

# 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

**This is an unpublished pre-print. Please check the authors' websites for updated versions and publication information.**

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 diverse 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 public health professionals and individuals with lived experience to unpack their concerns, situated across three distinct and critical public health issues that demand high-quality information: infectious disease prevention (vaccines), chronic and well-being care (opioid use disorder), and community health and safety (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 offer example reflection questions to help practitioners adopt a risk-reflexive approach. We discuss the need to revisit and adjust pre-existing mental models of information-seeking and complement evaluations with external validity and domain expertise through lived experience and real-world practices. Together, this work contributes 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.

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. 2025. A Risk Taxonomy and Reflection Tool for LLM Adoption in Public Health. 1, 1 (June 2025), 30 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

---

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

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2025 Association for Computing Machinery.

Manuscript submitted to ACM

Manuscript submitted to ACM

1

## 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, 84], LLMs hold the potential to influence how we interact with information. LLMs have quickly gathered hundreds of millions of active users [76] 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.”<sup>1</sup>

This technology enthusiasm extends into public health, a public sector that aims to protect and improve the health of people and their communities by promoting healthy lifestyles, researching disease and injury prevention, and detecting, preventing and responding to infectious diseases [24, 82]. Studies have presented evidence of LLMs’ potential in assisting collection, analysis, and dissemination of public health information [45, 46, 93], developing public health policies [22, 30, 36], scaling care to population [4, 42], and supporting community health workers [75]. In practice, LLMs have already been implemented in community- or population-level systems or technologies, such as health record infrastructure [51], virtual agents for dealing with domestic violence [59], and public health monitoring instrument [42]. While this proliferation seems promising, experts and scholars caution against exaggerating LLM capabilities [20, 92]. Specifically, concerns have been raised about the potential of generating harmful content [28, 55, 99], perpetuating biases [69], compromising data privacy [97], limited generalizability of models [92], and challenges in real-world implementation [42, 89]. Given the high-stakes nature and data sensitivity of population-level health, it warrants careful consideration of when it is appropriate to employ LLM capacities and how we can mitigate potential harm when we do.

Yet, despite this recognized need for caution, evaluating the risks of LLM adoption in public health real practice remains a complex challenge. The first gap lies in the lack of domain understanding, as the majority of risk taxonomies lack granularity to be applied for specific uses [56] or are created within the computer science community and tend to leave out domain experts and real users [6, 80, 88]. 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 [77]. Identifying the presence of generated harmful content is only the first step; we also need to contextualize the documented possibility of low-quality information within specific public health issues and relevant populations to evaluate the consequences 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 public health professionals and individuals with lived experience who actively seek online health information. Specifically, we ask: **What negative consequences might arise from adopting LLMs to meet informational needs in public health?**

We conducted focus group sessions with ten public health professionals and ten individuals with lived experience to uncover potential negative influences of using LLMs as a communication tool or information source in public health. We selected vaccines, opioid use disorder, and intimate partner violence 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

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

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 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, including those with professional experience and lived experience. To our knowledge, this is the first work to comprehensively explore the potential risks of LLMs for public health. Second, by offering a shared vocabulary through a risk taxonomy and reflection questions for operationalization, this work paves the path for future collaborations between experts in computing and public health fields. As public interest gathers in LLMs as an emerging technology, the risk taxonomy and reflection questions allow for careful consideration of potential negative consequences and prevent reckless adoption that could result in unbearable disruptions to individuals and communities and further erode trust in public health responses.

**Content Warning:** We caution 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 Digital Support and Threats to Health Information Seeking

The integration of information and communication technology has transformed how care seekers and providers seek, access, and interact with health information. About 59% of adults in the U.S. have searched for health or medical information online in the past 12 months [87]. Notable technology examples include search engines, social media platforms and online communities, as well as applications for health education and intervention [65]. Various types of digital platforms democratize health information access [15, 98], provide safe spaces for individuals facing stigma, legal concerns, or social judgment [16, 43, 98], and connect people with individuals who share similar experiences [3, 37].

Despite the easy access and benefits, finding high-quality health information and protecting privacy remains a challenge. For example, Hansen et al. found that only 69% of observed searches in their study resulted in a correct and useful answer [32]. Online users, especially younger generations, do not always verify sources or assess information reliability [32, 34], while echo chambers can further amplify biased perspectives [39]. Online platforms can facilitate the spread of inaccurate health information, whether shared innocently or with deceptive intent, and the prevalence of low-quality information can gradually erode trust and discourage online health-seeking behaviors [3, 98]. Privacy concerns also emerge. While online anonymity allows individuals to seek sensitive health information [1], it also poses risks of individuals being identified when their data is aggregated across multiple online sources [54].

The nature of digital interactions presents additional challenges in further marginalizing, isolating, or even silencing individuals. First, disparities in digital access and literacy can contribute to marginalization. Individuals with lower health literacy face greater challenges in finding, evaluating, and understanding online health information [58]. Internet access is also unevenly distributed, with fewer online health information seekers in developing regions compared to developed ones [38]. Meanwhile, online platforms, despite being a “safe space” to some, can lack the interpersonal trust-building processes inherent in traditional healthcare interactions [43]. In some cases, technology can even be misused to silence marginalized voices rather than empower them [25].

These challenges become even more pronounced with the rise of generative AI that can rapidly produce human-like content at scale with inconsistent quality [14, 55, 99] and potential influences on people’s critical thinking [52]. While prior work has examined various digital threats to health information, there remains a gap in capturing changes brought by AI’s generative abilities and comprehensively understanding the potential negative consequences of LLMs in this area. Our study builds on prior scholarship and addresses this gap by situating LLMs within representative public health issues, positioning them alongside traditional health communication, and engaging individuals outside the computing field who bring professional knowledge or lived experience.

## 2.2 Potential and Challenges of LLMs in Public Health

A growing body of work has explored the potential of large language models in supporting the three core goals of public health agencies: assessment, policy development, and assurance [67, 82]. In assisting systematic collection, analysis, and dissemination of health information, LLMs can help assisting infectious disease surveillance [93], promoting health interventions and healthy lifestyles [46], and combating misinformation or myths in facing public health crisis [45, 63]. In supporting the development of public health policies, LLMs can be used to understand public health discourse [22], while generative agents can also simulate human behavior to inform public health policy [36]. In assuring delivering personal and community-wide health services to every member of the community, LLMs hold the potential to scale provider to populations [4, 42], monitor social determinants of health (SDoH) at the population level [30], as well as support community health workers [75]. In practice, LLMs have already been implemented in community- or population-level systems or technologies, such as health record infrastructure [51], virtual agents for dealing with domestic violence [59], and public health monitoring instrument [42].

Orthogonally, research has also highlighted challenges for adopting LLMs for informational needs and public health support. When providing informational assistance, LLMs face documented issues in data privacy [97], language and geographic disparities [41, 64], and creation of inaccurate, biased, or toxic content [99]. These limitations could lead to exaggerated or new risks in public health, a field that often involves both sensitive or high-stakes issues and serves diverse populations [82]. In high-stakes contexts, there remain questions about LLMs’ ability to provide reliable and effective information. For instance, Harrer found that only half of the LLM-generated messages met the standards for direct adoption for use [33]. When used for communication assistance, LLMs are found to have difficulty in being customized or controlled [89] and lack of sensitivity in managing emotional needs and emergencies [86]. Tailoring language to suit diverse demographics is another challenge [42], particularly for marginalized groups, such as the LGBTQ+ community [57, 70] and people with disabilities [27], where cultural nuances and sensitive topics may be overlooked. In addition, a core goal of public health is to reduce health disparities and deliver interventions at the population level [82]. However, the tendency of LLMs to propagate inaccurate and biased race-based suggestions [69] may actually hinder broader public health efforts.

