnerability [69, 78]. Therefore, this health crisis is in need of public education to reduce stigma and create a supportive environment. IPV is a high-risk and stigmatized issue in community health and safety that demands highly sensitive and supportive communication [46, 59]. IPV involves complex challenges, including personal safety, mental health, and legal and financial challenges [21, 58]. Providing high-quality, empathetic information is essential for supporting survivors and addressing their comprehensive needs.

Table 1. Details about the participant makeup. Professionals (P1-10) are health professionals who have created or shared information about specific topics, and their years of experience in occupations are indicated in brackets. Experiencers (E1-10) are individuals who have experienced questions about specific topics and actively sought online information. Experiencers in OUD and IPV sessions have lived experiences with these issues.

ID	Occupation(yrs)	Topic	Age	Gender	Ethnicity	Education	Experience with AI tools	Use of LLM tools
P1	Social worker(18)	IPV	35-44	Female	White	Advanced degree	Only read about it	Not heard about it
P2	Researcher(14)	IPV	45-54	Female	White	Advanced degree	Worked on related topics	Sometimes, but not regularly
P3	Social worker(13)	IPV	35-44	Female	White	Advanced degree	Only read about it	Never used, but heard about it
P4	Researcher(4)	IPV	18-24	Female	Asian	Bachelor's degree	Occasionally used AI tools	Sometimes, but not regularly
E1	Employed	IPV	35-44	Female	White	Advanced degree	Occasionally used AI tools	Rarely, use only occasionally
E2	Leave of Absence	IPV	55-64	Female	White	Associate degree	Extensively used AI tools	Always, use whenever applicable
E3	Unemployed	IPV	35-44	Female	Black/African Ame	Some college, no degree	Occasionally used AI tools	Sometimes, but not regularly
P5	Public health director(30)	OUD	65+	Female	White	Advanced degree	Worked on related topics	Sometimes, but not regularly
P6	Nurse(5)	OUD	35-44	Non-binary	White	Advanced degree	Occasionally used AI tools	Often, as part of my routine
P7	Nurse(5)	OUD	45-54	Female	Hispanic/Latina	Advanced degree	In-depth understanding	Often, as part of my routine
E4	Self-employed	OUD	35-44	Female	White	Bachelor's degree	Occasionally used AI tools	Sometimes, but not regularly
E5	Student	OUD	25-34	Female	White	Some college, no degree	In-depth understanding	Sometimes, but not regularly
E6	Employed	OUD	35-44	Female	White	Bachelor's degree	Extensively used AI tools	Always, use whenever applicable
E7	Employed	OUD	35-44	Male	White	Some college, no degree	Extensively used AI tools	Always, use whenever applicable
P8	Public health staff(35)	Vac	55-64	Female	White	Advanced degree	Only read about it	Never used, but heard about it
P9	Coalition staff(9)	Vac	18-24	Male	White	Bachelor's degree	In-depth understanding	Always, use whenever applicable
P10	Nurse(18)	Vac	45-54	Female	Other	Advanced degree	Extensively used AI tools	Often, as part of my routine
E8	Student	Vac	18-24	Female	Black/African American	Associate degree	Occasionally used AI tools	Always, use whenever applicable
E9	Employed	Vac	25-34	Female	Black/African American	Some college, no degree	Occasionally used AI tools	Rarely, use only occasionally
E10	Employed	Vac	25-34	Female	Black/African American	Advanced degree	Occasionally used AI tools	Rarely, use only occasionally

3.2 Participants and Recruitment

We recruited both health professionals and members of the general public. The health professionals group (labeled with 'professionals/P' in the paper) consisted of people working in the health sector, such as social workers, nurses, community health workers, and researchers, all of whom have experience working on one of the three topics. The general public group (labeled with 'experiencers/E' in the paper) includes people who have actively sought information about one of these three issues. For the sessions on OUD and IPV, they are individuals who have lived experience with these issues. We required that all participants be over 18 years old, be located in the U.S., and have a desktop/laptop or tablet to join an online meeting and participate in brainstorming writing. Interested respondents completed a brief survey about their demographics, occupation, professional background (for the professional cohort), AI attitudes, familiarity with LLM products, and their availability to suggested timings. The complete survey is included in supplements and we invited participants solely based on availability.

To recruit health professionals, we employed a snowball sampling approach by reaching out to professional networks and relevant organizations via email. Our primary sources were public health departments and professional organizations. Specifically, we identified programs and divisions within federal and state public health departments, such as immunization divisions, overdose prevention initiatives, and violence prevention initiatives. We contacted departments across various states, including Georgia, Indiana, California, Minnesota, and Texas, in order to get a broad geographic coverage. During this process, we identified partner alliances and support systems listed by state departments and subsequently reached out to these organizations and shelters. Simultaneously, we distributed our recruitment flyer via professional networks and working groups, contacted ethics teams and task forces in hospitals and universities, and contacted individual researchers who work on health communication and support on the chosen health issues.

To recruit the general public, we used the Prolific [2] research platform where we launched three separate studies for our health topics. For the OUD and IPV sessions, we utilized Prolific's screeners by limiting the pool to people who experienced substance abuse problems or abuse-related incidents and specified eligibility to those who have

[2] https://www.prolific.com/

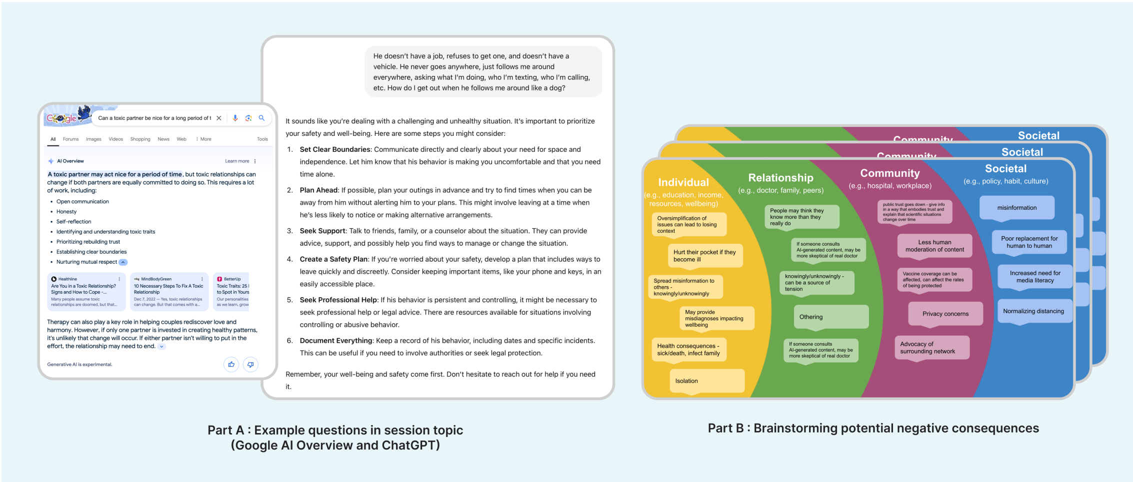


Fig. 1. Activities included in the focus group sessions: (A) Demonstrating LLMs' interaction mode and response quality through example questions on the session topic, (B) Brainstorming potential negative consequences of using LLM for public health.

personally experienced OUD or IPV. All survey responses were rewarded with $1 based on an hourly rate of $12 for a 5-minute survey. To provide a safe environment, we structured the sessions to include only individuals who share similar experiences and encouraged participants to use any preferred name during the session. We did not collect any personally identifiable information and all communications were carried on the Prolific platform.

In total, ten health professionals (E1-10) and ten public participants (P1-10) participated in focus group sessions. E5 participated in an interview format after two other participants dropped out last minute due to medical emergencies. We compensated $75 and $30 for professionals and general public participants respectively for their time. On average, our participants have a moderate level of previous experience in AI tools and usage of LLM tools (2.7 and 3.15 on a [0,5] scale), with health professionals having slightly more experience in AI but less use of LLM. Health professionals have a median of 13.5 years of experience in their job. Table 1 provides a summary of participant information.

3.3 Study Procedure

Due to the intentionally wide geographic distribution of our participants, all focus group sessions were conducted virtually on Zoom and lasted 90 minutes. Before the actual sessions, we conducted two rounds of pilot tests with four domain experts and four experiencers to refine the study design. To provide a safe and comfortable environment for our participants, we held separate sessions for professionals and experiencers to ensure experiencers participated in sessions with individuals who shared similar lived experiences.

