Implementation Science

METHODOLOGY

Open Access

The Practical, Robust Implementation and Sustainability (PRISM)-capabilities model for use of Artificial Intelligence in community-engaged implementation science research

Nabila El-Bassel[1*], James David[1], Trena I. Mukherjee[1], Maneesha Aggarwal[2], Elwin Wu[1], Louisa Gilbert[1], Scott Walters[3], Redonna Chandler[4], Tim Hunt[1], Victoria Frye[1], Aimee Campbell[5], Dawn A. Goddard-Eckrich[1], Katherine Keyes[6], Shoshana N. Benjamin[1], Raymond Balise[7], Smaranda Muresan[8], Eric Aragundi[9], Marc Chen[2], Parixit Davé[2], David Lounsbury[10], Nasim Sabounchi[11], Dan Feaster[6], Terry Huang[11] and Tian Zheng[9]

Abstract

Background Community-engaged research (CER) leverages knowledge, insights, and expertise of researchers and communities to address complex public health challenges and improve community well-being. CER fosters collaboration throughout all research phases, from problem identification and implementation to evaluation. Artificial Intelligence (AI) could enhance the collaborative process by improving data collection, analysis, insight, and engagement, while preserving research ethics. By integrating AI into CER, researchers could enhance their capacity to work collaboratively with communities, making research more efficient, inclusive, and impactful. However, careful consideration must be given to the ethical and social implications of AI to ensure that it supports the goals of CER. This paper introduces the PRISM-Capabilities model for AI to promote a human-centered approach that emphasizes collaboration, transparency, and inclusivity when using AI within CER.

Methods The PRISM-Capabilities model for AI includes six components to ensure that ethical concerns are addressed, trust and transparency are maintained, and communities are equipped to use and understand AI technology. This conceptual model is specifically tailored for community-engaged implementation science research, facilitating close collaboration between researchers and community partners to guide the use of AI throughout. This paper also proposes next steps to validate the model using the HEALing Communities Study (HCS), the largest community-engaged research study to date, which aimed to reduce fatal overdose deaths in 67 highly impacted communities in the United States.

Case study The PRISM-Capabilities model consists of six components: Optimizing engagement of implementers, settings, and recipients; characteristics of intervention implementers, settings, and recipients; equity assessment and risk management; implementation and sustainability infrastructure; external environment; and ethical assessment

*Correspondence:
Nabila El-Bassel
ne5@columbia.edu
Full list of author information is available at the end of the article

and evaluation. Although AI was not initially used during the HCS, we highlight how AI will be leveraged to complete post-hoc analyses of each of the six components and validate the PRISM-Capabilities model.

Conclusion The application of AI to CER relies on human-centered principles that prioritize human-AI collaboration, allowing for the operationalization of responsible AI practices. The PRISM-Capabilities model provides a framework to account for the complexities of real-world social science problems and explicitly positions AI tools at bottlenecks experienced with conventional approaches.

Keywords AI, Community-engaged, Framework, Methodology, Ethics, Opioid

Contributions to the literature (96/100)

• Highlights how AI could strengthen community-engaged implementation science research by improving data processes and engagement, while maintaining a collaborative and ethical approach.

•Introduces the PRISM-Capabilities model for AI, which emphasizes practical, ethical, and socially responsible research.

•Provides a plan and checklist for ethically driven AI in community implementation research.

•The model offers a potential tool to guide the use of AI in future implementation science studies and validates the model using the HEALing Communities Study (HCS), to provide robust real-world context, strengthening the model's applicability and relevance in addressing public health issues through implementation science research.