Accordingly, scholarship has started to critically reflect on the way we approach LLM evaluation – an effort of significant importance to public health that is grounded in real practice and community efforts. Researchers have called for better reporting transparency with contextualized and empirically grounded evidence and claims [20], assessment reliability that accounts for practical deployment and data representation [92], diverse assessment methodologies and raters in surfacing health equity-related biases [72], and attention to patient-provider dynamics in practice [90]. For instance, Wornow et al. [92] found in their survey study that the generalizability of most clinical foundational models was limited by an overreliance on structured codes and non-representative datasets. At the same time, most evaluations do not offer meaningful insights for real-world applications, instead focusing on traditional natural language processing

(NLP) tasks [92]. As a result, they advocate 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 issues.

### 2.3 AI and LLMs Risk Taxonomies

In response to the growing literature on the ethics and limitations of AI and LLMs, researchers have proposed various taxonomies to address risks posed by these models. Some specifically examine the potential risks of LLMs but take a general perspective on their impacts. One of the earliest efforts is the position paper by Bender et al. [6] that highlights the dangers in environmental and financial costs, overrepresentation of hegemonic viewpoints and encoded biases, hype or misunderstanding in language model capabilities, misconception of synthetic text as meaningful. In a more structured approach, Weidinger et al. [88] summarized six risks areas based on computing expert discussion and supporting literature, which include discrimination, hate speech and exclusion, information hazards, misinformation harms, malicious uses, human-computer interaction harms, and environmental and socioeconomic harms. Other studies adopt a more focused view on specific areas of impacts, but not specifically limited to LLMs. For instance, Shelby et al. [80] conducted a scoping review on sociotechnical harm of algorithmic systems and identified five major themes: representational, allocative, quality-of-service, interpersonal harms, and social system/societal harms. Similarly, Domínguez Hernández et al. [19] conducted a review using snowball and structured search to study the social, political, and environmental impacts of foundation models, which leads to 14 categories of risks — including less-studied ones of unreliable performance, lock-in and opacity risks, and value misalignment. In the realm of health, De Choudhury et al. [17] 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. Hamid and Brohi [31] reviewed prior literature and summarized seven open research challenges: misinformation, resource limitations, bias and fairness in healthcare LLMs, interpretability and transparency, integration within clinical workflows, ethical considerations and patient privacy, and clinical validation and real-world performance.

Despite these contributions, there is no established vocabulary to specifically map the risks in public health contexts, a domain that is featured by its high-stakes nature and population-level impacts [82]. Additionally, public health encompasses a broad scope for care-seeking and -providing that extends beyond health encounters, involving a wide range of stakeholders such as social workers, researchers, and community planners. However, the expertise built through professional knowledge and lived experience remains largely missing from current discussions, despite public health's emphasis on real-world practice and field involvement [24, 82]. 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 relying solely on NLP metrics and expertise [72, 92]. Consequently, research in HCI and CSCW has called for human-centered approaches and stakeholder involvement, as certain harms may be difficult to capture solely through existing AI literature [2, 48, 90]. A recent effort in health is the work by Antoniak et al. [2], which involved diverse stakeholders including healthcare workers and birthing people to propose practical principles for ethical use of NLP and LLMs for maternal health. Our work seeks to address this gap by proposing a shared vocabulary grounded in public health issues, informed by perspectives of both public health professionals and individuals with lived experience.

### 3 Method

#### 3.1 Public Health Topic Selection and Context

To contextualize and evaluate the risks of using large language models (LLMs) for informational needs in public health, we chose 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 [49] – infectious disease prevention, chronic and well-being care, and community health and safety. Each topic underscores the importance of high-quality information in public health communication, which often involves both sensitive or high-stakes issues and diverse populations [82]. Specifically, vaccines play a crucial role in infectious disease prevention, while misinformation and public distrust significantly contribute to vaccine hesitancy [18, 73, 74]. Effective counter-speech is necessary to combat these misconceptions and enhance trust [96, 100]. OUD is a major issue in chronic and well-being care, which is exacerbated by stigma [21, 68] and vulnerability [83, 95]. This health crisis needs public education to reduce biases 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 [62, 79]. IPV involves complex challenges, including personal safety, mental health, and legal and financial challenges [26, 78]. Providing high-quality, empathetic information is essential for supporting survivors and addressing their comprehensive needs.

#### 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<sup>2</sup> research platform where we launched three separate studies for each of 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

<sup>2</sup><https://www.prolific.com/>

Table 1. Participant information. Professionals (P1-10) are public health professionals who engage with community-based health intervention or prevention efforts and have created or shared information about selected topics. Professionals' years of experience in occupations are indicated in brackets. Experiencers (E1-10) are individuals who have experienced questions about selected 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

have 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 public health professionals (P1-10) and ten individuals who have experienced questions about selected topics (E1-10) participated in focus group sessions. P5 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 participants' experience and practices in meeting informational needs—either as an information seeker or creator. Specifically, we asked about how they 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.

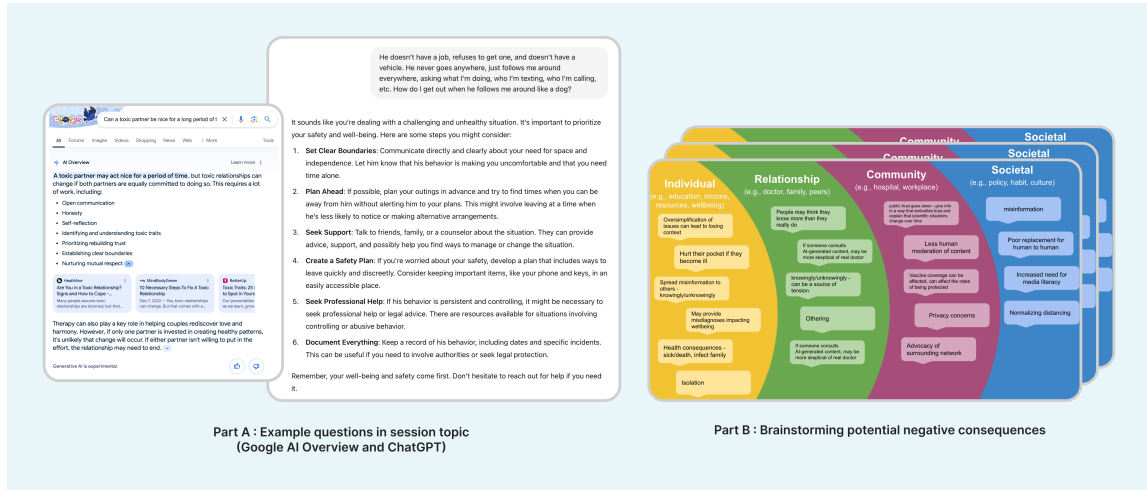


Fig. 1. Activities included in the focus group sessions: (A) Demonstrating LLMs' interaction mode and response quality through example questions on the session topic, (B) Brainstorming potential negative consequences of using LLM for public health.

