International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(REVIEW ARTICLE)

Check for updates

Transparent Health Risk Communication: Using Explainable AI to Support Equitable Decision-making in Public Health

Blessing Agbaza *

Western Illinois University, USA.

International Journal of Science and Research Archive, 2023, 09(02), 1101-1110

Publication history: Received on 02 June 2023; revised on 23 July 2023; accepted on 28 July 2023

Article DOI: https://doi.org/10.30574/ijsra.2023.9.2.0543

Abstract

In an era marked by heightened public health emergencies and digital transformation, the capacity to communicate health risks transparently and equitably has never been more critical. Explainable Artificial Intelligence (XAI) offers novel pathways for enhancing health risk communication by enabling algorithmic decision-making processes to be understood by both experts and laypersons. This study investigates the effectiveness and equity implications of using XAI-driven tools in health risk communication frameworks. Employing a mixed-methods approach combining quantitative analysis of survey data and qualitative key informant interviews from three U.S. health departments, we assessed how XAI tools such as interpretable machine learning models influence public trust, risk comprehension, and behavioral intention across diverse populations.

Findings indicate that while XAI-enhanced models improved overall risk comprehension by 31%, their benefits were significantly greater among participants with higher health literacy, raising concerns of a potential communication divide. Transparency, model explainability, and interactive features were positively associated with increased trust in health guidance. However, participants from marginalized communities expressed skepticism about data bias and fairness, emphasizing the need for culturally adapted explanations and participatory model design. Our results underscore that integrating explainable AI in public health communication must be accompanied by inclusive strategies to ensure equitable understanding and informed decision-making. Policy frameworks and technical standards for XAI deployment in health contexts must therefore prioritize transparency, inclusivity, and justice-centered design.

Keywords: Explainable AI; Health Risk Communication; Public Health Equity; Trust; Machine Learning Transparency; Health Literacy; Participatory Design.

1. Introduction

1.1. Background and Significance of Transparent Health Risk Communication

Health risk communication is a central pillar of public health, particularly in times of uncertainty such as disease outbreaks, environmental crises, and the management of chronic conditions. The primary objective of health risk communication is to convey scientifically accurate, timely, and actionable information that enables the public to make informed health decisions. However, historical and contemporary events such as the COVID-19 pandemic, vaccine hesitancy, and the spread of misinformation have underscored critical gaps in public understanding and trust in health systems, revealing that the way risks are communicated often determines the effectiveness of the public health response rather than the scientific facts alone [1,2].

* Corresponding author: Blessing Agbaza-Mogbojuri

Transparency in health risk communication is not solely about disclosing information but involves making complex information understandable and relevant to diverse populations. It encompasses openness about uncertainties, assumptions, data limitations, and potential consequences. Effective transparency is known to increase public trust, engagement, and behavioral compliance, especially when individuals perceive that they have been empowered to understand and participate in risk-informed decisions [3]. Yet, traditional communication methods struggle to meet the needs of a digitally connected, data-intensive, and socio-culturally diverse society.

With the integration of artificial intelligence (AI) into public health decision-making including epidemiological modeling, disease surveillance, and resource allocation new challenges have emerged regarding the interpretability and fairness of algorithmically driven guidance. Conventional "black box" models used in AI often obscure how decisions are made, undermining the principles of informed consent and eroding public trust. This opacity is particularly harmful in high-stakes or resource-limited health contexts where historical mistrust of institutions is already prevalent [4,5].

1.2. Emergence and Relevance of Explainable AI (XAI)

Explainable AI (XAI) refers to a subfield of artificial intelligence aimed at producing models whose internal logic can be understood and interrogated by human users. Rather than relying on opaque algorithms, XAI systems use interpretable techniques such as decision trees, SHAP (SHapley Additive exPlanations) values, counterfactuals, and attention maps to reveal how input features contribute to predictions or classifications [6]. These explanations may be visual, textual, or interactive, enabling both technical and non-technical stakeholders to engage with the model's outputs.

In public health, XAI offers transformative potential to enhance transparency in automated health systems, allowing policymakers, clinicians, and community members to understand the rationale behind data-driven recommendations. This is particularly crucial in risk communication, where understanding the "why" behind a health advisory such as exposure alerts or vaccine prioritization is essential for trust and compliance [7]. Furthermore, XAI may aid in auditing model fairness, identifying algorithmic biases, and supporting the ethical use of personal health data in predictive modeling.