Introduction

Artificial intelligence (AI) is transforming social science research by enabling large-scale data analysis, simulation of complex systems, and new understanding of human behavior [1]. Initially limited to automating data processes [1], AI has since advanced social science by supporting sentiment analysis, predictive modeling, and pattern recognition [2, 3], thereby expanding the reach, precision and power of research [4]. Early AI applications in social science were often top-down, relying on existing datasets and excluding community perspectives, which led to algorithmic bias and lack of cultural nuance [5–7]. In response, social science researchers have adopted participatory approaches in AI that prioritize community co-design, transparency, and ethical oversight [1, 8, 9]. Parthasarathy and Katzman (2024) emphasize that integrating marginalized communities into AI design not only improves equity but aligns with grassroots knowledge to address the needs of the community [6]. Tools like model auditing and feedback loops now bring stakeholders to identify and correct AI bias during the research development phase [10]. This participatory shift is visible in fields like healthcare and implementation science, where stakeholders shape the use of AI in diagnostic technology tools [11–14]. For instance, in diabetes research, predictive models have been developed through community input to include variables like food insecurity and transportation [15]. In substance use and mental health studies, natural language processing (NLP) tools co-created with communities have helped identify stigma and tailor interventions for different cultural contexts [8, 16]. Achieving equitable, contextual and community-driven use of AI, however, requires community engagement frameworks. This includes building shared language, decision-making frameworks, and design processes that bridge data science, implementation science, and lived experiences.

Ultimately, AI's role is not to automate decision-making but to augment human input and judgment, thereby enhancing adaptability, reducing implementation fatigue, and supporting ethical and sustainable community-engaged research (CER). By centering human-AI partnerships and prioritizing transparency, researchers could ensure that AI supports outcomes that are culturally and contextually responsive. The literature underscores the considerable potential of AI to enhance CER [15, 17], but also highlights a significant gap in conceptual models to guide the ethical application of AI [18]. This raises the need for a conceptual model that emphasizes a human-centered approach to AI use which minimizes bias in each step of CER [19]. In this paper, we introduce the PRISM-Capabilities model for AI as a conceptual model to guide the integration of AI and CER that is grounded in local knowledge and expertise.

Methods

An integrated conceptual model to guide the use of AI in community-engaged implementation science research: the PRISM-Capabilities Model for AI

The PRISM-Capabilities model for AI integrates the Practical, Robust Implementation and Sustainability Model (PRISM) [20] with the Capabilities Approach [21, 22] (Fig. 1). When combined, this model addresses historical shortcomings of AI in social science and CER by promoting ethical, human-centered collaboration. This ensures that research aligns with the values, morals, and

Fig. 1 This figure illustrates the six interconnected and mutually reinforcing components of the PRISM-Capabilities model for AI, as applied to community-engaged research (CER)

needs of the communities being served, so that AI is a complement, rather than a replacement to human efforts. This approach also fosters participatory processes, shared learning, co-design, and co-ownership to ensure that AI-enabled CER is guided by community voice and lived experiences [23–25].

PRISM is an implementation science framework for designing, delivering, and evaluating interventions [26], that incorporates the RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance) conceptual model [20]. We selected the PRISM framework because it explicitly incorporates organizational characteristics (e.g., culture, leadership), external environments (e.g., policy, funding) and perspectives of multiple stakeholders (patients, providers, administrators, funders), making it particularly well-suited for complex, real-world settings, and practical for implementation in diverse environments and at various levels (local community, nationally). In addition to implementation outcomes, PRISM focuses on sustainability and continuous feedback loops to support long-term change and ongoing improvement, which other implementation science frameworks may overlook. The domains of PRISM directly correspond to the types of data and decision points where AI methods (such as NLP, fairness audits, and simulation modeling) excel by enabling continuous learning, multilevel monitoring, and rapid feedback. PRISM allows for a holistic assessment of implementation efforts, including both process and outcome measures across multiple levels (patient, provider, organization, system) [20, 26]. The PRISM-Capabilities model for AI emphasizes iterative feedback and systems thinking, [29, 30] making PRISM the most suitable implementation science framework for use with AI tools. Moreover, PRISM guided the HEALing communities Study [31], which will be used to illustrate the PRISM-Capabilities model for AI in this paper.