Following the discussion of participants' current practices in meeting informational needs, 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 levels of prior experience with LLMs, we demonstrated their interaction mode and response quality by testing two example questions relevant to the session topic (Fig. 1 Part A) encouraging participants to try out their own questions as well. These questions were paraphrased from online communities of *r/abusive relationships*, *r/OpiatesRecovery*, and *r/HPV*<sup>3</sup> 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 (GPT-4o-mini<sup>4</sup>) and Google AI Overview (Gemini 1.5<sup>5</sup>) to show different interaction styles. We emphasized to the participants that the goal was not to compare these tools directly, as the questions were inherently different. During the sessions, we also encouraged participants to ask questions where participants asked questions such as the potential of “poisoning” models with synthetic extreme views.

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 [17, 23, 71, 81], 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.

<sup>3</sup>We used HPV vaccines as a specific example of vaccines to generate LLM responses.

<sup>4</sup>ChatGPT Release Notes. OpenAI. <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>

<sup>5</sup>While Google did not disclose specifics about model updates behind Google AI Overview, based on the time of focus groups and the Google Search team's announcement blog, it is likely to be Gemini 1.5. See: Expanding AI Overviews and introducing AI Mode. Google Search. Mar 05, 2025. <https://blog.google/products/search/ai-mode-search/>



### 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 that constitute the dimensions of our risk taxonomy, 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. 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 Risk Taxonomy for LLM Adoption in Public Health

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 4.1 to Sec 4.4, we unpack specific risks for each dimension. While a number of risks were identified, it is important to recognize that public health serves broad and diverse populations, and the emphasis on risks will naturally vary.

Overall, individuals with lived experience 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. Public health

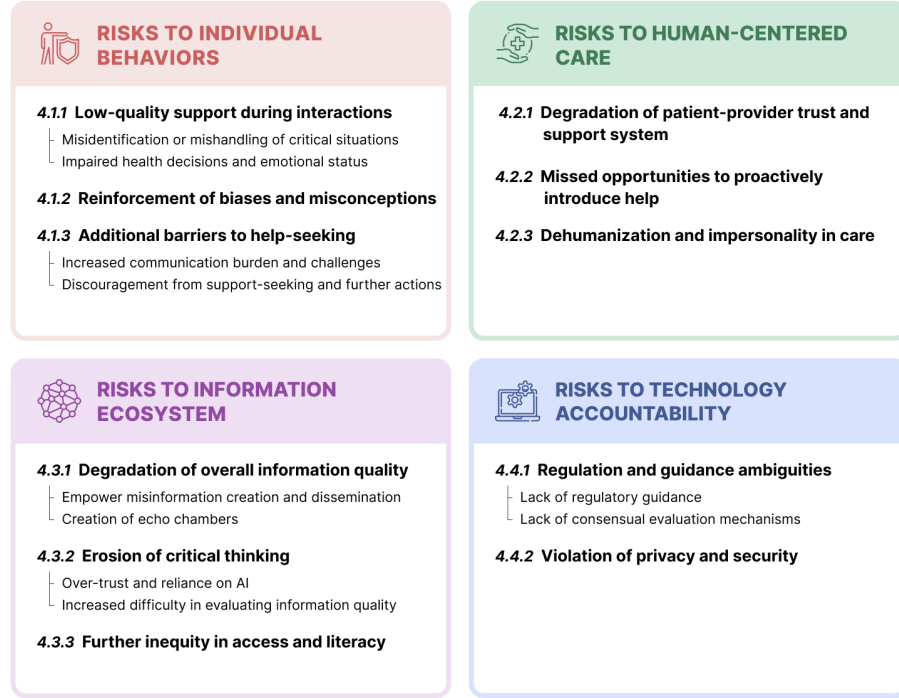


Fig. 2. We summarized four dimensions of risks that LLMs pose to public health: (1) Risks to Individual Behaviors, (2) Risks to Human-centered Care, (3) Risks to Information Ecosystem, (4) Risks to Technology Accountability.

professionals, on the other hand, tend to highlight more on 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.

#### 4.1 Dimension 1: Risks to Individual Behaviors

Participants reflected that individuals using LLMs to seek health information may experience harm during and after the interaction. They pointed out the potential of 1) low-quality support that mishandles critical situations or impairs health decisions and emotional well-being, 2) reinforcement of biases and misconceptions, or 3) additional barriers to help-seeking, such as communication burdens or discouragement from taking further actions.

##### 4.1.1 Low-quality support during interactions

**Misidentification or mishandling of critical situations:** Both professionals and general participants expressed worry that LLMs may fail to recognize and react appropriately to high-risk critical situations where immediate and effective responses are needed. These situations can be medical emergencies such as overdoses, or sensitive issues like

domestic violence. In the IPV sessions, we tested one rephrased question on Reddit, “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 professional and general participants reacted strongly to this answer as it failed to identify the potential severity that requires follow-up questions or at least to acknowledge the spectrum of toxic behaviors. Our participants explained that *“it doesn’t understand that this is an abusive situation (...) 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).

Our professional participants added that even if a specific case was recognized as critical, it was equally challenging to determine whether escalation was needed or how to properly approach it. 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. Our participants believe that 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 and professional support (e.g., in cases of suicidal risk): *“(in some situations) someone should be redirected to like a national line or someone else to talk to”* (P5).

**Impaired health decisions and emotional status:** Our participants voiced concern that people’s physical health could be jeopardized by misguided health decisions, while some public health professionals worried preventable issues would end up with costly or serious outcomes. They pointed out that guidance on OUD from LLMs could be dangerously inadequate, potentially resulting in severe repercussions of overdose or death. Participants also discussed the risk that people’s emotional status could also be negatively affected if LLMs caused 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: *“if I say ‘hey, what causes chest pains’, it’s gonna give you a list of things that could possibly be causing chest pains, and with health anxiety, you’re gonna go right to worst case scenario, and (...) be in a state of panic, and nothing can kind of talk you down from that.”* When it comes to mental health-related issues, our professional participants emphasized that it is nearly impossible to avoid triggering individuals with unknown backgrounds and histories, despite the considerable effort and training they invest in understanding the communities they serve and the accumulated experience through interactions — let alone for LLMs that lack all of these critical lived experience.

#### 4.1.2 Reinforcement of biases and misconceptions

Participants were apprehensive that misleading or biased narratives could be amplified and even internalized by individuals, especially if LLMs 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, some public health professionals worried that 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, our participants highlighted that both victim blaming and racial discrimination could prohibit people from seeking help: *“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.”* (P2). 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.”*

Our public health professional participants also stressed the need for culturally sensitive communication in mitigating biases and misconceptions; without it, counterspeech may trigger backfire effects. These professionals emphasized that all types of individual nuances demand efforts ranging from language adaptation to community understanding. One

example mentioned by P1 (IPV) is that certain religions may hold a belief that divorce is wrong and women should tolerate abuse, which requires additional care to sensitively address biases without turning help-seekers away. Public health professionals highlighted that, in practice, they need special training or 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. P5 (OUD), a public health director, explained this need of communication strategy: *“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.”*

#### 4.1.3 Additional barriers to help-seeking

**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, with P1 (IPV) 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.”*

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.”*

**Discouragement from support-seeking and further actions:** Participants feared that help-seekers may experience stronger reluctance to seek assistance and tendencies toward self-isolation, while LLMs may outright deny assistance or provide unactionable or unempathetic guidance. They explained that if individuals in difficult situations already withdraw from social relationships and LLMs offer human-like interactions as a seemingly “easier” alternative, 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. IPV experiencers worried 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, participants also worried that when LLMs provide oversimplified suggestions, on top of the low-quality support during the interaction itself, it also has further effects of leaving 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.”*

## 4.2 Dimension 2: Risks to Human-centered Care

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.