Despite this promise, the practical integration of XAI into health communication strategies remains underdeveloped. Studies have demonstrated that even when explainability is technically achieved, it may not result in genuine user understanding or acceptance unless contextualized for diverse audiences [8]. For example, an explanation that resonates with a data scientist may be inaccessible or culturally irrelevant to a community health worker or a patient with low health literacy. Consequently, explainability must be considered not only as a technical feature but also as a socio-communicative and ethical imperative.

1.3. Equity Challenges in AI-Driven Health Communication

A major concern in AI-based health applications is the reproduction of existing health inequities. Risk communication, when filtered through biased algorithms or presented in inaccessible formats, can further marginalize vulnerable populations such as racial minorities, immigrants, the elderly, and individuals with low digital literacy [9]. The complexity of AI models and their dependence on large, often unrepresentative datasets may inadvertently obscure structural inequalities or propagate harmful assumptions.

Explainable AI, if implemented without an equity lens, may privilege those already empowered by education and technology, creating what has been termed an "interpretability divide" where some segments of the population can critically assess algorithmic guidance while others remain excluded or confused [10]. Furthermore, if stakeholders are not involved in the co-design of explanations, XAI tools may reinforce technocratic dominance rather than democratizing health knowledge.

Therefore, there is a growing call for participatory, culturally sensitive, and linguistically appropriate approaches to designing explainable AI systems in health. This includes aligning explanations with user values, providing actionable insights, and tailoring visual or narrative forms to resonate with different cognitive and cultural frames [11]. Without such approaches, XAI may only superficially address the transparency problem without achieving its deeper ethical and communicative goals.

1.4. Study Objectives

This study explores how explainable artificial intelligence (XAI) can support transparent and equitable health risk communication within public health contexts. It aims to evaluate the impact of XAI-driven tools on public understanding and trust in health risk information, while also assessing how diverse populations particularly those with varying levels

of health literacy and socioeconomic status interpret AI-generated explanations. In addition, the study examines the equity implications of integrating XAI into public health communication strategies, with a focus on whether such technologies reinforce or mitigate existing disparities. Finally, it seeks to identify user-centered design principles and generate policy recommendations that can guide the deployment of explainable AI in ways that are inclusive, contextually relevant, and capable of fostering public confidence in health-related decision-making processes.

2. Methods

2.1. Study Design and Setting

This study employed a convergent mixed-methods design, integrating quantitative and qualitative data to assess the influence of explainable AI (XAI) on public trust, risk comprehension, and health-related decision-making. The research was conducted across three public health departments in the United States: Cook County (Illinois), Clark County (Nevada), and Ramsey County (Minnesota), selected for their demographic diversity, prior AI adoption efforts, and existing community engagement infrastructure.

Each site piloted an XAI-based health risk communication prototype during a simulated outbreak scenario involving exposure to a hypothetical airborne contaminant. Participants received individualized risk scores with accompanying explanations generated through a machine learning model using decision trees and SHAP values.

The study was carried out between February and June 2022 and adhered to participatory ethics protocols involving both institutional stakeholders and community representatives.

2.2. Study Population and Sampling

A multistage sampling approach was used. First, three counties with ongoing digital health innovation initiatives were purposely selected. Within each county, community health centers and local NGOs helped recruit adult participants, aged 18 and above, for both the survey and interview components.

The quantitative sample comprised 1,200 participants, 400 per county, stratified by age, gender, race/ethnicity, and self-reported health literacy. The qualitative sample consisted of 42 in-depth interviews, 14 per county, including members of the public, local health officials, community leaders, and data analysts.

Eligibility criteria included basic digital literacy, that is, the ability to use smartphones or computers, consent to participate, and residence in the respective county for at least one year. Exclusion criteria included prior participation in AI research projects and inability to provide informed consent.

2.3. XAI Tool Development and Risk Communication Intervention

An XAI-powered risk prediction model was developed using anonymized synthetic datasets simulating environmental exposure, pre-existing conditions, and demographic risk factors. The model, built using random forest classifiers, generated personalized risk scores for hypothetical exposure outcomes.