The Capabilities Approach focuses on the freedoms and conditions that enable individuals and

communities to achieve their goals [21, 27, 28]. When applied to CER, the Capabilities Approach emphasizes ethical imperatives rooted in autonomy and human dignity [32], which are critical when AI influences decisions and outcomes. It underscores the need to ground AI in local realities and ensure that individuals have a hand in shaping the data and insights that affect their communities [22, 28]. It also enhances transparency and accountability by embedding community voices in every phase of AI development and use [33]. This"bottom-up"approach elevates community contributions through shared ownership and local knowledge [34]. The PRISM-Capabilities model thus ensures that AI solutions are culturally relevant and tailored to community priorities, fostering equitable and effective outcomes.

As a non-linear model, PRISM-Capabilities supports iterative feedback aligned with human-centered design (HCD), where rapid cycles of human-AI collaboration refine implementation in real time. By incorporating systems thinking [35, 36], the model addresses the interconnected influences that shape CER outcomes. Co-creation and shared goals, such as clarifying the benefits of AI use, addressing bias, and transparency are central to this process. Ultimately, the model positions community members as co-designers, co-analysts, and co-stewards of AI-enabled CER.

This paper first presents an overview of the PRISM-Capabilities model for AI (Fig. 1). Next, this paper uses the HEALing Communities Study (HCS), the largest implementation science study ever funded to address substance use [31], as a retrospective use case to demonstrate the PRISM-Capabilities model. Although AI was not widely available during the implementation phase of HCS, it could have enhanced the CER process. This paper strengthens the empirical foundation of the PRISM-Capabilities model for AI by describing post-hoc analyses that will be completed to simulate the real-time utility of AI during HCS implementation to fully capitalize on the extensive dataset generated by the HCS, while also presenting limitations. Finally, we describe the potential technical limitations of AI such as hallucinations, explainability challenges, automation risks and algorithmic bias, which could undermine ethical CER implementation, while also proposing safeguards.

The interconnected components of the PRISM-Capabilities model for human-AI collaboration in CER

By delineating the model's six components (Table 1), we offer a practical blueprint for translating the conceptual model into action. The model supports real-world application by detailing specific data types and analytic techniques (e.g., NLP, fairness audits, simulation modeling), promoting transparent human-AI collaboration, and

Table 1 PRISM-capabilities model components for human-AI collaboration in community-engaged research

Purpose	Example AI Tools & Methods	Example Implementation Questions
Optimizing Engagement Ensure early, inclusive co-definition of research problems with stakeholders; identify engagement gaps and community priorities	NLP (sentiment analysis, topic modeling); ML for engagement forecasting	Who are key partners? What engagement gaps exist? What early barriers can AI detect?
Characteristics of Implementers, Settings, Recipients Adapt interventions to organizational readiness and local context; enable real-time adjustments	NLP (readiness signals); ML (site clustering); SHAP, LIME for transparency	What's the organizational context? How do interventions align with site-specific needs?
Equity Assessment & Risk Management Monitor disparities in implementation and outcomes; ensure real-time fairness auditing	NLP (bias detection); ML (risk prediction, fairness audits); dashboards	Are disparities emerging? How do AI tools support equitable resource allocation?
Implementation & Sustainability Infrastructure Support long-term planning, assess fidelity, and optimize resources	Simulation (system dynamics, agent-based); NLP (session analysis); ML (forecasting)	What resources are needed long-term? How can drift in intervention fidelity be detected early?
External Environment Anticipate how policy, organizational, or regulatory shifts influence CER success	NLP (policy analysis); ML (trend detection); geospatial mapping	How do structural factors support or limit implementation? What external threats exist?
Ethical Assessment & Evaluation Build procedural justice, transparency, and accountability into all AI-supported activities	NLP (ethical flagging); SHAP, LIME; participatory audit tools	Are AI decisions explainable? How are community values integrated?

surfacing key questions for participatory co-design. The guide is tailored to help research teams, AI experts, and community partners use AI in ways that enhance trust, contextual responsiveness, and ethical accountability throughout all phases of CER implementation.