### 4.2.1 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. They worried that if people start to take LLM responses as absolute facts, these people 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 suggestions 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.”*

Many participants raised concerns that people may start to rely on flawed technical support and choose not to consult with doctors or go to hospitals. Even more troubling, participants feared users could become *“anchored”* to technology instead of seeking out or building social support systems: *“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.”*

### 4.2.2 Missed opportunities to proactively introduce help

Our professional participants doubted if LLMs can recognize and adjust to 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 would 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 about their life 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, this proactive role is diminished, and individuals are less likely to receive help unless they actively and intentionally seek it.

### 4.2.3 Dehumanization and impersonality in care

Our public health professional participants were troubled by the possibility that professionals may unintentionally adopt similar impersonal and pragmatic communication styles of LLMs if they become accustomed to the language style. 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 this adoption can hurt the empathy and trust building with patients, which 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 based on the during-session demos and their own experience, 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. As an example, 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.”*

### 4.3 Dimension 3: Risks to Information Ecosystem

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.

#### 4.3.1 Degradation of overall information quality

**Empower misinformation creation and dissemination:** Participants fretted that with LLMs, misinformation can gain traction with high volume and presented credibility, allowing malicious individuals to distort words to their advantage. E1 (IPV) 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.”* Our public health professional participants also worried the lagged media literacy could exacerbate the damage: *“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.”* (P9, VAC).

**Creation of echo chambers:** Some participants imagined that echo chambers created by LLMs could perpetuate misconceptions, especially when users do not know how to ask the right and non-leading questions. E5 (OUD) said *“I think a lot of times they (users) ask the wrong questions or [...] lead it on to give a specific answer that they’re not necessarily looking for.”* In the OUD sessions with insiders, we tested a question about long-term opioid use for chronic pain and LLM emphasized benefits like enhanced physical function and improved quality of life but downplayed risks. A domain expert noted that was perhaps the most positive tone she had ever encountered in any informational materials on the subject. Notably, none of the experienter participants raised concerns, which we believe indicates a gap in recognizing the potential harm that such overly positive portrayals may cause.

#### 4.3.2 Erosion of critical thinking

**Over-trust and over-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, potentially leading them to believe they fully grasp a topic or issue. P6 (OUD) explained 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, as 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 them. As a matter of fact, 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.”*

**Increased difficulty in evaluating information quality:** Our participants also shared that evaluating the quality of LLM responses is not easy when their previous ways of verifying the author and reference no longer apply. 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.

#### 4.3.3 Further inequity in access and literacy

Our professional participants were concerned that inequities in the information ecosystem can be exacerbated, considering disparities in language proficiency and technology access. 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 tried to use LLMs to adjust certain health information for native Spanish speakers with 4th-grade



reading levels but was unable to generate messages that would work in practice. 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.”*

#### 4.4 Dimension 4: Risks to Technology Accountability

Participants pointed out the uncertainties and concerns about technology accountability, highlighting 1) ambiguities in regulation and guidance due to lack of policies and understanding, and 2) violations of privacy and security.

##### 4.4.1 Regulation and guidance ambiguities

**Lack of regulatory guidance:** Our participants expressed concerns about the gray area in determining responsibility when LLMs create low-quality content and poor outcomes. Our general user participants worried that LLMs could be misused for malicious purposes to justify harmful behaviors or beliefs. On the other hand, public health professionals believed that the lack of regulatory guidance on LLM use would prohibit 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 major 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?”*

**Lack of consensual evaluation mechanisms:** Our public health professional 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.”*

##### 4.4.2 Violation of privacy and security

Our participants raised another concern 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). They 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) shared her concern with any AI being: *“Where is this information going? Where is it being stored? Is any of this being resold?”* Even if unintentionally, sensitive details that individuals thought they shared in a safe space can still be leaked: *“even a data breach could happen (...) the developers have access to all the questions you ask (...) then that (personal information) would not be secure”* (E5, OUD).



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. Participants highlighted that these safety precautions and privacy design are commonly seen in practice and specialized technologies (as for violence prevention), but are missing in a tool that is promoted as serving all purposes.

## 5 Operationalizing the Risk Taxonomy

To facilitate the operationalization of this risk taxonomy, in this section, we present (1) a listing of distinctive LLM characteristics from traditional health information sources that contribute to the above-identified risks, which grounds the risks into the technical affordances and features; and (2) a set of reflection questions that contextualize each risk into discussion probes for practical utilization. These two were drawn from our engagement and interpretations of participant input and data analysis. As we envision the risk taxonomy as a bridge to help both AI and health practitioners chart potential consequences in adopting LLMs in public health, these tools can support different stakeholders in understanding the technology's capabilities and limitations, reflecting the web of vulnerabilities in specific public health contexts, and probing benefit-risk tradeoff with a risk-reflexive mindset.

### 5.1 Distinctive LLM Characteristics

To ground the risk taxonomy in the technical affordances and features of contemporary LLMs, we summarize a list of characteristics (Fig. 3) that contribute to the identified risks and differentiate LLMs from traditional health information sources. We believe this list can serve as basic introductory material for the general public to understand the mechanisms and constraints of the technology, as well as for interdisciplinary stakeholders with varying levels of AI knowledge to support informed discussions about what risks may emerge and how they might be mitigated. 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. We acknowledge that some characteristics may not apply to all models and product deployments — such as for fine-tuned or specialized models and emerging interaction modes — and mark these characteristics with asterisks (\*).

- **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.
- **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).
- **Positioned to serve general purposes\*:** Most LLMs claim to serve general purposes, but they tend to lack domain-specific knowledge and practice-based expertise.

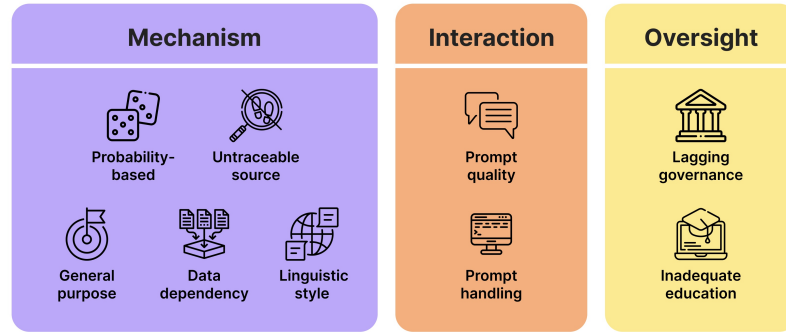


Fig. 3. Distinctive LLM characteristics from traditional health information sources that contribute to identified risks.

- **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.
- **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.

## 5.2 Reflection Questions

To support the practical application of this risk taxonomy in light of the LLM characteristics presented above, we offer a set of reflection questions. These reflection questions demonstrate how the taxonomy can be translated into an actionable evaluation tool: they can serve as practical probes to evaluate specific benefit-risk tradeoffs in a hypothetical LLM deployment scenario, by reflecting on whether the LLM's risks outweigh its potential benefits and where further mitigation may be needed. For each risk, we have noted the relevant LLM characteristics that could contribute to it, marked with icons in accordance with Fig. 3; however, we acknowledge that these relationships are not definitive or fixed, as they may vary depending on specific contexts and implementations. We also emphasize that these questions are not exhaustive; rather, they are illustrative prompts intended to support discussions. Appropriate questions will vary depending on the specific use cases and populations served.