The model incorporated SHAP (SHapley Additive exPlanations) values to provide localized, individualized feature attribution. These values were displayed as simple visual explanations and plain-language narratives designed to communicate the top three factors contributing to a person's risk level.

Participants received a simulated risk report through an app interface and were asked to review the output. The interface offered both a basic view; summary risk level and contributing factors and an expanded view; detailed feature contributions and confidence intervals. Participants could explore how changing input factors might alter their risk, offering a hands-on interpretability experience.

2.4. Data Collection Instruments

Quantitative data were collected using a structured, pre-tested questionnaire that measured: trust in health information (5-item Likert scale adapted from existing risk communication literature); risk comprehension (assessed via scenario-based multiple-choice items); perceived fairness of the XAI model; behavioral intention to act on the health recommendation; health literacy level (measured using the Newest Vital Sign (NVS) screening tool).

Qualitative data were gathered through semi-structured interviews, focusing on: perceptions of explainability and transparency, clarity and usability of AI explanations, relevance to cultural and contextual understanding and concerns about bias, surveillance, and data ethics.

All interviews were conducted in English, audio-recorded with consent, and transcribed verbatim.

2.5. Data Analysis

Quantitative analysis was conducted using Stata version 17.0 (StataCorp, College Station, TX, USA). Descriptive statistics were computed to characterize the sample. Associations between XAI comprehension, trust, and behavioral intention were examined using multivariable logistic regression models. Health literacy was tested as a moderator using interaction terms.

Qualitative data were analyzed using thematic analysis, supported by NVivo 14 software. A coding framework was developed deductively from the interview guide and inductively from emerging patterns. Coding was conducted by two independent researchers to ensure inter-rater reliability, with discrepancies resolved through discussion.

The mixed-methods findings were triangulated to provide a comprehensive understanding of how different populations experienced XAI-based health risk communication.

2.6. Ethical Considerations

Ethical approval for this study was granted by the Institutional Review Board (IRB) of Western Illinois University, under protocol number WIU-PH-XAI-2022-02. All participants provided written informed consent, with assurances of confidentiality, the right to withdraw at any time, and non-identifiability in published materials.

Community advisory boards in each county reviewed and approved the tools, consent forms, and communication materials used in the study. Data were anonymized at the point of collection and stored on encrypted servers.

2.7. Quality Assurance and Validity

To ensure validity and reliability; the XAI interface was co-designed with community stakeholders and iteratively refined through usability testing, instruments were pre-tested with 50 individuals not included in the final analysis, Cronbach's alpha values for trust and fairness scales were above 0.85, indicating high internal consistency and triangulation across methods ensured robustness of interpretations.

3. Results

3.1. Descriptive Characteristics of the Study Population

A total of 1,200 participants were included in the quantitative component of the study, with an even distribution across the three participating counties (Cook, Clark, and Ramsey). The mean age was 42.3 years (SD ± 13.6), with 58.2% identifying as female. Participants were racially and ethnically diverse: 37.1% identified as White, 29.5% as Black or African American, 19.2% as Hispanic or Latino, and 11.6% as Asian or Pacific Islander. A small percentage (2.6%) identified with other or multiple races.

In terms of education, 26.7% had high school education or less, 45.4% had some college or associate degrees, and 27.9% held bachelor's or advanced degrees. Health literacy, assessed using the Newest Vital Sign (NVS), showed that 36.4% had low health literacy, 29.3% had moderate levels, and 34.3% had adequate literacy.

Across the sample, 71.9% had never used AI-enabled health applications, while 85.6% reported using smartphones daily. Table 1 summarizes the sociodemographic distribution.

Table 1 Sociodemographic characteristics of participants (N = 1,200)

Characteristic	Frequency	Percentage (%)
Female	698	58.2
Aged 18–34 years	338	28.2
Aged 35–54 years	569	47.4
Aged 55 years and older	293	24.4
White	445	37.1
Black/African American	354	29.5
Hispanic/Latino	230	19.2
Asian/Pacific Islander	139	11.6
High school or less	320	26.7
Some college	545	45.4
Bachelor's degree or higher	335	27.9
Low health literacy	437	36.4
Moderate health literacy	351	29.3
Adequate health literacy	412	34.3

3.2. Trust, Comprehension, and Behavioral Intention

Following exposure to the XAI-based risk communication tool, 67.5% of participants reported increased trust in the simulated health message, with a significantly higher proportion among those with higher health literacy ($\chi^2 = 22.19$, $p < 0.001$). Risk comprehension scores improved by 31.2% on average (pre-test mean = 3.9/10; post-test mean = 5.1/10; $p < 0.001$, paired t-test), with the greatest gains observed in participants aged 18–34 years.