Optimizing engagement of implementers, settings, and recipients

The PRISM-Capabilities model for AI begins with gathering data from key implementers, organizational leaders, community members, individuals with lived experience, and policymakers to identify challenges and priorities to improve intervention acceptability. Community engagement at this stage aims to co-define the research question, identify barriers, and ensure that diverse stakeholder voices are included from the outset of CER implementation. Human-AI collaboration in this phase could generate real-time insights using NLP, sentiment analysis and other AI tools when drawing from qualitative data to support more inclusive and effective implementation. Topic modeling could also be applied to meeting transcripts to identify recurring themes in engagement and trust, helping to tailor implementation strategies to local needs.

AI tools could help answer questions like: *Are there emotional tone, morale, or participation gaps across stakeholder groups? Who are the key community partners?* (identified via NLP in meeting transcripts)? *What is the state of organizational infrastructure* (assessed through partner feedback and documents)? *How is the intervention perceived* (measured through sentiment analysis)? *What prior experiences shape implementer perspectives? What external factors, policy, funding, or local support might be barriers to implementation? What skills and training gaps exist among implementers* (identified through performance records)? When answering such questions, topic modeling could uncover recurring themes to enhance an understanding of implementation challenges. Additionally, ML methods could support responsive and equitable decision-making by synthesizing diverse datasets such as demographic trends, local health outcomes, and economic indicators to construct a dynamic, data-driven model of the implementation context.

Characteristics of implementers, settings, and recipients

This component considers the skills, capacities, readiness, and contextual factors of the individuals and systems involved in CER implementation to ensure alignment with local needs, contexts and available resources. This is achieved by incorporating feedback from all stakeholders early in the CER process and enabling continuous refinement of core components and implementation strategies. To ensure contextual fit, all stakeholders must assess whether interventions align with community needs, values, and available resources. SDM could be used to visualize variations across sites using NLP to enhance the process of understanding site-level differences in organizational readiness and capacity. SDM data sources may include in-depth interviews, focus group discussions, administrative records, and surveys. Real-time sentiment analysis could support timely adjustments by addressing questions like: *How are participants responding to the intervention?* Or *What changes could increase impact?*

Importantly, readiness indicators and other features used in AI models should be co-developed with community input. Tools like SHapley Additive exPlanations (SHAP) or Local Interpretable Model-agnostic Explanations (LIME) could help make AI outputs interpretable and actionable [37]. By quantifying how much each feature contributes to an individual prediction, SHAP and LIME enable transparent, consistent, and locally accurate explanations of complex ML models and could be used for auditing AI models**,** identifying bias, or building trust with stakeholders. AI models could also automate routine tasks and improve decision-making, and engagement.

Equity assessment and risk management

The next component identifies potential disparities in implementation and outcomes and ensures inclusive access to benefits across diverse populations. When used in CER, AI could enhance and ensure equity assessment and risk management by continuously analyzing performance data, detecting trends in real time and supporting equitable intervention distribution [38, 39]. These tools could uncover disparities in participation, access and outcomes, particularly when implemented in collaboration with communities.

NLP methods, including supervised classification and unsupervised clustering, analyze meeting transcripts, interviews, and narratives to detect linguistic biases, exclusionary framing, and disparities in how underrepresented groups are being discussed by various stakeholders. These analyses could help identify patterns of disparities and disproportionate burden, prompting timely adaptations. AI could also integrate demographic and other contextual data to guide equitable resource allocation and performance using indicators such as race, income, geography, or criminal-legal system involvement [40]. AI dashboards and fairness audits that are stratified by these variables could be used to visualize emerging inequities and track subgroup disparities to shape inclusive and effective interventions [40, 41].