### 5.2.1 Risks to Individual Behaviors

#### Low-quality support during interactions

LLMs for public health need to be prepared to handle both critical situations and the risks of low-quality information. Critical events such as overdoses or domestic violence require timely identification, appropriate response, and escalation or redirection to professional support. Equally important is the need to prevent harm from low-quality responses, which

can jeopardize users' physical health or emotional well-being. Since LLMs are probability-based, often positioned as general-purpose tools, and highly dependent on users' prompts, their use in public health must be guided by professional standards through aligning with established guidelines for identifying and managing high-risk cases and actively monitoring the credibility of generation information.

*To better identify or handle critical situations:*

- *What are the best practices or guidelines for health professionals in identifying and handling critical situations?*
- *How can technical experts translate best practices into technologies to identify and escalate or redirect situations that require intervention or professional help?*
- *What are the relative risks of under-identification and over-identification of critical situations?*

*To better support health decisions and emotional status:*

- *How may users' health decisions and emotional status be hurt if the tool produces low-quality information? 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?*

### Reinforcement of biases and misconceptions

LLMs for public health need to consider potential biases or misconceptions, as public health topics are often sensitive and complex, and mistakes or non-cultural sensitive communication can backfire. Since LLMs are trained on broad and imperfect datasets, they may unintentionally reflect prevailing stereotypes or follow users' inclinations. Culturally tailored communication is needed in facing diverse populations. Yet, LLMs tend to produce standardized language and lack the contextual awareness needed, while there is limited governance and insufficient mechanisms for oversight. Preventing these harms requires intentional design choices, alignment with community-informed practices, and ongoing efforts to ensure outputs are accurate, respectful, and attuned to the needs of those most affected.

- *What are the common biases and misconceptions regarding the intended use cases? How to prevent the output information from reinforcing them?*
- *What are the potential cultural or personal variances in approaching the intended use cases? How do health professionals adjust accordingly to these differences?*
- *How can technical experts incorporate a feedback loop to continuously reflect on and improve support quality?*

### Additional barriers to help-seeking

LLMs may create further barriers to help-seeking, particularly for those who withdraw from social interactions or are hesitant to take action. The ease and immediacy of LLM can make them appealing substitutes to real social interactions, and discourage people from reaching out to real-world support networks or professionals. When LLMs offer vague or oversimplified suggestions, deny certain requests, or lack bidirectional conversations, users may feel dismissed or stalled. These shortcomings are tied to LLM's inherent qualities: sources are untraceable even if not hallucinated, non-specialized models have challenges of providing specialized knowledge or local resources, and cannot engage in proactive dialogue. Therefore, LLMs for public health should support users in forming effective questions, provide actionable and digestible information, and scaffold pathways toward additional help.

*To lower the communication burden:*

- *How do health professionals find their role in communication when serving specific goals? When do health professionals need to lead the conversations? In these situations, what are the recommended communication strategies in professional practice?*
- *How can technical experts translate professional practice into rules, and support users in formulating questions that accurately and fully describe their needs and situations?*

*To facilitate support-seeking and further actions:*

- *When and how do health professionals scaffold users to seek real-world help or take necessary actions?*
- *What requests are beyond the tool's capabilities? In these cases, how can technical experts communicate our limitations and refer users to relevant and accessible resources?*
- *Do health professionals believe that users can act on the information provided by the tool?*

### 5.2.2 Risks to Human-centered Care

#### Degradation of patient-provider trust and support system

LLMs could undermine trust in health professionals, especially when their advice conflicts with AI responses or when it's unclear whether recommendations come from expertise or technology. As LLMs appear authoritative and human-like, they may be misused as replacements for support systems or even discourage long-term investment in human-centered services. These risks are amplified by LLMs' untraceable sources, limited domain understanding, and the lack of sufficient public education and governance. Therefore, LLMs in public health need to clearly communicate their limitations, support disclosure practices when used alongside human care, and proactively redirect users toward qualified providers and supportive systems when needed.

- *Compared to health professionals, what are the tool's limitations? When and how to clarify the difference between technology mechanisms and professional expertise?*
- *What is this tool's role alongside actual care and health professionals? If it will support health professionals, how can they effectively explain the tool's assistance to patients?*
- *What are the potential long-term impacts if users rely solely on this tool? What relevant community or professional support can be provided to users?*

#### Missed opportunities to proactively introduce help

LLMs may struggle to recognize when individuals are ready for support or how to respond with appropriate timing. In practice, public health professionals tend to take a proactive role in introducing help when it is most likely to be accepted, by assessing a person's mental state, life circumstances, and readiness for intervention. However, LLMs, particularly general-purpose ones, depend on user initiation and prompt quality, and they lack continuous interactions or embedded real-world practices. To better support users in vulnerable moments, LLMs should be guided by domain-specific practices that scaffold help-seeking and provide proactive and context-sensitive help.

- *How do health professionals identify a person's readiness for support and how might those insights inform system design? What are the appropriate times and situations to introduce help?*
- *When do health professionals proactively reach out to help-seekers, and how can systems incorporate similar check-in or follow-up conversations?*

#### Dehumanization and impersonality in care

When used in public health communication, LLM outputs may unintentionally set a pragmatic and impersonal tone. If users or providers grow accustomed to this style, it can undermine the emotional care and empathy that is essential to trust-building in health encounters. LLMs are probability-based and thus lack true empathy or bi-directional engagement,

and are not yet governed by clear standards for public sector use. Thus, their neutral or generic responses, especially when addressing sensitive issues, can fail to affirm people's feelings or situations and discourage them from recognizing harm or seeking help. Therefore, LLMs should consider the emotional needs of users and create feedback mechanisms that evaluate human experience.

- *What are the potential emotional needs of users? How do health professionals address these needs and assess empathy and compassion in communications?*
- *Are there recommended training or guidelines from corresponding domain experts on approaching communication?*
- *How can technical experts create a feedback loop for the system and continuously evaluate the emotional quality of information?*

### 5.2.3 Risks to Information Ecosystem

#### Degradation of overall information quality

LLMs can decrease the overall quality of public health information due to their capacity to generate large volumes of convincing content without understanding the information to guarantee accuracy. Without proper oversight and public education, their accessibility may enable falsehoods to appear credible or widely accepted. Because of LLMs' sycophancy tendency, they risk reinforcing beliefs, limiting exposure to alternative perspectives, and creating informational echo chambers. Therefore, LLMs should clearly communicate their limitations, support content moderation, and align with evidence-based practices.

*To mitigate misinformation creation and dissemination:*

- *How can technical experts communicate risks related to AI-generated misinformation (such as hallucination) to health professionals and end users?*
- *How do health professionals and previous systems monitor and mitigate the spread of misinformation? What mechanisms (e.g., fact-checking, cross-referencing) can be built into the design to detect, correct, or prevent harmful content?*

*To counter 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:

LLMs may erode critical thinking abilities with responses that appear organized, confident, and authoritative, which can give people a false sense of mastery over a topic. This standardized language style can also be mistaken for accuracy and completeness. In traditional information environments, people often rely on cues such as authorship, references, or domain reputation to assess credibility. However, these credibility indicators no longer apply to LLMs, making it difficult for users to evaluate the quality. Therefore, LLMs should actively communicate their mechanism and limitations and incorporate means that encourage users to critically engage with the generated content.