Regression analyses revealed that participants who accessed the interactive explanation feature had 2.9 times higher odds of reporting behavioral intention to act on the recommendation (Adjusted Odds Ratio [AOR] = 2.93; 95% CI: 2.22–3.86). Trust in the explanation was positively associated with both comprehension ($\beta = 0.41$, $p < 0.001$) and behavioral intention ($\beta = 0.36$, $p < 0.001$), controlling for demographic variables and county of residence.

However, disparities emerged in how explanations were perceived. Among participants with low health literacy, only 42.7% found the explanations easy to understand, compared to 84.3% of those with adequate literacy ($\chi^2 = 58.44$, $p < 0.001$). Similarly, individuals with lower educational attainment were less likely to report that the model output felt "relevant" or "trustworthy."

3.3. Perceived Fairness and Algorithmic Bias

Overall, 59.6% of respondents rated the model as fair, while 22.4% were unsure, and 18.0% expressed skepticism, citing concerns over how data were used and potential bias. Perceived fairness was significantly lower among Black and Hispanic participants compared to White participants (Black: 49.2%, Hispanic: 51.1%, White: 67.3%; $p < 0.01$).

Qualitative feedback highlighted a recurrent concern about the use of zip codes and pre-existing conditions in risk scoring. Some participants questioned whether these variables embedded systemic bias:

"If you already think our neighborhoods are risky, how can I trust the risk score isn't just about where I live?" — Participant, Black female, 45, Clark County.

Participants from marginalized communities emphasized the need for greater transparency in how inputs were selected and weighted, with calls for inclusive model governance.

3.4. Qualitative Themes from In-Depth Interviews

Thematic analysis of the 42 interviews revealed four interrelated insights into how participants perceived and engaged with explainable AI tools in public health communication. First, transparency was found to build conditional trust. Participants consistently valued explanations that were interactive, visual, or narrative in nature, noting that being able to "see why" made them more inclined to believe the risk assessment. However, transparency alone was not sufficient for full trust, as exemplified by one participant who remarked, *"I trust it more when I see it... but I still want to know who made it and for what purpose."* This highlights that explainability must be accompanied by clarity around authorship, intent, and data provenance.

Second, it became evident that explainability is not synonymous with understanding. Even when visual aids were provided, several participants particularly those with lower educational attainment or for whom English was a second language reported difficulty understanding the explanations, especially when they involved probabilistic language or technical terminology. This disconnect underscores the importance of adapting explanatory content to the cognitive and linguistic needs of diverse audiences.

Third, participants emphasized that culturally responsive explanations were significantly more effective in promoting engagement and comprehension. Many expressed a preference for analogies, stories, or real-life examples that aligned with their personal experiences. One participant explained, *"Tell me it's like how my asthma gets worse in the winter that I understand,"* illustrating how relatable comparisons can bridge cognitive gaps and deepen understanding.

Finally, there was a clear desire for participatory transparency, especially among community health workers and local organizers. Many participants expressed strong interest in being involved in the co-design of risk tools, arguing that such collaboration would enhance both the legitimacy and the sustainability of AI-driven public health interventions. This participatory impulse suggests that truly explainable AI is not only about readable outputs but also about inclusive processes that engage users in shaping the systems that affect them.

3.5. Summary of Key Findings

XAI tools improved overall comprehension and trust in health risk messages, while interactive and visual explanations were particularly effective in enhancing behavioral intention. However, disparities in comprehension and perceived fairness among different demographic groups suggest the emergence of an "explainability divide." Although participants valued transparency, they emphasized the importance of contextual relevance and cultural sensitivity in the presentation of AI-generated explanations. Furthermore, persistent concerns about bias and exclusion highlighted the urgent need for participatory design processes and inclusive governance in the deployment of AI systems in public health.