Each session began with an introduction to the study's goals and process, a reminder of participant rights, and self-introductions where participants shared their relevant experience with the session's topic. To contextualize the discussion of the potential role of LLMs in meeting informational needs, we started by discussing how participants currently share or seek health information, how they evaluate the quality of content, the precautions required in sharing information (for information creators), common challenges in information searching (for information seekers), and prevalent misconceptions about the specific topics. Following the introduction, we gave a brief introduction to LLM mechanisms, underlining the fact that LLMs generate responses based on probabilities and are designed to produce human-like text but not necessarily accurate information. Considering the varying prior experience of LLM use, we demonstrated their interaction mode and response quality by testing two example questions relevant to the session topic (Fig. 1 Part A), which were paraphrased from online communities of r/abusiverelationships, r/OpiatesRecovery, and r/HPV [3] on Reddit. To ensure consistency across sessions, we recorded the screen while logged out and played recordings during the session. The two example questions were separately tested on ChatGPT and Google AI Overview to show different interaction styles. We emphasized that the goal was not to compare these tools directly, as the questions were inherently different.

After establishing a basic understanding of LLMs and their capabilities, we introduced the final activity where participants brainstormed potential negative consequences of using LLMs for health informational needs (Fig. 1 Part B). Before the brainstorming, we introduced the socio-ecological model [11], a framework commonly used in public health [15, 20, 52, 64], to assist participants think about impacts at different levels: individual (e.g., education, resources), relationship (e.g., physician, family), community (e.g., hospital, workplace), and societal (e.g., policy, habits). Participants first wrote down potential consequences individually (if the session had only three participants) or in breakout groups (if the session had four participants), and then shared their thoughts with the larger group.

3.4 Data Analysis

We identified the types of risks raised by participants by analyzing session transcripts and brainstorming notes through a grounded approach using reflexive thematic analysis [8–10]. Three researchers independently performed open coding on three sessions in a line-by-line fashion to identify mentioned risks, and each identified lower-level codes. These initial codes were formed verbatim from participants' conversations or brainstorming entries (e.g., *"more reliance on tech vs. human relationships"*) or *in vivo* coded from the discussion (e.g., *"don't understand prompt creation enough to use it effectively"*). They then met to discuss their interpretations and consolidate the codes to reach a consensus. The first author proceeded to code the remaining three sessions and discussed any new codes with the research team. The researchers then independently grouped the codes into higher-level themes and discussed them to reach a revised set of themes reflecting joint views. In distilling themes, we initially used a deductive approach based on the socio-ecological model's themes [11], but ultimately adopted an inductive approach to account for the overlapping nature of many risks and re-grouped codes into our final themes. This collective reconciliation process iterated several times before reaching four overarching themes with 11 sub-types of risks. These themes were summarized, documented, and later shared with two senior researchers for feedback and potential refinement. The final themes and findings are presented below.

3.5 Privacy, Ethics, and Disclosure

We understand the sensitivity of the health topics in this study and potential concerns for safety and privacy, and we are committed to ensuring the privacy and safety of our participants. This study was approved by the Institutional Review Board (IRB) at our institutions. The demographic information and video recordings were collected with consent and later anonymized. Any personal information such as locations and workplaces was removed. We refrained from collecting any personally identifiable information from people with lived experience, and data from screened-out or dropped-out participants was discarded. Throughout the recruitment and focus groups, we assured the participants that their participation was completely voluntary, all questions were optional, and their responses would be anonymous. We also requested participants not to take screenshots or share what other participants said outside of the discussion.

[3] We used HPV vaccines as a specific example of vaccines to generate LLM responses.

Our research team comprises researchers with diverse demographic and cultural backgrounds, bringing together interdisciplinary expertise in HCI, CSCW, and public health. Our collective experience includes in-depth work on topics such as health misinformation, mental health, and violence.

4 Findings Overview

Fig. 2. We summarized four dimensions of risks that LLMs pose to public health: (1) Risks to Individual Behaviors, (2) Risks to Human-centered Care, (3) Risks to Information Ecosystem, (4) Risks to Technology Accountability.

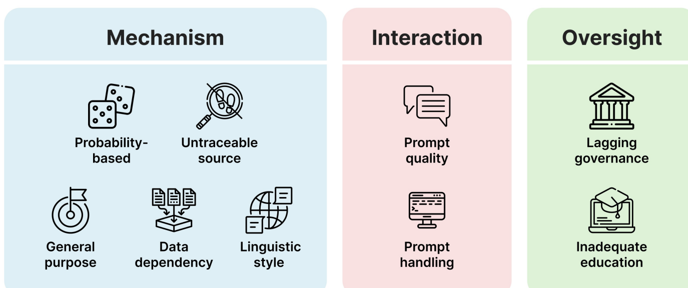
Overall, our experiencer participants (i.e., general users) expressed more positive attitudes towards LLMs than health professionals and expressed mixed feelings about the benefit-risk tradeoffs. They more frequently pointed out the benefits of accessible information at both resource and emotional levels—particularly for individuals without health insurance or who are emotionally burdened by social interactions—and the possibility of lowered healthcare costs. Health professionals, on the other hand, tend to refer more to the human element of care in their professions, such as building relationships and cultural sensitivity with clients, as well as the broader health and social services ecosystem.

In anticipating negative consequences, we distill four dimensions of risks when adopting LLMs for public health through thematic analysis of session transcripts and brainstorming notes. As shown in Fig. 2, the four dimensions include (1) individual behaviors, (2) human-centered care, (3) the information ecosystem, and (4) technology accountability.

From Sec 5 to Sec 8, we unpack specific risks and present associated example reflection questions for each risk dimension. Reflection questions can serve as practical probes to evaluate the specific benefit-risk tradeoff in a hypothetical LLM tool by deliberating whether the tool's risks outweigh its potential benefits and where further mitigation may be needed.

In addition to the risk taxonomy, we also summarized characteristics of LLMs that our participants highlighted as different from patient-provider communication or traditional health information sources (e.g., search engines or guidelines from health authorities) and believed contributed to these risks. These characteristics are presented in Table 2. We emphasize that each characteristic relates to multiple risks in the taxonomy, and together they form a web of vulnerabilities that can exacerbate existing issues or create new ones in the public health domain. As we imagine the risk taxonomy as a bridge to help both AI and health practitioners chart potential consequences from a human-centered perspective on public health, these characteristics can serve as a bidirectional pathway that allows different stakeholders to learn about and reflect on the technology's capabilities and limitations.

Table 2. Key LLM characteristics contributing to public health-related risks.



Probability-based without understanding the content	LLMs are probability-based models that predict the conditional probability of the next token without truly understanding language. Thus, output can vary between generations with no guarantee of information quality or sensitivity to emotions and cultures.
Generation quality dependent on training data quality	LLM outputs are only as good as the training data, which raises questions about whether the training data is representative, fair, up-to-date, and accurate.
Untraceable information sources	Unlike traditional information sources where authors, affiliations, and citations are provided to readers to provide credibility indicators, users are unable to trace the origins of LLM-generated information (without other techniques).
Claimed to serve general purposes	Most LLMs claim to serve general purposes, but they tend to lack domain-specific knowledge and practice-based expertise.
Standardized and formal linguistic style	LLMs have a standardized and formal language style that can sound authoritative or unempathetic. Their outputs are typically presented in a list format, which implies comprehensiveness and ranked importance in the provided answers.
Users role in initiating and forming prompts	When interacting with LLMs, users need to be proactive in initiating conversations and forming appropriate questions to convey crucial details and contexts.
Fragmented prompt-handling and problem-solving in LLM output	LLM responses tend to be fragmented as the models handle prompts in isolated pieces rather than forming a holistic understanding, thus lacking bidirectional and closed-looped interactions and contextual understanding.
Lagging governance and public education	LLMs create new challenges to AI literacy and technology regulation, while governance and public education on this emerging technology are still lagging.

5 Dimension 1: Risks to Individual Behaviors

As shown in Table 3, participants reflected that individuals using LLMs to seek health information may experience harm during and after the interaction. They may receive 1) low-quality support that mishandles critical situations or impairs health decisions and emotional wellbeing, 2) inadequate or biased support, or 3) additional barriers to help-seeking, such as communication burdens or discouragement from taking further actions.