*To limit over-trust and reliance on AI:*

- *How should technical experts and health professionals communicate this tool's limitations and prevent users from mistaking efficiency for correctness and comprehensiveness?*
- *Could the tool unintentionally create an impression of authority or reinforce an illusion of knowledge? What design cues or transparency measures can be implemented to mitigate this risk and promote users' critical engagement with the information?*

*To lower the difficulty in evaluating information quality:*

- *How transparent are model developers about the tool's training data sources and limitations?*
- *How can system designers and public educators explain the uncertainties in LLM generations, persuasiveness in standardized language, and mechanism differences from other information sources?*
- *What external resources or user reminders could be integrated into the tool to assist users in verifying information?*

### Further inequity in access and literacy

LLMs could exacerbate inequities in access and literacy, such as for individuals with lower digital and health literacy, non-English speakers, or people without reliable access to the internet or devices. The standardized language style and lack of adaptation to different literacy levels may further limit understanding or usability. Therefore, LLMs for public health should consider inclusive language practices and be accompanied by alternative information sources and equitable policies to ensure accessibility across populations.

- *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?*
- *What are the implications for individuals who may need this tool but lack access to it?*
- *How can designers and developers support users with limited reading or digital literacy in interacting with this tool? How would health professionals customize the presented information in considering individual differences?*

### 5.2.4 Risks to Technology Accountability

#### Regulation and guidance ambiguities

LLMs for public health remain in a regulatory gray area, where responsibility for harm caused by low-quality outputs is still ambiguous. In the absence of shared standards and established practices, the application of LLMs in public health should proactively manage risks by setting clear red flags and safeguards, while incorporating user feedback and domain expertise to refine evaluation processes.

*To cope with lagging regulatory guidance:*

- *What current regulations apply to the use of AI tools in the intended area? Are there any risky use requests that should be referred to professionals?*
- *How could this tool be misused? What safeguards can designers and developers implement to prevent or limit such misuse?*
- *How could the lack of regulatory guidance affect the evaluation and management of responsibilities and risks?*

*To navigate the lack of consensual evaluation mechanisms:*

- *What existing evaluation mechanisms and practices in health can be applied to assess the effectiveness and reliability of this tool? Can technical experts translate these practices into tests or benchmarks?*
- *How can health professionals and technical experts collaboratively gather user feedback to evaluate this tool, and how can that feedback be integrated to update evaluation practices and refine safeguards?*

#### Violation of privacy and security

LLMs may lead to privacy and security risks, particularly when used in sensitive public health contexts. Without built-in safeguards, users (both patients and professionals) may unknowingly share personal or protected health information. Unlike specialized technologies governed by strict privacy standards, general-purpose LLMs often lack the privacy-aware design required in healthcare. Given their probabilistic nature and opaque data sourcing, these LLM systems may accidentally expose or connect sensitive data. Therefore, LLMs for public health should incorporate responsible data collection and management practices, while guiding users in safe and privacy-conscious practices.

- *What types of information does this tool collect, and how is that information stored, used, and managed? What measures should technical experts implement to protect users' privacy?*
- *What are the applicable privacy regulations, and how can technical and law experts ensure compliance with them? What procedures are in place for handling privacy violations or potential data breaches?*
- *How can designs and communications convey privacy risks and users' ability to manage their data? What guidance should be provided about the types of information users should or shouldn't share? Could users unknowingly disclose sensitive information, and how can this risk be minimized?*

## 6 Discussion

### 6.1 LLMs for Informational Needs: Revisiting and Adjusting Pre-existing Mental Models of Information-Seeking Approaches

One of the highly discussed LLM application domains is to integrate them into informational tools – as has been seen in the growing framing of “next-generation search engine” that has gained traction since the beginning of LLM breakthroughs and associated hype [29, 61, 66]. We have seen this direction already being put into practice; sometimes in clearly identifiable ways, such as with new products like Perplexity, and other times more subtly, as with Google Search’s AI Overview feature embedded into search engines. 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 [56]. 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 [97]. 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 Google AI overview is powered by LLMs.

Regardless of whether being approved or recognized by the public or not, individuals are beginning to engage with LLMs or LLM-empowered tools for informational-seeking behaviors in ways that mirror their previous habits and experience with traditional informational sources [47, 94]. Yet, we are at a time where public communication and education around LLMs remain lagging and even limited, leaving many users to rely on pre-existing mental models – that are formed and based on approaches which are inherently different from LLMs – to make sense of and interact with this new form of technology. As a result, users may approach and conceptualize these tools with inaccurate expectations and misplaced trust.

In the online world, the most common approaches for health information used to be search engines and online communities [16, 65]. Search engines can retrieve information from numerous sources that can be ranked by relevance or promotion priority, and often provide clear attribution to websites or authors [53]. Online communities and social media platforms offer a more social experience where people can seek emotional support or long-term connections or find people with similar backgrounds or experiences [16, 44]. However, our findings reveal that LLMs do not fit perfectly into either framework, yet users may still mistakenly apply old assumptions to making sense of LLMs. Unlike search engines, LLMs typically do not offer transparent reference attribution or a list of resources for people to choose and select. Instead, they provide a synthesized and consolidated output where the tasks of selection, interpretation, summarization, and evaluation are collapsed into a single response, with very limited user agency and control if any.

Our participants also noted that because LLMs often present information in a list form, it was easy to assume an implied ranked importance, even though no such ranking is intended or meaningful in the context of LLM responses. At the same time, while the chatbot-like interface may resemble interactions in online communities, LLMs cannot have the lived experience and interpersonal reciprocity that characterize real-human social exchanges [13, 44]. In addition, verifying information quality can be additionally challenging in LLM interactions. LLM responses tend to sound authoritative and comprehensive and sometimes even make up non-existing sources and evidence [99], while credibility indicators such as authorship and language style [7, 35, 40, 50] no longer apply.

This potential mismatch between user expectations shaped by pre-existing mental models and actual LLM characteristics can lead to misunderstanding, mistrust, or even harmful outcomes when LLMs are used for sensitive or high-stakes informational needs as in public health. As such, it is essential to understand the mental models people may based on and actively support the adjustments of these mental models. Therefore, 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 to 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 – whether through interface cues, system disclosures, or public education – to empower users with a fair understanding of such tools for safe and appropriate utilization.

## 6.2 LLMs for Public Health and Other Public Sectors: Complementing Evaluations with External Validity and Domain Expertise

Public health is a domain grounded in accountability and population-scale responsibility, with values deeply rooted in ethical and practical obligations. As emphasized by the public health professionals in our study, introducing LLMs into this space raises serious questions about liability and reliability. These public health professionals expressed fears of becoming the ones bearing responsibility and losing their professional license if LLM-generated information leads to harm. They also raise broader concerns about how individual or incidental failures might result in cascading failures when deployed to large populations.

However, these professionals’ voices are largely missing in LLM design and development today. While some research has begun to involve clinicians and healthcare professionals, the participation of public health experts – who bring critical perspectives from community-level care and population health – remains limited. Despite the growing body of work on benchmarking LLMs and examining their impacts, most assessment approaches remained centered within the computer science field. These approaches often rely on metrics that access internal validity, such as the frequency of hallucinations or toxic language [14, 55], which fall short in capturing how such failures play out in real-world settings. In high-stakes public sector domains like public health, this limitation can be significant.