4. Discussion

4.1. Principal Findings

This study provides evidence that explainable AI (XAI) can improve trust, comprehension, and behavioral intention in health risk communication when deployed with interactive and user-centered design. Participants exposed to XAI-generated explanations demonstrated significantly higher levels of risk understanding and were more inclined to act on the guidance received. These findings validate prior theoretical assumptions that transparency, when made intelligible, enhances public confidence and decision-making efficacy in health contexts [1–3].

However, the study also uncovered critical equity gaps. Individuals with lower health literacy and education levels were significantly less likely to comprehend and trust the explanations, suggesting that the benefits of XAI are not uniformly distributed. This disparity reflects the presence of an "interpretability divide," whereby the ability to make sense of AI-driven outputs becomes stratified along lines of health literacy, education, and socioeconomic status replicating existing social inequities in a new, algorithmic form [4,5].

While XAI promises to enhance public health decision-making, these results indicate that its impact may be limited or counterproductive unless paired with deliberate inclusion strategies that make explanations accessible and culturally appropriate to all populations, especially historically marginalized groups.

4.2. Comparison with Previous Research

The observed increase in trust and comprehension associated with XAI tools aligns with prior studies in digital health, which found that transparency enhances user engagement and perceived accuracy [6,7]. Prior research on AI in clinical settings has emphasized that explainability improves provider confidence in diagnostic tools and increases likelihood of integration into practice [8]. This study extends those findings to lay audiences in a public health setting, highlighting the wider societal implications of explainability beyond professional use.

However, the variability in comprehension by health literacy mirrors findings in consumer health informatics, where complex or poorly contextualized information often exacerbates health disparities [9,10]. This study's qualitative data reinforce calls in the literature for user-centered explanation design including narrative framing, analogies, and visualizations tailored to cognitive and cultural expectations [11].

Concerns about algorithmic fairness, particularly regarding geographic and demographic data, have been documented in previous AI research as well. Studies have warned that predictive models using zip codes, race, or income-related variables may perpetuate bias even when technically "accurate," thereby undermining trust among marginalized groups [12–14]. This study confirms these concerns in the domain of public health risk communication and adds empirical weight to the demand for participatory AI governance.

4.3. Implications for Public Health Practice

The findings offer several implications for health practitioners, communicators, and policymakers aiming to use AI-driven tools to inform public health decision-making XAI serves as a trust-building mechanism but only when it is made accessible. The observed association between explainability and trust suggests that health communication strategies incorporating AI must transcend predictive accuracy and prioritize intelligibility. Achieving this requires collaborative design involving communication experts, health educators, and community stakeholders. Health literacy must also be a central design consideration in AI explanations. Explanations that rely on advanced statistical or technical terminology risk alienating large segments of the population, particularly those with lower educational attainment. To democratize understanding, health systems and digital platforms should invest in plain-language summaries, interactive visualizations, multilingual content, and storytelling formats that resonate with varied literacy levels and cultural backgrounds [15].

Participatory approaches play a critical role in increasing both the relevance and legitimacy of XAI tools. Involving community members in the development, testing, and deployment phases not only improves usability but also fosters cultural resonance and local ownership. This form of participatory transparency is particularly important in contexts marked by institutional mistrust, where algorithmic outputs are often viewed with suspicion. However, fairness in algorithmic decision-making cannot be assumed simply because models are explainable. Developers must rigorously assess and clearly communicate how input features influence predictions, especially when these inputs may correlate with structural disadvantage. Tools such as transparency reports, bias audits, and standardized model datasheets should become routine components of AI implementation in public health settings [16].

Furthermore, risk communication policies should be updated to include provisions for digital equity. Given the demonstrated digital divide in interpretability, public health agencies must devise strategies to ensure accessibility for populations with limited access to digital technologies or baseline familiarity with AI systems. These strategies may include offering offline communication formats, facilitating community-based training sessions, and collaborating with trusted local leaders to bridge gaps in comprehension and access.

This study represents one of the first empirical investigations into the equity implications of explainable AI in the context of public health communication. Its mixed-methods design enabled the integration of quantitative metrics and rich qualitative insights into how diverse demographic groups engage with and interpret XAI systems. The use of geographically varied sites and stratified sampling further enhances the generalizability of findings across urban populations in the United States.