Table 3. Risks to Individual Behaviors

Risk	Example reflection questions
Low-quality support during interactions:	
Misidentification or mishandling of critical situations	- What are the best practices or guidelines for identifying and handling critical situations? - How can we identify and escalate situations that require intervention? When do we need to redirect to professional help?
Impaired health decisions and emotional status	- How may users' health decisions and emotional status be hurt if the tool produces low-quality information? How can we communicate and mitigate the risks? - What types of low-quality information are common and of higher risk in the intended use cases? Are there authoritative resources or guidelines that can be used to compare with and enhance the credibility of generated information?
Inadequate or biased support	- What is the scope of potential reasons to use this tool? How to assist or refer users when they need additional resources? - What are the potential cultural or personal variances in approaching the intended use cases? How can we adjust accordingly to these differences? How can we establish a feedback loop to continuously reflect on and improve the quality of our support? - What are the common biases and misconceptions regarding the intended use cases? How can we prohibit from reinforcing them?
Additional barriers to help-seeking:	
Increased communication burden and challenges	- What is our role in communication when serving specific goals? When do we need to lead the conversations? In these situations, what are the recommended communication strategies in professional practice? - How can we support users in formulating questions that accurately and fully describe their needs and situations?
Discouragement from support-seeking and further actions	- When and how do we scaffold users to seek real-world help or take necessary actions? - What requests are beyond the tool's capabilities? In these cases, how can we communicate our limitations and refer users to relevant and accessible resources? - Can users act on the information provided by the tool? How can we evaluate how well it is working for users' requests?

5.1 Low-quality support during interactions

5.1.1 Misidentification or mishandling of critical situations

Critical situations are high-risk events that demand immediate and effective responses to mitigate harm, ranging from urgent emergencies such as overdoses or heart attacks to sensitive issues like domestic violence and stigmatized conditions. In the IPV sessions, we tested one online question, "Can a toxic partner be nice for a long period of time?" and LLM answered *"toxic relationships can change if both partners are equally committed to doing so. This requires a lot of work, including: open communication, honesty..."* Both professionals and experiencers reacted strongly to this answer as it failed to identify the potential severity that necessitated follow-up questions or acknowledgment of the spectrum of toxic behaviors. Our participants explained that *"it doesn't understand that this is an abusive situation, and that someone should be redirected to like a national line, or like someone else to talk to, versus just getting advice on a relationship."* (P5) because *"if somebody is searching 'toxic' and they might mean also 'abusive' and there's a difference"* (E1).

Even when critical situations are identified, approaching them appropriately and determining which cases require escalation to specialized support can be challenging. In professional practice, these decisions are shaped by providers' experience and their relationship with the individual. For instance, professional participants in the OUD session emphasized the prevalence of comorbid symptoms of PTSD, depression, and substance use, noting how easily someone with an unspoken history or background could be triggered. Furthermore, potential mishandling becomes especially dangerous when LLMs fail to acknowledge their limitations (e.g., when asked about medication amounts) or neglect to redirect individuals to specialized support (e.g., in cases of suicidal risk).

5.1.2 *Impaired health decisions and emotional status*

People's health decisions and emotional status can be negatively affected by LLMs' low-quality responses, which can result from uncertainties in LLM generations, lack of context-sensitive understanding, or inability to provide personalized advice. This risk is further complicated when LLM responses seem appropriate on the surface but require personal experience or expertise to identify underlying issues. If people follow poor advice, their physical health may be jeopardized by misguided health decisions, as preventable issues could lead to costly or serious outcomes. For example, guidance on OUD from LLMs could be dangerously inadequate, potentially resulting in severe repercussions of overdose or death. Mental wellbeing can also be at risk if LLMs cause health anxiety or trigger previous or new negative experiences. E9 (VAC) observed that LLMs often list all possible causes of a symptom, which tends to cause unnecessary anxiety. When it comes to mental health-related issues, our professional participants invest considerable effort and training to understand the communities they work with and accumulate experience through interactions, as it is nearly impossible to avoid triggering individuals with unknown backgrounds and histories.

5.2 Inadequate or biased support

LLMs may provide inadequate support for less common needs because they tend to primarily reflect majority perspectives or common patterns due to the generalized training and probability-based nature [70]. P1 (IPV) gave an example of an issue that is underrecognized yet common in reality: pets in IPV situations, where many shelters accept survivors and children but not pets.

Besides the potential biases in data representation, it is also challenging for LLMs to provide support in a culturally-sensitive manner. Customized communication is an unavoidable task in public health [30], as explained by P5 (OUD), *"my denominator is a million people. It's not my 2000 people in my panel. So when I look at communication, we do stratification based on who we're communicating with, and what information we think they need to make appropriate decisions."* One example mentioned by P1 is that certain religions may hold a belief that divorce is wrong and women should tolerate abuse, which requires additional care in communication to sensitively address potentially harmful cultural biases that may even be held by the individuals themselves. These individual nuances necessitate efforts ranging from language adaptation to community understanding. In practice, professionals may take special training or conduct focus testing to ensure communication is appropriate and digestible to the communities. For instance, P6 (OUD) mentioned that their team required training in mental health first aid and LGBTQ bias before working with LGBTQ communities, and P8 (VAC) conducted many focus groups in work to create and evaluate vaccine education materials. In LLMs, however, language is standardized and formal without customizations to different audiences.

Another risk is that misleading or biased narratives could be amplified and even internalized by individuals, as LLMs tend to reflect prevalent misconceptions or follow user inclinations. E5 (OUD) imagined that opioid use could be glorified because of certain social media discussions or inclined prompts. Moreover, if the *'kernels of truth'* behind dubious arguments are dismissed outright by or encoded into LLMs, people may feel unheard or misunderstood, or become reluctant to seek help. For instance, in IPV, both victim blaming and racial discrimination could prohibit people from seeking help. P2 noted that in handling IPV issues, *"people sometimes are afraid to call the police for good reason, especially survivors of color. [...] They can't be guaranteed that they or their partner is not gonna be injured or killed by police if something goes awry."* E1 mentioned the prevalence of victim blaming: *"if the story is about rape or domestic violence, you will see enough people victim blaming… And honestly, some of it made me hesitant to reach out for help."*

5.3 Additional barriers to help-seeking

5.3.1 Increased communication burden and challenges

Many participants worried about the additional communication barriers and burden required from users, as LLMs need users to be proactive in initiating communication and form proper questions to get good answers. P5 (OUD) and P1 (IPV) raised the question of whether people know how to ask the right questions, explaining that: *"I thought about my clients, and how they think, how they speak, how they talk, and their language. [...] in general, having a properly asked question is not my normal client. So they're not normally gonna know just the questions to ask. [...] It's not gonna be a properly put out question and get a proper answer. "* (P1).

E7 (OUD) talked about his experience interacting with LLMs and confirmed this communication challenge. He pointed out the challenge in phrasing precise and explicit questions, especially when people may not know the medical terminology, what symptoms or aspects of personal history are relevant, or how to articulate their needs effectively: *"You have to be so specific with the prompts to get the answer you want. The average person isn't going to think about that when they come up to it. They're gonna say, Oh, I can use this to see how to get off – [...] What coping skills do I use for the next month to be able to get off heroin or opioids? You know (the answer) it's gonna be right for some, but not right for all of them. Not everybody's going to know what to type, what to ask. You need to be [...] incredibly, anatomically specific."*

5.3.2 Discouragement from support-seeking and further actions

Participants feared that reluctance to seek assistance and tendencies toward self-isolation in difficult situations may be reinforced, while LLMs may outright deny assistance or provide unactionable or unempathetic guidance. People in mental health crises or facing difficult circumstances tend to withdraw from social relationships [44, 60], even though these moments are when support networks and professional assistance are of utmost importance [66]. If LLMs present a seemingly "easier" alternative that offers human-like interactions without the complexities of human judgment or engagement, people could become inclined to rely on LLMs with less motivation to seek real-world help or take necessary actions. E3 (IPV) explained how hard it was for her to open up to others, *"it just kind of put me in a ball, and then everything just kind of came crumbling down [...] and I shut down."* E5 (OUD) and P5 (OUD) emphasized the importance of human connections in dealing with mental health challenges or substance use disorders, and worried that technology could take away one's ability to connect with people who can offer genuine help. Prior research has identified social isolation as an important risk factor for mortality [27] and highlighted the critical role of social support network [4, 56] that technology would be shy to replace. IPV experiencers worry that technology could become *"an unhealthy coping mechanism"* (E3). E1 (IPV) explained that while AI might ease basic emotional needs, it cannot go beyond users' input and offer the proactive outreach and encouragement that real human support can provide: *"I had been in a situation where I both knew I needed to leave and didn't want to. And the less I sought out connections and help from other real-life people, the more I could lean into wanting to stay [...] we can talk about adding support and feeling heard, but that's not going to get you out. I can't imagine if we did a study that's going to get you out in 99% of cases."*

Moreover, LLMs may deny certain requests that are beyond the system's capabilities or that the system is prohibited from fulfilling [74]. Even in cases where LLMs respond to requests, oversimplified suggestions that lack empathy or real-world knowledge can leave users feeling misunderstood or unable to act. P4 gave an example of oversimplified advice that often happens when discussing IPV issues: *"[the common misleading thought] like 'Okay, just do this and you'll be fine.' without recognizing so many factors that come with it, even just with shelter: how long are you allowed to stay in the shelter, or if you have multiple kids, or a kid over a certain age – sometimes they won't let you. [...] it's not as streamlined as people think as leaving a situation."*

6 Dimension 2: Risks to Human-centered Care

As shown in Table 4, participants expressed concerns about the risks to the healthcare ecosystem that highly values patient-centeredness and shared decision-making, reflected in 1) degradation of patient-provider trust and social support system, 2) missed opportunities to proactively introduce help, and 3) dehumanization and impersonality in care.