Our findings illustrate this point by revealing a range of risks that are previously underexplored in the literature but rooted in everyday practices in public health professionals and the lived experience of individuals. These overlooked risks primarily concern help-seeking and help-providing processes, as well as liability issues. This extension of help-seeking and -providing is related to the expansive scope of care in public health, which extends beyond health encounters and involves a wide range of stakeholders like community and social workers. Meanwhile, liability is a core concern for practical fields where professionals are accountable for harm — especially in public health, where practitioners



often work on the front lines and policies can affect populations. Specifically, the missing risks include additional communication burdens that may create new barriers to help-seeking, the discouragement of support-seeking and further actions due to impersonal or unhelpful responses, and the erosion of critical thinking — not simply due to over-reliance on AI, but because users are left without tools or context to evaluate information quality. There is also the potential of missed opportunities for LLMs to proactively introduce help, a role that public health professionals routinely play in practice. Moreover, the absence of consensual evaluation mechanisms for LLM performance leaves public health directors and frontline workers without shared standards for assessing appropriateness or safety.

Together our work underscores the need to complement existing evaluations with external validity and domain expertise that is built through lived experience or situated knowledge. Instead of viewing LLMs as a universal one-for-all solution, they should serve as the base for vertical adaptations [85] that can be contextualized and adapted for specific public health needs. This adaption will help align systems with the practices, values, and constraints of public health professionals and affected communities. In doing so, we join others in calling for participatory approaches with health professionals [2, 72, 90], to include also a wider range of public health professionals, and call for involvement of domain experts not as last-stage validators but as co-creators who will help shape the scopes and tasks of the technology. We also echo Winner’s views on reflecting technology change and their social influences [91], arguing that adoption and evaluation decisions should be approached in two parts. First: should LLMs be used at all for a given public health task? There may be scenarios where the stakes or relational complexity simply make them unsuitable [5]. Second: if LLMs are used, how should they be designed and evaluated to reduce harm and serve the communities? After all, we emphasize that LLMs are not standalone solutions; they should be integrated with other application layers and safety protocols to ensure responsible, context-sensitive use.

Importantly, our proposed risk taxonomy offers a starting point for addressing external validity and domain expertise. We view it not as a checklist for compliance, but a tool for proactive reflection. Combined with the LLM characteristic list and reflection questions, our work provides a practical reflection tool to support interdisciplinary experts, regardless of their level of LLM expertise, in conducting case-specific evaluations. Together, our work offers a shared vocabulary for researchers and practitioners in both computing and public health to identify who might be at risk, under what conditions, and how risks might be mitigated through design and deployment decisions. Many participants in our study emphasized 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 social and 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 issues and involved communities to know *what to evaluate*. We believe that a responsible and comprehensive risk evaluation requires insights from both perspectives, and we hope this work can support joint efforts in defining what to evaluate and how to evaluate, translating real-world practice into the assessment of emerging technologies.

### 6.3 Limitations and Future Work

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. Similarly, we do not claim the list of distinctive characteristics or the mappings between characteristics and risks as definitive, especially since the landscape of LLM capabilities is ever-emergent and evolving. 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. We acknowledge that our brief introduction to LLM mechanisms may not have provided sufficient technical background for participants to identify every possible risk. However, we intentionally recruited individuals with varying levels of LLM familiarity to reflect real-world conditions, where only 23% of U.S. adults had ever used ChatGPT in 2024 [60]. We believe that participants can still offer valuable insights into potential consequences in actual-use contexts drawing from their practice-based and lived expertise. Future work can explore risks by analyzing real-world LLM usage histories donated by users, or by inviting participants to engage with LLMs over time and examining their interaction data and reflections through follow-up interviews. 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.

## 7 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 hope 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] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 3906–3918.
- [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] Laima Augustaitis, Leland A Merrill, Kristi E Gamarel, and Oliver L Haimson. 2021. Online transgender health information seeking: facilitators, barriers, and future directions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [4] 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.
- [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] Mohit Chandra, Siddharth Sriraman, Gaurav Verma, Harneet Singh Khanuja, Jose Suarez Campayo, Zihang Li, Michael L Birnbaum, and Munmun De Choudhury. 2024. Lived Experience Not Found: LLMs Struggle to Align with Experts on Addressing Adverse Drug Reactions from Psychiatric Medication Use. *arXiv preprint arXiv:2410.19155* (2024).
- [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] Kaitlin L Costello. 2016. Impact of patient-provider communication on online health information behaviors in chronic illness. *Proceedings of the Association for Information Science and Technology* 53, 1 (2016), 1–10.
- [16] Munmun De Choudhury, Meredith Ringel Morris, and Ryan W White. 2014. Seeking and sharing health information online: comparing search engines and social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1365–1376.
- [17] 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).
- [18] 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.
- [19] 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 *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 776–796.
- [20] Joanneke Drogts, 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.
- [21] 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.
- [22] Laura Espinosa and Marcel Salathé. 2024. Use of large language models as a scalable approach to understanding public health discourse. *PLOS Digital Health* 3, 10 (2024), e0000631.
- [23] Centers for Disease Control, Prevention, et al. 2015. Models and frameworks for the practice of community engagement.
- [24] U.S. Centers for Disease Control and Prevention (CDC). 2025. Introduction to Public Health. <https://www.cdc.gov/training-publichealth101/php/training/introduction-to-public-health.html>
- [25] Jasmine C Foriest, Shrivika Mittal, Kirsten Bray, Anh-Ton Tran, and Munmun De Choudhury. 2024. A Cross Community Comparison of Muting in Conversations of Gendered Violence on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–29.
- [26] 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.
- [27] 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.
- [28] Samuel Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A Smith. 2020. Realtotoxicityprompts: Evaluating neural toxic degeneration in language models. *arXiv preprint arXiv:2009.11462* (2020).
- [29] Google. 2023. Supercharging Search with generative AI. <https://blog.google/products/search/generative-ai-search/>
- [30] Marco Guevara, Shan Chen, Spencer Thomas, Tafadzwa L Chaunzwa, Idalid Franco, Benjamin H Kann, Shalini Moningi, Jack M Qian, Madeleine Goldstein, Susan Harper, et al. 2024. Large language models to identify social determinants of health in electronic health records. *NPJ digital medicine* 7, 1 (2024), 6.
- [31] Rida Hamid and Sarfraz Brohi. 2024. A Review of Large Language Models in Healthcare: Taxonomy, Threats, Vulnerabilities, and Framework. *Big Data and Cognitive Computing* 8, 11 (2024), 161.
- [32] Derek L Hansen, Holly A Derry, Paul J Resnick, and Caroline R Richardson. 2003. Adolescents searching for health information on the Internet: an observational study. *Journal of medical Internet research* 5, 4 (2003), e900.
- [33] 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).
- [34] Amelia Hassoun, Ian Beacock, Sunny Consolvo, Beth Goldberg, Patrick Gage Kelley, and Daniel M Russell. 2023. Practicing information sensibility: How Gen Z engages with online information. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [35] 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.
- [36] Abe Bohan Hou, Hongru Du, Yichen Wang, Jingyu Zhang, Zixiao Wang, Paul Pu Liang, Daniel Khashabi, Lauren Gardner, and Tianxing He. 2025. Can A Society of Generative Agents Simulate Human Behavior and Inform Public Health Policy? A Case Study on Vaccine Hesitancy. *arXiv preprint arXiv:2503.09639* (2025).