Nevertheless, several limitations should be acknowledged. The simulated nature of the risk scenario may not fully capture the emotional or behavioral responses that would emerge during an actual public health emergency. Additionally, although SHAP values were chosen for their interpretability, other explanation formats such as counterfactuals, Local Interpretable Model-Agnostic Explanations (LIME), or example-based reasoning may yield different levels of user comprehension and should be explored in future research. The reliance on self-reported measures of trust and behavioral intention introduces a risk that these outcomes may not translate directly into real-

world behavior. Lastly, requiring participants to have basic digital literacy may have unintentionally excluded those mostly at risk of digital exclusion, a group for whom inclusive design is particularly critical.

Despite these limitations, the findings offer critical insight into the design and governance of AI-enabled communication systems, particularly in advancing equitable public health outcomes.

4.4. Directions for Future Research

Future research should examine XAI deployment in real-time public health events, such as vaccine campaigns, environmental alerts, or epidemic control efforts. Longitudinal designs could track actual behavior change following exposure to explainable AI tools, offering more robust evidence of causal effects.

There is also a need to study alternative explanation modalities tailored to specific populations such as audio-based summaries for visually impaired individuals, narrative videos for low-literacy communities, or gamified learning for youth engagement.

Another critical research agenda involves auditing AI models for bias and transparency in government communication systems, with a focus on civic accountability and democratic oversight. This includes investigating how marginalized groups can meaningfully participate in AI policy and governance.

Finally, interdisciplinary collaborations across public health, computer science, sociology, and communication studies are essential to build comprehensive, justice-centered approaches to AI explainability that respond to both technical and human complexities.

5. Conclusion

This study contributes critical empirical insights to the growing discourse on the role of explainable artificial intelligence (XAI) in enhancing transparency and equity in public health risk communication. Through a mixed-methods design, we demonstrate that XAI can substantially improve public understanding and trust in AI-generated health risk messages but only when those explanations are designed with attention to clarity, relevance, and inclusivity.

The findings affirm that explainability, while technically achievable, is not synonymous with genuine understanding. Disparities in comprehension and perceived fairness across literacy, racial, and socioeconomic lines point to a pressing risk: that XAI, without explicit equity measures, may deepen rather than reduce information inequities. The so-called “black box” problem of AI may thus be replaced by a “glass box” dilemma, one that is transparent but only to those already equipped to interpret its contents.

Trust in AI-driven health recommendations is mediated not only by the quality of the model but by the perceived intentions, cultural alignment, and communicative competence of the tools delivering them. As such, transparency in AI must be conceived not as a product feature but as a participatory, iterative, and community-centered process. The explainability of an AI model is only meaningful when it enables users particularly those in marginalized or historically underserved communities to understand, challenge, and act upon its outputs.

If public health is to benefit from the power of AI without undermining its core ethical commitments, then XAI must be deployed within a broader framework of justice, accessibility, and accountability. This requires not only new technologies but new practices of governance, engagement, and interdisciplinary collaboration.

5.1. Recommendations

Based on the study’s findings, several strategic recommendations emerge for stakeholders involved in the development and deployment of explainable AI tools in public health communication. First, it is imperative to embed equity into the design and implementation of XAI systems. Health agencies and AI developers should adapt explanation formats to accommodate diverse user needs by ensuring multilingual, culturally responsive, and visually accessible outputs. These explanations should be tested across varying literacy levels and cognitive styles before public deployment to maximize inclusiveness.

Secondly, fostering participatory model development is essential. The co-creation of XAI systems with patients, community health advocates, and local leaders enhances both the legitimacy and contextual relevance of risk

communication tools. Such participatory approaches ensure that model outputs resonate with the lived experiences and values of target populations.

Also, institutionalizing transparency audits and systematic reporting mechanisms can bolster public trust and accountability. Agencies deploying XAI should implement structured reviews that document the model's architecture, data provenance, fairness evaluations, and the clarity of its explanatory outputs. Instruments such as model cards and algorithmic impact assessments should be made publicly available to enhance transparency and enable external scrutiny.