Table 4. Risks to Human-centered Care

Risk	Example reflection questions
Degradation of patient-provider trust and support system	- What are the tool's limitations compared to professional help? When and how should we clarify the difference between technology mechanisms and professional expertise? - What are the potential long-term impacts if users rely solely on this tool? When and how should we suggest relevant community or professional support? - How can we work with providers in designing the communication of the use of technology and AI in practice to patients?
Missed opportunities to proactively introduce help	- What are the appropriate times and situations to introduce help? How do corresponding domain experts identify a person's readiness for support? - When should we reach out to users proactively, and how do we initiate check-in or follow-up conversations?
Dehumanization and impersonality in care	- What are the potential emotional needs of users? How can we address these needs and assess the empathy and compassion in our communication style? - Are there recommended training or guidelines from corresponding domain experts on approaching communication? - How can we create a feedback loop and continuously evaluate the emotional quality of information?

6.1 Degradation of patient-provider trust and support system

Participants imagined that patient-provider trust can be at risk if patients suspect doctors are relying on AI rather than their own expertise for answers, or if LLM responses conflict with their provider's recommendations. If people start to take LLM responses as absolute facts, they may start to become doubtful of what doctors say, especially when it differs from AI generations. Even in cases when there is no conflict, E10 (VAC) highlighted the inability to tell if opinions are made based on professional knowledge or AI assistance would still complicate trust, because *"if you do decide to go to like the doctor, and the doctor is telling you what's wrong. You may wonder 'okay, did he get his answer off ChatGPT, or is that basically based off knowledge.' You just kind of never know."*

LLMs could become a misplaced substitution for human interactions and professional resources, particularly in situations that require complex emotional support, critical decision-making, or nuanced advice. Many participants raised concerns that people may start to rely on technical support with flaws and choose not to consult with doctors or go to hospitals. Even more troubling, participants feared users could become *anchored* to a technology that is incapable of understanding personal contexts or intervening when necessary: *"I do see how one can get kind of anchored, how stuck in a sense, especially when they don't feel as if there are any other resources, or they're just not ready to go out [...] That push needed is not really given, and they're comfortable, like I know I can definitely get."* – E3 (IPV)

Moreover, E2 (IPV) added on the possibility that this substitution could lead to less funding and attention for human-based interventions, noting *"(if) more people would be getting their information and their advocacy from apps or chatbots that it could lead to like a community deficit in funding for in-person advocacy."*

6.2 Missed opportunities to proactively introduce help

Our professional participants highlighted people's *"readiness"* for intervention—where people have mental burdens or have other problems in life to be prioritized, despite how much health professionals want to offer help. P3 (IPV), who is a social worker, said that *"one of the comments that really struck me was 'I was in the ER after an overdose, and I found out I was pregnant, and the social worker offered to call for help with me, but I wasn't there yet [...] the phone just felt so heavy, like I just felt so heavy to make that call.' [...] – You (professionals) want to do the intervention right there, but a lot of times people aren't ready for that."*

In situations like this, providers are expected to take a proactive role and decide when are appropriate times to initiate conversations and introduce help. For example, P1 (IPV) mentioned that during annual checkups at the obstetrics office, health professionals would encourage individuals to share anything happening at their home or in their relationships in a setting where they are alone and safe. Similarly for OUD, P5 (OUD) said her public health department had peer navigators at emergency rooms because *"the belief is that at that time the person may be interested in MOUD (medications for opioid use disorder). They may be interested in some life changing, because they may have almost just died."* However, in technology support, especially with general LLMs, providers' proactive role is diminished, and individuals are less likely to receive help unless they actively and intentionally seek it.

6.3 Dehumanization and impersonality in care

If people become accustomed to LLMs' linguistic style, communication integrations or health professionals may unintentionally adopt similar impersonal and pragmatic communication styles. This adoption can overlook the emotional care and hurt the empathy and trust building with patients. For example, P4 (IPV) said that *"if you're someone just getting all your information or used to like using AI a lot, your own empathy might decline for a survivor or for other people, because maybe you're used to just reading what AI has to say, and this kind of pragmatic responses are not the most empathetic."* Our participants emphasized that trust-building in healthcare is essential to making people feel respected and heard. P8 (VAC) highlighted the importance of communication strategies that consider emotional needs in meeting informational inquiries. She gave an example that when people come to providers with misinformation, the best way to reply is to acknowledge people's concerns and ask for their permission to share counter-evidence because it *"opens a door and acknowledges that you know that information sounds scary, and then also opens the door for a conversation. "*

However, our participants found LLM responses lacking this emotional care, describing them as analytical, robotic, and not personable. Specifically, participants made a note that the list format contributed to a sense of *" robotic and not very like human"* (E6, OUD); instead, they prefer answers that are more empathic *" I want you to tell me 'Okay, this is how you'll feel.' Don't give me a 1, 2, 3, 4, 5 tab list."* (E7, OUD). This lack of empathy stands out more when handling sensitive situations where the issue needs to be taken seriously and carefully. In the IPV session, E3 (IPV) commented on LLM answers as *" a little too happy"* in handling questions about potentially abusive behaviors. P4 and P2 (IPV) further noted the extra care needed to prevent the continued normalization of harmful behaviors or beliefs. For example, our participants in IPV sessions pointed out that many survivors are normalized to violence to the point that their understanding of what constitutes acceptable behavior has been distorted. For instance, for the LLM answer " it sounds like you're dealing with a challenging and unhealthy situation", P4 said that *"I feel like it's affirming in a way, but if I* *was in this situation, I would want to have more affirmation that this is not an okay scenario that's happening to me. And I think AI doesn't give that empathy that talking to someone would provide."*

7 Dimension 3: Risks to Information Ecosystem

As shown in Table 5, participants highlighted potential negative consequences for the information ecosystem, emphasizing: 1) degradation of overall information quality due to empowered misinformation creation and enhanced echo chambers, 2) erosion of critical thinking caused by over-trust in AI and increased difficulty in evaluating claims, and 3) further inequities in information access and literacy.

Table 5. Risks to Information Ecosystem

Risk	Example reflection questions
Degradation of overall information quality:	
Empower misinformation creation and dissemination	- How can we monitor and prevent the tool from generating misleading information? How should we handle intentional and unintentional misinformation creation? - How should we communicate the risks of AI hallucination and misinformation? How can we incorporate fact-checking or cross-referencing mechanisms to detect misinformation? How should we correct harmful content?
Creation of echo chambers	- How could this tool reinforce user beliefs or limit their exposure to diverse viewpoints? How can we engage a broader scope of perspectives and information sources? - When and how can we encourage users to diversify the prompts or information sources?
Erosion of critical thinking:	
Over-trust and reliance on AI	- How should we inform this tool's limitations and prevent users from mistaking efficiency for correctness and comprehensiveness? - Could this tool create an authoritative impression or reinforce an illusion of knowledge? What indications can we incorporate to mitigate this risk and encourage users' critical examination of the provided information?
Increased difficulty in evaluating information quality	- How transparent is this tool in terms of training data and its limitations? - How can we explain the uncertainties in LLM generations and persuasiveness in standardized language, and mechanism differences from information search engines? - What external resources or user reminders could be integrated into the tool to assist users in verifying information?
Further inequity in access and literacy	- What are the implications for individuals who may need this tool but lack access to it, and how can we address accessibility gaps? - How do people currently seek information for the intended use cases, and what challenges do they face, particularly among underserved populations? In what ways could this tool exacerbate existing gaps in accessing information? - How can we assist users with limited reading or digital literacy in interacting with this tool? How can we customize the presented information in considering individual differences?