Ultimately, equity assessment in this model is not just about data accuracy, but also about participatory

oversight and actionable insights that reduce harm for all populations. To ensure equity metrics are transparent, accountable and meaningful, communities must co-define risks by selecting which disparities to track, how to interpret subgroup errors, validate algorithmic outputs, and what thresholds warrant action. Fairness-aware modeling (e.g., demographic parity checks and disparate impact audits) must be implemented as a continuous auditing mechanism that is governed collaboratively and continuously, rather than as a one-time evaluation. This transforms equity assessments into a dynamic, corrective mechanism that moves beyond static disparity reporting to enable actionable, real-time mitigation. Furthermore, researchers could develop AI-driven dashboards to allow for transparency and data-informed outputs to enable CER implementation teams to respond quickly.

Implementation and sustainability infrastructure

The PRISM-Capabilities model for AI highlights the importance of contextual factors in building and sustaining implementation systems [27, 42]. This component of the model evaluates organizational systems, resource flows, training, and operational supports to ensure effective intervention delivery and sustainability. It provides a framework for optimizing workflows, training, and resource planning through simulation and forecasting tools that incorporate diverse stakeholder inputs [43].

AI tools such as ML models, agent-based modeling (ABM) and system dynamics modeling (SDM) could be used with community input to simulate or estimate needs such as resources, staffing, fidelity, or community engagement necessary for achieving the desired outcomes [44, 45]. This allows for informed decision-making, intervention planning [46], and maintenance of standards throughout the implementation process [19]. AI tools could also support implementation fidelity and sustainability by analyzing multiple data sources including meeting transcripts, session recordings, and technical assistance (TA) logs detailing the type of support offered, frequency of interactions, and specific implementation challenges addressed to assess intervention fidelity and community responsiveness to the intervention. For example, NLP could identify procedural drift or flag low engagement by analyzing language use, while ML could integrate fidelity reports with demographic data to detect where implementation may falter [22]. Importantly, researchers and community members should co-specify thresholds for acceptability to enable AI models to reflect shared expectations around fidelity and performance and empirically test these thresholds and the corresponding responses. This iterative testing is critical to developing data-informed decision rules when observed variables change in a community; and determine the type of response that is warranted and appropriate thresholds. Applying this strategy in studies like the HCS would allow communities and researchers to calibrate actions based on evidence, strengthen planning, training, and mid-course corrections through AI-informed learning cycles. Moreover, AI tools could detect early warning signs of resource strain (e.g., reduced meeting participation or burnout indicators) and simulate future implementation needs under various scenarios to support sustainability planning. ABM and SDM could test how variations in coalition leadership, staffing, or funding affect implementation success over time [19], and AI-driven forecasting tools ensure local relevance and accuracy, when developed with stakeholder input.

External environment

The PRISM-Capabilities model for AI incorporates how external factors such as policies, regulations, community assets and broader socio-political factors shape the success and sustainability of CER [20, 27]. PRISM focuses on how systemic structures (e.g., laws, reimbursement systems, resource availability) affect intervention delivery and sustainability, while the Capabilities Approach examines how those same forces constrain or enable individuals' abilities to achieve desired outcomes. Together, they provide a complementary lens to assess how broader conditions impact equity and feasibility in CER. AI tools could enhance this by processing large volumes of unstructured and structured data. NLP could be used to analyze policy documents, clinical guidelines, legislative records, and media content to extract relevant shifts in regulation, reimbursement, or political sentiment that may affect intervention implementation. Geospatial mapping could help identify gaps in local infrastructure (e.g., healthcare or educational facilities), while image-recognition tools could assess geographic disparities in service delivery [47, 48]. However, all AI-generated interpretations of policy or resource data should be validated through community and expert review, particularly in contexts with contested or historically exclusionary policies. Ultimately, the value of AI lies not only in monitoring regulatory, political or economic shifts, but in ensuring such insights are interpreted collaboratively and used to design ethically and practically grounded interventions that address real-world problems.