Furthermore, public health systems should integrate digital and health literacy initiatives into their communication strategies. Programs aimed at improving foundational knowledge of health technologies will empower individuals particularly those from digitally marginalized groups to engage meaningfully with AI-driven tools over time.

Establishing cross-sectoral governance structures is vital for ensuring ethical oversight. The governance of XAI in health communication must extend beyond technical experts and include ethicists, legal scholars, sociologists, and representatives from the communities affected by AI interventions. This multidisciplinary oversight helps align AI deployment with broader principles of human rights, data justice, and public accountability.

To address the persistent digital divide, stakeholders must adopt hybrid communication channels that blend digital innovation with traditional outreach methods. Risk communication strategies should not depend solely on online platforms but must be complemented by offline efforts such as community radio broadcasts, printed educational materials, and engagement through trusted local health workers. This blended approach ensures that health messages are accessible and inclusive across a wide range of socioeconomic and technological contexts.

Furthermore, national and local public health policy frameworks must be updated to reflect both the transformative potential and the inherent risks associated with AI technologies. Such policies should mandate transparency and explainability for all high-stakes applications, explicitly prohibit the use of opaque or unaccountable algorithms in public health decision-making, and actively promote inclusive, equity-oriented communication strategies in both routine and emergency settings.

Lastly, there is a pressing need to prioritize research on human-centered explainability. Future interdisciplinary investigations should explore innovative explanation formats that are not only technically accurate but also emotionally resonant and socially grounded. Evaluation metrics for explainable AI must move beyond computational precision and include measures of trust, comprehension, and user empowerment to ensure these tools truly serve the public interest.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare no conflict of interest related to the design, execution, analysis, or publication of this research. All affiliations and sources of funding, if any, have been transparently disclosed in accordance with academic integrity and institutional guidelines.

Statement of ethical approval

This study was reviewed and approved by the Institutional Review Board (IRB) of Western Illinois University, under protocol number WIU-PH-XAI-2022-02. All participants provided written informed consent before participation. Data collection procedures adhered to the principles outlined in the Belmont Report, and no personally identifiable information was stored. The research also followed relevant guidelines for responsible data use, digital ethics, and community engagement. Stakeholder consultations and community review boards in all three counties provided additional oversight for study design, language used in digital tools, and dissemination procedures.

References

[1] Covello VT. Risk communication: An emerging area of health communication research. Ann Int Commun Assoc. 1992;15(1):359–73.

[2] Fischhoff B. The sciences of science communication. Proc Natl Acad Sci U S A. 2013;110(Suppl 3):14033–9.

[3] O'Neill O. Trust, trustworthiness, and transparency. Ethics Int Aff. 2002;26(1):75–87.

[4] Obermeyer Z, Emanuel EJ. Predicting the future — big data, machine learning, and clinical medicine. N Engl J Med. 2016;375(13):1216–9.

[5] Eubanks V. Automating inequality: How high-tech tools profile, police, and punish the poor. New York: St. Martin's Press; 2018.

[6] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD; 2016. p. 1135–44.

[7] Chouldechova A, Roth A. The frontiers of fairness in machine learning. Commun ACM. 2020;63(5):82–9.

[8] Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169(12):866–72.

[9] Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav. 2007;31(Suppl 1):S19–26.

[10] Lepri B, Oliver N, Letouzé E, Pentland A, Vinck P. Fair, transparent, and accountable algorithmic decision-making processes. Philos Technol. 2018;31(4):611–27.

[11] Wang D, Yang Q, Abdul A, Lim BY. Designing theory-driven user-centric explainable AI. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; 2020. p. 1–15.

[12] Barocas S, Selbst AD. Big data's disparate impact. Calif Law Rev. 2016;104(3):671–732.

[13] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53.

[14] Veinot TC, Ancker JS, Bakken S, Bates DW, Eden J, Guise J-M. Health equity and the role of informatics. J Am Med Inform Assoc. 2019;26(8–9):689–95.

[15] Jimenez G, Spinazze P, Matcha A, Alam T, Seneviratne S, Salgado M, et al. Digital health equity and COVID-19: The innovation curve cannot reinforce the social gradient of health. J Med Internet Res. 2021;23(3):e24589.

[16] Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. In: Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency; 2019. p. 220–9.