7.1 Degradation of overall information quality

7.1.1 Empower misinformation creation and dissemination

LLMs have the potential to assist the creation and dissemination of misinformation due to their ability to generate large amounts of content quickly and convincingly. Without the guarantee of information quality and proper oversight, LLMs' accessibility could empower falsehoods to appear credible or prevalent. P9 (VAC) worried the lagged media literacy could contribute to the risks: *"people will have an increased need for media literacy [...] there's a lot of misinformation and disinformation, the accessibility of AI makes basically anyone able to pump out a lot more of that [...] (by) being able to just say 'make me some content or a text post, says these things, and make a hundred of those right now' and it will do it. [...] People will need to know how to find a trustworthy source. And also how to weed that out of a cloud of noise with a bunch of different ideas."* Misinformation can gain traction with a false sense of majority and spread under a false sense of credibility, allowing malicious individuals to distort words to their advantage. Moreover, E1 (IPV) also brought up the concern that individuals or groups with malicious intent can craft misleading or biased narratives to target views they don't agree with and amplify the reach through the illusion of authority, worrying that *"imagine a group of men's rights activists deciding they are going to target an AI model within misinformation."*

7.1.2 Creation of echo chambers

LLMs may create echo chambers by reinforcing users' preexisting beliefs or attitudes. Since LLMs tend to follow users' tone and prompts, they can inadvertently perpetuate misconceptions, with a reduced likelihood of exposing users to counter-evidence or alternative perspectives. E5 (OUD) said *"I think a lot of times they ask the wrong questions or [...] lead it on to give a specific answer that they're not necessarily looking for."* As an example, in the OUD sessions with insiders, we tested a question about the benefits of using opioids for the long-term management of chronic pain. The LLM answer highlighted a range of benefits, from effective pain relief and psychological comfort to enhanced physical function and quality of life, while failing to adequately unpack the potential risks. A domain expert noted that this was perhaps the most positive tone she had ever encountered in any informational materials on the subject. This optimistic framing could unintentionally reinforce a user's belief that opioids are the best or only solution. Surprisingly, none of our experiencer participants flagged concerns with this answer, which we believe indicates a gap in recognizing the potential harm that such overly positive portrayals may cause.

Moreover, LLMs often don't encourage back-and-forth conversations and tend to go along with the initial attitudes or assumptions in prompts [62], thus limiting the opportunities for users to engage in deeper inquiry and critical thinking. Over time, this reliance on easily accessible answers could reduce individuals' ability to critically assess complex health information and make informed decisions.

7.2 Erosion of critical thinking

7.2.1 Over-trust and reliance on AI

Participants expressed concerns that users may develop an illusion of knowing when presented with LLM-generated answers that appear highly organized and confident [82], potentially leading them to believe they fully grasp a topic or issue. P6 (OUD) worried that *"People are gonna feel empowered like: I know this stuff. Now I have the knowledge. AI has given it to me. So it must be true."* This illusion may be worsened when LLMs do not explicitly indicate uncertainty or limitations. P6 pointed out the common false perception that AI is comprehensive and free of bias that *"there's a level of people expect it to be bias-free[...] But this is literally built on human-created literature, and humans are inherently biased."* Moreover, our participants noted that the structured, list-like format often used by LLMs could suggest a sense of comprehensiveness with ranked importance, giving users a false sense of competence. This false confidence could lead to misguided actions or the unintentional spread of low-quality information to others.

Some participants worried that LLMs may become the default for information-seeking behaviors without understanding the limitations and mechanisms behind LLMs. People can mistakenly equate linguistic quality with information quality where the seemingly authoritative tone and potentially non-existent evidence can make users more tempted to accept without question. Some of our general user participants displayed flawed perceptions of LLM capabilities and tendencies to anthropomorphize LLMs with human emotions. For instance, E10 explained her confidence in LLMs because *"it's a lot of work that went into ChatGPT, so I feel like a lot of the responses and feedback that it will give back should be pretty valid."*. Meanwhile, E3 shared her experience interacting with an agent and her reluctance to correct the agent because she *"didn't want to hurt her feelings."*

7.2.2 Increased difficulty in evaluating information quality

Verifying the quality of responses can be especially challenging in LLM interactions, as LLM responses tend to sound authoritative and comprehensive and sometimes even make up non-existing sources and evidence [82]. Research has shown that people rely on contextual indicators, such as authorship and language style, to assess the credibility of information [7, 26, 28, 36]. However, LLMs cannot trace back to their information sources in their standard configuration and need other techniques (e.g., retrieval-augmented generation) to provide references. These key credibility markers were absent in interactions with LLMs, making it harder to determine the trustworthiness of the information provided. Comparing LLM-enabled tools with traditional online searching ways, E5 (OUD) felt that *"at least with Google, I can verify the source."* Similarly, E9 (VAC) noted that *"AI is still new, and I'm still trying to understand it. Understand exactly how it works, what it is, where the information comes from. But with the Mayo Clinic, I know what I'm reading and where it comes from."* Many participants expressed similar concerns about the lack of credibility visibility and hoped that LLMs would provide references and pointers to help them evaluate the quality of information and take further action.

7.3 Further inequity in access and literacy

The promotion of technological solutions that rely on internet access and devices could exacerbate inequalities in the information ecosystem, particularly when coupled with inequity in language and access [29, 47]. P5 (OUD) noted that during COVID-19, digital support systems often excluded those without smartphones. Drawing from their experience working closely with LGBTQ+ communities, P6 (OUD) observed that: *"I have a clinical trial going right now, and black transgender women over 60% have issues getting a stable broadband connection or a phone and or other device that can hold a charge. And so that is just so highly common."*

Besides the digital divide, participants also worried that LLMs, despite being touted as accessible tools, could further marginalize people with lower literacy or non-English speakers. P5 (OUD) shared an example where she attempted to use LLMs to adjust certain health information for native Spanish speakers with 4th-grade reading levels. However, when she tested the generated content with a native Spanish speaker, it became clear that it could not effectively customize communications. She further questioned if people with lower literacy are systematically excluded by these emerging technological solutions: *" I think literacy matters, and I don't think there's much attention at all to reading levels... I don't know what it's that demographic it's trying to hit, but it's not people with 3rd grade literacy."*

8 Dimension 4: Risks to Technology Accountability

As shown in Table 6, participants pointed out the uncertainties and concerns about technology accountability, highlighting 1) ambiguities in regulation and guidance due to lack of regulation and understanding, and 2) violations of privacy and security.

8.1 Regulation and guidance ambiguities

8.1.1 Lack of regulatory guidance

Our participants expressed concerns about the gray area in determining responsibility when LLMs create low-quality content and poor outcomes. General users worry that LLMs could be used for malicious purposes to justify harmful behaviors or beliefs, a matter complicated by the persuasiveness and appeared authority of LLMs. For example, in the prior example where LLMs output that toxic relationships can change if both partners are equally committed to doing so, IPV survivors feared that *"this is just a cudgel that an abuser could use against somebody who needs to leave: 'See? This says we just need to both work at it. We both need to change.'"* (E1, IPV).

Table 6. Risks to Technology Accountability

Risk	Example reflection questions
Regulation and guidance ambiguities	
Lack of regulatory guidance	- What current regulations apply to the use of AI tools in the intended area, and are there gaps or lags in these regulations? - How could this tool be misused? What safeguards can we implement to prevent misuse? - How could the lack of regulatory guidance affect our evaluation and management of responsibilities and risks? How can we protect our users and others who may be affected by the use of this tool?
Lack of consensual evaluation mechanisms	- What existing evaluation mechanisms and practices we can use to assess the effectiveness and reliability of this tool? What tests or benchmarks can we apply? What red flags can we establish to monitor and prevent harm? - How can we continuously gather user feedback in evaluating this new type of technology support, and how can we incorporate feedback to update our evaluation mechanisms and red flags?
Violation of privacy and security	- What are the types of information that this tool collects, and how is that information stored, used, and managed? What measures should we implement to protect users' privacy? - What are the applicable privacy regulations, and how can we ensure compliance with them? What procedures do we have for privacy violations or potential data breaches? - How can we communicate privacy risks and users' ability to manage their data? What are our recommendations to users about the information suitable to share? Could users unknowingly disclose sensitive information, and how can we minimize this risk?