Ethical assessment and evaluation

Researchers conducting CER must prioritize ethical AI use across all six components of the PRISM-Capabilities model to ensure inclusivity, equity, safety, data privacy, and accountability across all phases of CER. In this area, NLP could analyze large volumes of feedback (e.g., outcome data, meeting transcripts, social media sentiment

Table 2 Checklist to guide the ethical use of AI in CER

Checklist item	Example Questions
Safeguarding data confidentiality	What cybersecurity measures are in place to safeguard data?
Compliance with regulations and policies	Has IRB approval for the study been granted?
Promoting equity and fairness by including diverse communities and considering their unique characteristics	Are members of the community included on the research team? Is equity included in the research goals?
Maintaining data integrity to ensure accuracy and reliability	Are the proper measures in place to ensure data is properly handled?
Adopting and implementing privacy protocols for transparent data use	What privacy protocols have been developed to protect data and confidentiality?
Implementing robust data security measures	How is data being stored and handled?
Ensuring transparency in AI processes and outcomes	What dissemination practices will the research group use throughout?
Building trust and confidence by aligning AI applications with community expectations	How are the researchers ensuring community voices are heard?

etc.) to identify ethical concerns or outcomes that partners, participants, and community members may miss during human review. For instance, AI algorithms could detect early-stage implementation biases such as unequal access across sociodemographic groups and generate ethical impact reports to guide decision-making.

While AI tools are powerful for synthesis, they must not replace human judgment. Oversight is essential to contextualize AI outputs, especially given the complexity of behavior, cultural differences, and structural inequities across communities. Including data from diverse datasets (e.g., policies, administrative data, meeting minutes) enhances ethically grounded interpretations. In addition, researchers should use clear, well-contextualized prompts and integrate fact-checking to reduce hallucinations (i.e., inaccuracies that arise from overgeneralized or misaligned patterns in the training dataset) in AI-generated content [49, 50]. Though not eliminated entirely, hallucinations could be minimized through retrieval-augmented generation (RAG), in which AI retrieves real information from an external sources (meeting minutes, survey data, focus group discussions, etc.) while generating its answer [51, 52]. In this process, community members could be actively involved when reviewing AI-generated outputs for accuracy. Furthermore, AI tools must be deployed alongside strong data protection measures. This includes informed consent, clear explanation of AI's role, compliance with ethical and legal standards (e.g., HIPAA, GDPR), and enterprise-level safeguards like secure platforms, encryption, role-based access, and audit logs. Additional protections such as text and voice anonymization and differential privacy techniques are also crucial when working with sensitive data. Researchers should systematically evaluate the intended and unintended consequences of AI-supported decisions as they evolve over time, integrating this into real-time monitoring. Sociodemographic overlays should be used in conjunction with feedback and outcome data to identify disparities that may not be visible in raw performance metrics. Ethical safeguards must include embedded de-identification protocols, differential privacy layers, and automated audit trail systems within AI pipelines to ensure procedural justice throughout the data lifecycle.

Explicit mechanisms should also be in place to uphold transparency in AI decision-making, supported by real-time explainability features. Algorithmic bias stemming from data and representational imbalances is also a critical issue, and AI models trained on biased data may produce harmful outcomes. To mitigate this, researchers must use diverse datasets, conduct fairness audits, and implement interpretable models. Explainability tools such as SHAP or LIME could help explain how a ML model made a specific prediction, especially when the model itself is complex and not directly interpretable [37]. This could clarify how decisions are made and help stakeholders verify their logic. Participatory AI-checking ensures diverse voices, including researchers, implementers, and people with lived experience are engaged throughout CER. Finally, researchers should also support the development of open-source explainability tools and community-governed AI systems [53].

To ensure ethical and equitable CER, we propose that all stakeholders involved in CER adopt an ethical checklist guided by the six phases of the PRISM- Capabilities model for AI (Table 2). This checklist helps establish a foundation for ethical accountability, fosters trust, and promoting responsible AI integration during CER.