- [37] Maia Jacobs, Galina Gheihman, Krzysztof Z Gajos, and Anoopum S Gupta. 2019. "I think we know more than our doctors" How Primary Caregivers Manage Care Teams with Limited Disease-related Expertise. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–22.
- [38] Xiaoyun Jia, Yan Pang, and Liangni Sally Liu. 2021. Online health information seeking behavior: a systematic review. In *Healthcare*, Vol. 9. MDPI, 1740.
- [39] Julie Jiang, Xiang Ren, Emilio Ferrara, et al. 2021. Social media polarization and echo chambers in the context of COVID-19: Case study. *JMIRx med* 2, 3 (2021), e29570.
- [40] 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>
- [41] 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*.
- [42] 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.
- [43] Grace J Johnson and Paul J Ambrose. 2006. Neo-tribes: The power and potential of online communities in health care. *Commun. ACM* 49, 1 (2006), 107–113.
- [44] Allen C Johnston, James L Worrell, Paul M Di Gangi, and Molly Wasko. 2013. Online health communities: an assessment of the influence of participation on patient empowerment outcomes. *Information Technology & People* 26, 2 (2013), 213–235.
- [45] Saubhagya Joshi, Eunbin Ha, Andee Amaya, Melissa Mendoza, Yonaira Rivera, and Vivek K Singh. 2024. Ensuring accuracy and equity in vaccination information from ChatGPT and CDC: mixed-methods cross-language evaluation. *JMIR Formative Research* 8 (2024), e60939.
- [46] 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.
- [47] Thashmee Karunaratne and Adenike Adesina. 2023. Is it the new Google: Impact of ChatGPT on students' information search habits. In *Proceedings of the 22nd European Conference on e-Learning ECEL*. 147–155.
- [48] 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.
- [49] Jeffrey P Koplan and David W Fleming. 2000. Current and future public health challenges. *Jama* 284, 13 (2000), 1696–1698.
- [50] Bill Kovach and Tom Rosenstiel. 2011. *Blur: How to know what's true in the age of information overload*. Bloomsbury Publishing USA.
- [51] 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>
- [52] Hao-Ping Hank Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
- [53] Dirk Lewandowski. 2023. *Understanding search engines*. Springer.
- [54] Fengjun Li, Xukai Zou, Peng Liu, and Jake Y Chen. 2011. New threats to health data privacy. In *BMC bioinformatics*, Vol. 12. Springer, 1–7.
- [55] 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).
- [56] 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.
- [57] 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.
- [58] Jennifer Manganello, Gena Gerstner, Kristen Pergolino, Yvonne Graham, Angela Falisi, and David Strogatz. 2017. The relationship of health literacy with use of digital technology for health information: implications for public health practice. *Journal of public health management and practice* 23, 4 (2017), 380–387.
- [59] 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>
- [60] Colleen McClain. 2024. Americans' use of chatgpt is ticking up, but few trust its election information. Pew Research Center.
- [61] Microsoft. 2023. Reinventing Search with a New AI-Powered Microsoft Bing and Edge: Your Copilot for the Web. <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/>
- [62] 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.
- [63] Shravika Mittal, Hayoung Jung, Mai ElSherief, Tanushree Mitra, and Munmun De Choudhury. 2025. Online Myths on Opioid Use Disorder: A Comparison of Reddit and Large Language Model. *Proc. ICWSM* (2025).
- [64] 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.
- [65] Linda Neuhauser and Gary L Kreps. 2003. Rethinking communication in the e-health era. *Journal of health psychology* 8, 1 (2003), 7–23.

- [66] U.S. News. 2023. U.S. News Launches Generative AI Search Across USNews.com. <https://www.usnews.com/info/blogs/press-room/articles/2023-11-21/u-s-news-launches-generative-ai-search-across-usnews-com>
- [67] Division of Health Care Services and Committee for the Study of the Future of Public Health. 1988. *The future of public health*. Vol. 88. National Academies Press.
- [68] Yngvild Olsen and Joshua M Sharfstein. 2014. Confronting the stigma of opioid use disorder—and its treatment. *Jama* 311, 14 (2014), 1393–1394.
- [69] Jesutofunmi A Omiye, Jenna C Lester, Simon Spichak, Veronica Rotemberg, and Roxana Daneshjou. 2023. Large language models propagate race-based medicine. *NPJ Digital Medicine* 6, 1 (2023), 195.
- [70] Anaelia Ovalle, Palash Goyal, Jwala Dhamala, Zachary Jagers, 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.
- [71] Leah Perkinson, Kimberley E Freire, and Meredith Stocking. 2017. Using essential elements to select, adapt, and evaluate violence prevention approaches. (2017).
- [72] Stephen R Pfohl, Heather Cole-Lewis, Rory Sayres, Darlene Neal, Mercy Asiedu, Awa Dieng, Nenad Tomasev, Qazi Mamunur Rashid, Shekoofeh Azizi, Negar Rostamzadeh, et al. 2024. A toolbox for surfacing health equity harms and biases in large language models. *Nature Medicine* 30, 12 (2024), 3590–3600.
- [73] 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.
- [74] 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.
- [75] Pragnya Ramjee, Mehak Chhokar, Bhuvan Sachdeva, Mahendra Meena, Hamid Abdullah, Aditya Vashistha, Ruchit Nagar, and Mohit Jain. 2024. ASHABot: An LLM-Powered Chatbot to Support the Informational Needs of Community Health Workers. *arXiv preprint arXiv:2409.10913* (2024).
- [76] 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/>
- [77] 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.
- [78] 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.
- [79] 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.
- [80] 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.
- [81] 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.
- [82] B. Turnock. 2009. *Public Health: What It Is and How It Works*. Jones & Bartlett Learning. <https://books.google.com/books?id=Z8dMleOJOuYC>
- [83] 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.
- [84] 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).
- [85] Gaurav Verma, Jiawei Zhou, Mohit Chandra, Srijan Kumar, and Munmun De Choudhury. 2025. A Framework for Situating Innovations, Opportunities, and Challenges in Advancing Vertical Systems with Large AI Models. *arXiv preprint arXiv:2504.02793* (2025).
- [86] 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).
- [87] Xun Wang and Robin A Cohen. 2023. Health information technology use among adults: United States, July–December 2022. (2023).
- [88] 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.
- [89] 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.
- [90] 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.
- [91] Langdon Winner. 1980. Do artifacts have politics? *Daedalus* 109, 1 (1980).
- [92] 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.

- [93] Jiacheng Xie, Ziyang Zhang, Shuai Zeng, Joel Hilliard, Guanghui An, Xiaoting Tang, Lei Jiang, Yang Yu, Xiufeng Wan, and Dong Xu. 2025. Leveraging Large Language Models for Infectious Disease Surveillance—Using a Web Service for Monitoring COVID-19 Patterns From Self-Reporting Tweets: Content Analysis. *Journal of Medical Internet Research* 27 (2025).
- [94] Ruiyun Xu, Yue Feng, and Hailiang Chen. 2023. Chatgpt vs. google: A comparative study of search performance and user experience. *arXiv preprint arXiv:2307.01135* (2023).
- [95] 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.
- [96] 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.
- [97] 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.
- [98] Yuehua Zhao and Jin Zhang. 2017. Consumer health information seeking in social media: a literature review. *Health Information & Libraries Journal* 34, 4 (2017), 268–283.
- [99] 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>
- [100] 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.