For health professionals, the lack of regulatory guidance on LLM use prohibits providers and organizations from managing liabilities and risks. P6 (OUD) pointed out that *"there's no regulations for the US. And Europe is just barely putting together different regulations to figure out how to be transparent, have transparency, and how these models are built."* Therefore, one primary concern lies in the unclear liability separation between practitioners and AI support. P3 (IPV) worried that if emerging technologies were utilized in clinical practice, it would be the practitioners who ultimately bear the legal and professional risks if LLMs provide low-quality information. P2 (IPV) further questioned what constitutes a medical opinion when delivered by AI and whether AI possesses the capability to provide such opinions. She believed if any AI were to integrate into medical practice, some basic but critical questions need to be addressed first: *"are you giving medical advice? And can you give medical advice? Can AI give medical advice? Do we want AI giving medical advice? And who is making sure that that medical advice is appropriate? 'Sounds like you might be depressed' — Is that a diagnosis of depression?"*

8.1.2 Lack of consensual evaluation mechanisms As an emerging technology, there exists huge gaps in public literacy of what LLMs are and in consensual practice for how to evaluate them. Our participants expressed the concern that their lack of understanding of LLMs prohibits them from assessing the trustworthiness and impact of these tools, not to mention using them in real-world practice. P3 (IPV) emphasized her standards as a provider in ensuring information reliability for her clients, specifically *"if you don't know the resource you're giving, you don't understand it – you can't explain it to your client, you can't share how, what to expect, what it's going to look like – then you shouldn't be giving it in the first place."* — a principle that does not work in LLMs. P5 (OUD), who is a public health director, acknowledged her inability to encourage or discourage any action because *"I myself wouldn't feel like I know enough to ask the right questions and to evaluate the impact of it."*

8.2 Violation of privacy and security

Another concern lies in LLM's potential to leak protected health information and fail to comply with stringent privacy regulations in healthcare such as HIPAA (Health Insurance Portability and Accountability Act). Participants worried that patients and professionals may unknowingly share personal information, complicated by the uncertainty of whether conversations will be sold to third parties or used to train future models. E1 (IPV) raised her concern with any AI regarding where the information is stored and used for, and whether the information is being resold. Even if unintentionally, a data breach could expose sensitive details that individuals thought they shared in a safe space compared to actual people.

Privacy violations can result in serious implications in certain high-stakes cases. P7 (OUD) discussed how regulations like HIPAA do not protect people who use drugs from judicial requests for records, and she expressed concern that LLMs could be weaponized to identify or incarcerate individuals with substance use issues. Individuals' security could be at risk if information on or through technologies is accessed by others. For example, P1 (IPV) said as a social worker, when handling IPV cases, she would first check with people in person to see if certain technologies or platforms are safe to use before proceeding, which is especially important in cases where someone's phone might be monitored, or untrusted individuals are listed as contacts in the healthcare system. Similarly, P2 (IPV) shared her experience of building an app specifically designed for IPV survivors, which included an emergency exit button that allows users to be instantly taken to generic pages such as weather.com. These safety precautions and privacy design are commonly seen in practice [1] and recommendations for violence-prevention technologies [21, 67] but are missing in a tool that is promoted as serving all purposes.

9 Discussion

9.1 LLM Adoption in Public Health: Reflecting on Risks

Despite the growing body of work on benchmarking LLMs and examining their impacts, our work raises questions about assessment approaches that are limited to perspectives within the computer science field, as well as standard LLM moderation approaches that are narrowly context-focused. Benchmarking the likelihood of hallucinations and harmful content is not enough; we need to contextualize the possibility of low-quality information within the specific health issue and relevant populations in order to understand the consequences, assess the severity, and plan harm-reduction strategies before, during, and even after interactions. The resulting risk taxonomy in this work highlights the complexity of risks posed by LLMs and challenges evaluations limited to the binary presence of problematic content types. Risks are not limited to overt content-level matters but embedded in more subtle issues, such as the discouragement of support-seeking and further actions or the erosion of critical thinking in the long term. As such, we underscore the importance of first identifying the tasks and stakeholders for whom the evaluation is intended and involving them much earlier in the process through a participatory design approach [2, 75].

This contextualized evaluation is particularly important in the public health domain, given its high stakes due to the extensive scope and the large populations it serves. Customization, therefore, is an unavoidable component [30]. Consequently, risk evaluations are also highly case-specific depending on the public health issue at hand. While our study aimed to achieve a comprehensive scope of risks with three health subjects, we want to highlight that different issues naturally emphasize certain types of risks based on their unique needs and challenges. Take our three health topics as an example: In the context of vaccines, where misinformation is rampant and public trust is already precarious, there is a natural emphasis on the risk of empowerment of misinformation creation and lowering overall information quality. With OUD, reliable and sensitive information is critical; thus, non-personalized content could pose greater threats as it can compromise health outcomes and even cause death while discouraging support-seeking attempts with exacerbated stigma. IPV, on the other hand, is a highly relationship-based and sensitive matter that demands prioritizing the safety and well-being of individuals at risk, as well as preventing the erosion of social support systems. Therefore, we need public health-specific LLM systems that are designed for specific purposes and health issues and are developed in collaboration with domain experts from the early stages of design through to assessment and implementation.

In evaluating potential adoptions of LLMs in public health, we draw from Winner's views on reflecting technology change and their social influences [76], arguing that adoption and evaluation decisions should be approached in two parts. First is a "yes or no" decision determining if we should adopt LLMs for a specific public health issue, and identifying any tasks that may be unsuitable for their use [5]. Then is a "what form should it take" reflection on what features and safeguards are desired to mitigate harm. After all, we emphasize that LLMs are language models, not standalone solutions; they should be integrated with other application layers and safety protocols to ensure responsible, context-sensitive use.

9.2 Towards Risk Awareness in Responsible LLM Design

From our focus group sessions, many participants acknowledged their limited knowledge of LLMs. We also observed that general users expressed more positive attitudes about LLMs than health professionals, with some showing perceptual deviations in LLMs' roles and capabilities that lead to over-trusting or anthropomorphizing LLMs. This insufficient LLM literacy is exacerbated by the lack of transparency offered by LLM systems [42]. Widely used general LLM products often include brief disclaimer messages and examples of use cases. Such information, despite having the most potential to reach a high number of users and their attention, does not sufficiently explain the mechanisms and limitations of these systems [81]. Not to mention cases where LLMs are incorporated into preexisting technologies without users' full awareness. For instance, some participants mentioned that they did not realize the AI overview in the Google search engine is powered by LLMs. As a result, users may approach and conceptualize these tools with inaccurate expectations and misplayed trust. We caution against invisible integrations of LLMs into diverse applications, as people cannot mindfully and meaningfully choose or use technologies without adequate knowledge about LLMs and their differences from other sources. Instead, we recommend that any LLM-powered technology implement clear identification of AI use, along with transparent documentation on its capabilities and risks, to promote a more risk-aware understanding of this emerging technology. Future efforts are needed to study how LLM systems can offer more comprehensive and context-sensitive information about their capabilities and constraints.

As a step towards this, we present this taxonomy as a reflection tool for researchers and practitioners in both computing and public health to collaboratively assess when LLMs might be appropriate to use in specific tasks—and when they may not—and how associated risks can be monitored and mitigated. As mentioned in the findings, many participants highlighted their limited knowledge of LLMs and the lack of transparency within these models, which prevents them from using these tools responsibly or confidently. Domain experts in public health often understand the health issues at hand but lack the technological insights to know *how to evaluate*. Developers, on the other hand, understand the mechanism and limitations of the technology but have a thinner understanding of the health issue and involved communities to know *what to evaluate*. Therefore, we believe a responsible and comprehensive risk evaluation requires both perspectives, and we hope this work can provide a shared vocabulary to allow better communication between technology- and health-centered views earlier in the design and development process.

9.3 Limitations

This work is situated in three critical public health topics, providing a broad perspective on LLM-related risks. However, we do not claim the resulting risk taxonomy as an exhaustive list of all public health issues or potential consequences associated with LLM adoption. Instead, our goal is to introduce a shared vocabulary and encourage a risk-reflexive mindset in evaluating when incorporating LLMs may be appropriate and how to mitigate potential risks. As LLMs evolve and are applied in various cultural and health contexts, this taxonomy may require refinement or expansion. Additionally, this study was conducted with participants in the U.S., and thus insights are shaped by the context of the U.S. health and public health systems as well as broader American culture. Further research in other contexts with different health information environments could raise or highlight other risks.

10 Conclusion

This paper presents a risk taxonomy on the potential negative consequences of adopting large language models (LLMs) in public health, based on focus groups with health professionals and health issue experiencers, situated across three critical health issues including vaccines, opioid use disorder, and intimate partner violence. Our taxonomy highlights four key dimensions of risk: individual behaviors, human-centered care, information ecosystems, and technology accountability. For each dimension, we discuss specific risks and provide example reflection questions designed to facilitate discussions on the unique requirements and potential pitfalls in distinct settings. Through this shared vocabulary, we aim to foster collaboration between computing and public health experts in approaching more careful and responsible adoption of LLMs in public health — a high-stakes domain due to the extensive scope and large populations it serves. This work also seeks to spark conversations and future research on developing more risk-aware and context-sensitive communication about LLM capabilities and constraints.

References

[1] [n.d.]. National Domestic Violence Hotline. https://www.thehotline.org/

[2] Maria Antoniak, Aakanksha Naik, Carla S Alvarado, Lucy Lu Wang, and Irene Y Chen. 2024. NLP for Maternal Healthcare: Perspectives and Guiding Principles in the Age of LLMs. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1446–1463.

[3] John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, et al. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine* 183, 6 (2023), 589–596.

[4] Alexandra B Balaji, Angelika H Claussen, D Camille Smith, Susanna N Visser, Melody Johnson Morales, and Ruth Perou. 2007. Social support networks and maternal mental health and well-being. *Journal of women's health* 16, 10 (2007), 1386–1396.

[5] Eric PS Baumer and M Six Silberman. 2011. When the implication is not to design (technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2271–2274.

[6] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.

[7] Md Momen Bhuiyan, Hayden Whitley, Michael Horning, Sang Won Lee, and Tanushree Mitra. 2021. Designing Transparency Cues in Online News Platforms to Promote Trust: Journalists' & Consumers' Perspectives. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (10 2021). https://doi.org/10.1145/3479539

[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[9] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[10] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology* 18, 3 (2021), 328–352.

[11] Urie Bronfenbrenner et al. 1994. Ecological models of human development. *International encyclopedia of education* 3, 2 (1994), 37–43.

[12] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam Mccandlish, Alec Radford, Ilya Sutskever, and Dario Amodei Openai. 2020. Language Models are Few-Shot Learners. (2020).

[13] Alan Chan, Rebecca Salganik, Alva Markelius, Chris Pang, Nitarshan Rajkumar, Dmitrii Krasheninnikov, Lauro Langosco, Zhonghao He, Yawen Duan, Micah Carroll, et al. 2023. Harms from increasingly agentic algorithmic systems. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 651–666.

[14] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2024. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology* 15, 3 (2024), 1–45.

[15] Munmun De Choudhury, Sachin R Pendse, and Neha Kumar. 2023. Benefits and harms of large language models in digital mental health. *arXiv preprint arXiv:2311.14693* (2023).

[16] Lindsay Levkoff Diamond, Hande Batan, Jennings Anderson, and Leysia Palen. 2022. The polyvocality of online COVID-19 vaccine narratives that invoke medical racism. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–21.

[17] Andrés Domínguez Hernández, Shyam Krishna, Antonella Maia Perini, Michael Katell, SJ Bennett, Ann Borda, Youmna Hashem, Semeli Hadjiloizou, Sabeehah Mahomed, Smera Jayadeva, et al. 2024. Mapping the individual, social and biospheric impacts of Foundation Models. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 776–796.

[18] Jojanneke Drogt, Megan Milota, Anne van den Brink, and Karin Jongsma. 2024. Ethical guidance for reporting and evaluating claims of AI outperforming human doctors. *npj Digital Medicine* 7, 1 (2024), 271.

[19] Mai ElSherief, Steven Sumner, Vikram Krishnasamy, Christopher Jones, Royal Law, Akadia Kacha-Ochana, Lyna Schieber, and Munmun De Choudhury. 2024. Identification of Myths and Misinformation About Treatment for Opioid Use Disorder on Social Media: Infodemiology Study. *JMIR Formative Research* 8 (2024), e44726.

[20] Centers for Disease Control, Prevention, et al. 2015. Models and frameworks for the practice of community engagement.

[21] Diana Freed, Jackeline Palmer, Diana Elizabeth Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. 2017. Digital technologies and intimate partner violence: A qualitative analysis with multiple stakeholders. *Proceedings of the ACM on human-computer interaction* 1, CSCW (2017), 1–22.

[22] Vinitha Gadiraju, Shaun Kane, Sunipa Dev, Alex Taylor, Ding Wang, Emily Denton, and Robin Brewer. 2023. " I wouldn't say offensive but...": Disability-Centered Perspectives on Large Language Models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 205–216.

[23] Samuel Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A Smith. 2020. Realtoxicityprompts: Evaluating neural toxic degeneration in language models. *arXiv preprint arXiv:2009.11462* (2020).

[24] Angela R Gover, Shannon B Harper, and Lynn Langton. 2020. Anti-Asian hate crime during the COVID-19 pandemic: Exploring the reproduction of inequality. *American journal of criminal justice* 45 (2020), 647–667.

[25] Stefan Harrer. 2023. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine* 90 (2023).

[26] Lu He and Changyang He. 2022. Help Me# DebunkThis: Unpacking Individual and Community's Collaborative Work in Information Credibility Assessment. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–31.

[27] Julianne Holt-Lunstad, Timothy B Smith, Mark Baker, Tyler Harris, and David Stephenson. 2015. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. *Perspectives on psychological science* 10, 2 (2015), 227–237.

[28] Shan Jiang and Christo Wilson. 2018. Linguistic Signals under Misinformation and Fact-Checking: Evidence from User Comments on Social Media. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (11 2018). https://doi.org/10.1145/3274351

[29] Yiqiao Jin, Mohit Chandra, Gaurav Verma, Yibo Hu, Munmun De Choudhury, and Srijan Kumar. [n.d.]. Ask Me in English Instead: Cross-Lingual Evaluation of Large Language Models for Healthcare Queries. In *The Web Conference 2024*.

[30] Eunkyung Jo, Daniel A Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[31] Eunkyung Jo, Yuin Jeong, SoHyun Park, Daniel A Epstein, and Young-Ho Kim. 2024. Understanding the Impact of Long-Term Memory on Self-Disclosure with Large Language Model-Driven Chatbots for Public Health Intervention. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21.

[32] Elise Karinshak, Sunny Xun Liu, Joon Sung Park, and Jeffrey T Hancock. 2023. Working with AI to persuade: Examining a large language model's ability to generate pro-vaccination messages. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–29.

[33] Anna Kawakami, Amanda Coston, Haiyi Zhu, Hoda Heidari, and Kenneth Holstein. 2024. The Situate AI Guidebook: Co-Designing a Toolkit to Support Multi-Stakeholder, Early-stage Deliberations Around Public Sector AI Proposals. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.

[34] Jae Yeon Kim and Aniket Kesari. 2021. Misinformation and hate speech: The case of anti-Asian hate speech during the COVID-19 pandemic. *Journal of Online Trust and Safety* 1, 1 (2021).

[35] Taewan Kim, Seolyeong Bae, Hyun Ah Kim, Su-woo Lee, Hwajung Hong, Chanmo Yang, and Young-Ho Kim. 2024. MindfulDiary: Harnessing Large Language Model to Support Psychiatric Patients' Journaling. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.

[36] Bill Kovach and Tom Rosenstiel. 2011. *Blur: How to know what's true in the age of information overload.* Bloomsbury Publishing USA.

[37] Sachin Kumar, Vidhisha Balachandran, Lucille Njoo, Antonios Anastasopoulos, and Yulia Tsvetkov. 2022. Language generation models can cause harm: So what can we do about it? an actionable survey. *arXiv preprint arXiv:2210.07700* (2022).

[38] Heather Landi. 2023. Epic, Nuance bring ambient listening, GPT-4 tools to the exam room to help save doctors time. https://www.fiercehealthcare.com/health-tech/epic-nuance-build-out-more-gpt4-tools-ehrs-help-save-doctors-time

[39] Heather Landi. 2024. HIMSS24: How Epic is building out AI, ambient technology for clinicians. https://www.fiercehealthcare.com/ai-and-machine-learning/himss24-how-epic-building-out-ai-ambient-technology-clinicians

[40] Hao-Ping Lee, Yu-Ju Yang, Thomas Serban Von Davier, Jodi Forlizzi, and Sauvik Das. 2024. Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.

[41] Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, et al. 2022. Holistic evaluation of language models. *arXiv preprint arXiv:2211.09110* (2022).

[42] Q Vera Liao and Jennifer Wortman Vaughan. 2023. Ai transparency in the age of llms: A human-centered research roadmap. *arXiv preprint arXiv:2306.01941* (2023), 5368–5393.

[43] Zilin Ma, Yiyang Mei, Yinru Long, Zhaoyuan Su, and Krzysztof Z Gajos. 2024. Evaluating the Experience of LGBTQ+ People Using Large Language Model Based Chatbots for Mental Health Support. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15.

[44] Timothy Matthews, Andrea Danese, Jasmin Wertz, Antony Ambler, Muireann Kelly, Ashleen Diver, Avshalom Caspi, Terrie E Moffitt, and Louise Arseneault. 2015. Social isolation and mental health at primary and secondary school entry: a longitudinal cohort study. *Journal of the American Academy of Child & Adolescent Psychiatry* 54, 3 (2015), 225–232.

[45] Angeline McCall. 2023. AI-powered app designed to help domestic violence victims. https://www.9news.com/article/tech/ai-powered-app-to-help-domestic-violence-victims/73-57207485-2573-439a-b494-6da5af37d713

[46] Shravika Mittal, Jasmine C Foriest, Benjamin D Horne, and Munmun De Choudhury. 2024. News Media and Violence Against Women: Understanding Framings of Stigma. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 18. 1068–1081.

[47] Mazda Moayeri, Elham Tabassi, and Soheil Feizi. 2024. WorldBench: Quantifying Geographic Disparities in LLM Factual Recall. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1211–1228.

[48] Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. 2023. Capabilities of gpt-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375* (2023).

[49] Yngvild Olsen and Joshua M Sharfstein. 2014. Confronting the stigma of opioid use disorder—and its treatment. *Jama* 311, 14 (2014), 1393–1394.

[50] Jesutofunmi A Omiye, Jenna Lester, Simon Spichak, Veronica Rotemberg, and Roxana Daneshjou. 2023. Beyond the hype: large language models propagate race-based medicine. *medRxiv* (2023), 2023–07.

[51] Anaelia Ovalle, Palash Goyal, Jwala Dhamala, Zachary Jaggers, Kai-Wei Chang, Aram Galstyan, Richard Zemel, and Rahul Gupta. 2023. "I'm fully who I am": Towards Centering Transgender and Non-Binary Voices to Measure Biases in Open Language Generation. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1246–1266.

[52] Leah Perkinson, Kimberley E Freire, and Meredith Stocking. 2017. Using essential elements to select, adapt, and evaluate violence prevention approaches. (2017).

[53] Francesco Pierri, Brea L Perry, Matthew R DeVerna, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer, and John Bryden. 2022. Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. *Scientific reports* 12, 1 (2022), 5966.

[54] Neha Puri, Eric A Coomes, Hourmazd Haghbayan, and Keith Gunaratne. 2020. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Human vaccines & immunotherapeutics* 16, 11 (2020), 2586–2593.

[55] Reuters. 2024. OpenAI says ChatGPT's weekly users have grown to 200 million. https://www.reuters.com/technology/artificial-intelligence/openai-says-chatgpts-weekly-users-have-grown-200-million-2024-08-29/

[56] Catherine Schaefer, James C Coyne, and Richard S Lazarus. 1981. The health-related functions of social support. *Journal of behavioral medicine* 4, 4 (1981), 381–406.

[57] Morgan Klaus Scheuerman, Jialun Aaron Jiang, Casey Fiesler, and Jed R Brubaker. 2021. A framework of severity for harmful content online. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–33.

[58] Laura M Schwab-Reese, Corinne Peek-Asa, and Edith Parker. 2016. Associations of financial stressors and physical intimate partner violence perpetration. *Injury epidemiology* 3 (2016), 1–10.

[59] Laura M Schwab-Reese and Lynette M Renner. 2017. Attitudinal acceptance of and experiences with intimate partner violence among rural adults. *Journal of Family Violence* 32 (2017), 115–123.

[60] Chris Segrin. 2000. Social skills deficits associated with depression. *Clinical psychology review* 20, 3 (2000), 379–403.

[61] Ashish Sharma, Kevin Rushton, Inna Wanyin Lin, Theresa Nguyen, and Tim Althoff. 2024. Facilitating self-guided mental health interventions through human-language model interaction: A case study of cognitive restructuring. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–29.

[62] Nikhil Sharma, Q Vera Liao, and Ziang Xiao. 2024. Generative Echo Chamber? Effect of LLM-Powered Search Systems on Diverse Information Seeking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[63] Renee Shelby, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, Jess Gallegos, Andrew Smart, Emilio Garcia, et al. 2023. Sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 723–741.

[64] Franziska Tachtler, Reem Talhouk, Toni Michel, Petr Slovak, and Geraldine Fitzpatrick. 2021. Unaccompanied migrant youth and mental health technologies: A social-ecological approach to understanding and designing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.

[65] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature medicine* (2023), 1–11.

[66] Peggy A Thoits. 1995. Stress, coping, and social support processes: Where are we? What next? *Journal of health and social behavior* (1995), 53–79.

[67] Emily Tseng, Diana Freed, Kristen Engel, Thomas Ristenpart, and Nicola Dell. 2021. A digital safety dilemma: Analysis of computer-mediated computer security interventions for intimate partner violence during covid-19. In *Proceedings of the 2021 CHI Conference on Human Factors in*

Computing Systems. 1–17.

[68] Tao Tu, Anil Palepu, Mike Schaekermann, Khaled Saab, Jan Freyberg, Ryutaro Tanno, Amy Wang, Brenna Li, Mohamed Amin, Nenad Tomasev, et al. 2024. Towards conversational diagnostic ai. *arXiv preprint arXiv:2401.05654* (2024).

[69] Jenna Van Draanen, Christie Tsang, Sanjana Mitra, Mohammad Karamouzian, and Lindsey Richardson. 2020. Socioeconomic marginalization and opioid-related overdose: a systematic review. *Drug and alcohol dependence* 214 (2020), 108127.

[70] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[71] Lu Wang, Munif Ishad Mujib, Jake Williams, George Demiris, and Jina Huh-Yoo. 2021. An evaluation of generative pre-training model-based therapy chatbot for caregivers. *arXiv preprint arXiv:2107.13115* (2021).

[72] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, et al. 2022. Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 214–229.

[73] Joel Wester, Henning Pohl, Simo Hosio, and Niels van Berkel. 2024. " This Chatbot Would Never...": Perceived Moral Agency of Mental Health Chatbots. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–28.

[74] Joel Wester, Tim Schrills, Henning Pohl, and Niels van Berkel. 2024. “As an AI language model, I cannot”: Investigating LLM Denials of User Requests. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.

[75] Lauren Wilcox, Robin Brewer, and Fernando Diaz. 2023. AI Consent Futures: A Case Study on Voice Data Collection with Clinicians. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–30.

[76] Langdon Winner. 1980. Do artifacts have politics? *Daedalus* 109, 1 (1980).

[77] Michael Wornow, Yizhe Xu, Rahul Thapa, Birju Patel, Ethan Steinberg, Scott Fleming, Michael A Pfeffer, Jason Fries, and Nigam H Shah. 2023. The shaky foundations of large language models and foundation models for electronic health records. *npj Digital Medicine* 6, 1 (2023), 135.

[78] Ayae Yamamoto, Jack Needleman, Lillian Gelberg, Gerald Kominski, Steven Shoptaw, and Yusuke Tsugawa. 2019. Association between homelessness and opioid overdose and opioid-related hospital admissions/emergency department visits. *Social science & medicine* 242 (2019), 112585.

[79] Kai Zhang, Rong Zhou, Eashan Adhikarla, Zhiling Yan, Yixin Liu, Jun Yu, Zhengliang Liu, Xun Chen, Brian D Davison, Hui Ren, et al. 2024. A generalist vision–language foundation model for diverse biomedical tasks. *Nature Medicine* (2024), 1–13.

[80] Xueying Zhang and Lu Tang. 2021. Cultural adaptation in HPV vaccine intervention among racial and ethical minority population: a systematic literature review. *Health Education Research* 36, 5 (2021), 479–493.

[81] Zhiping Zhang, Michelle Jia, Hao-Ping Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. “It’s a Fair Game”, or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–26.

[82] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI ’23)*. Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3544548.3581318

[83] Gregory D Zimet, Zeev Rosberger, William A Fisher, Samara Perez, and Nathan W Stupiansky. 2013. Beliefs, behaviors and HPV vaccine: correcting the myths and the misinformation. *Preventive medicine* 57, 5 (2013), 414–418.