Machine learning/AI tools considered in the PRISM-capabilities model

Having outlined the six components of the PRISM-Capabilities model, we now turn to the specific AI and ML tools that enable real-world application. To move beyond conceptual guidance, the following section (Table 3)

Table 3 ML/AI tools considered in the PRISM-capabilities model for AI

AI/ML Tools	Data Scenarios	Strengths	Limitations and Open Problems	Verification Needs	Human-AI Collaboration Opportunities	PRISM-Capabilities Model for AI Components
Natural Language Understanding (e.g., BERT, BERTopic, GPTs)	Survey open-end questions; Interview data	Excellent language modeling, topic discovery, sentiment analysis	Hard to adapt to specialized domain	Coherence, explainability, fairness	Guided fine-tuning with expert reviews, human-designed XAI methods	Optimizing engagement Characteristics of implementers Equity assessment
LLM-based Embeddings	Text corpus; Unstructured text data; Document retrieval datasets	Strong semantic search capability	Hard to interpret embedding meaning	Embedding quality, drift detection	Human-guided cluster labeling	Characteristics of implementers Implementation and sustainability infrastructure
Natural Language Generation (e.g., GPTs, RAG)	Text generation such as summary explanation	Fast generation, transferable learning	Prone to hallucinations	Factuality checking, bias detection	Humans design quality control protocols and specialized knowledge base	Optimizing engagement Equity assessment Ethical assessment and evaluation
Predictive Modeling (e.g., Random Forests, XGBoost, GLM)	Tabular datasets; Prediction and estimation	Powerful and fast, interpretable feature attribution	Poor extrapolation to unseen scenarios; Prone to biases in data	Fairness, detection of overfitting, feature audit	Feature importance with human input	Characteristics of implementers Equity assessment Implementation and sustainability infrastructure
Bayesian Modeling	Integrating multiple, noisy data sources	Captures uncertainty, combine evidence from diverse data	Computationally intensive, sensitive to model setup and prior assumptions	Model convergence, predictive checks, model inference and interpretation	Humans contribute priors and model structure, interpret posterior distributions	Implementation and sustainability infrastructure
Causal Inference	Observational; Interventional datasets	Discover causal relationships from data	Sensitive to hidden confounders; Scaling to big systems	Sensitivity to causal assumptions, counterfactual validation	Human knowledge guides causal graph design, humans design follow-up experiments	Equity assessment Ethical assessment and evaluation
Network Analysis (e.g., GNNs, Node2Vec, Stochastic Block Models)	Network data (social or spatial graphs)	Captures relational structures and network-related features	Hard to explain; Prone to biases and shifts over time	Stability, explainability, fairness	Human-guided label correction, Human checking of communities	Optimizing engagement External environment
Temporal AI Models (e.g., LSTM)	Event sequences; Panel data	Good at long-range dependencies	Hard to interpret, prone to missing data	Sequence alignment, missing data	Human insights into sequence features	Implementation and sustainability infrastructure
AI-based Simulation-Based Modeling (e.g., Surrogate Models, intelligent agents)	Simulation-based scenario reasoning based on agents or systems models	Interpretable, informed by domain knowledge, fast	Hard to generalize under extrapolation, hard to evaluate using observed data	Robustness, error metrics for system-level performance, model validation by observation data	Human guides training domains, agent objectives, interventions, and evaluates predictions	Implementation and sustainability infrastructure Ethical assessment and evaluation

operationalizes each of the six components with concrete AI methods, typical data sources, strengths and limitations, and human-AI collaboration points. This mapping is critical for researchers seeking to apply the model in practice, particularly in studies like HCS [31].

NLP and sentiment analysis could rapidly synthesize qualitative feedback, while ML models could predict patterns and disparities, supporting equity and engagement goals. Generative AI and topic modeling could accelerate reporting and thematic analysis, but require human oversight to maintain accuracy and contextual sensitivity. Simulation and forecasting models could assist in sustainability planning, while explainable AI (XAI) methods and privacy-preserving technologies strengthen transparency, ethical oversight, and data protection. Across all applications, human-AI collaboration remains foundational, ensuring AI complements rather than replaces community knowledge and decision-making. This concrete synthesis moves beyond speculation to offer an actionable, ethically grounded roadmap for AI integration into CER that is fully aligned with the PRISM-Capabilities model for AI.

Case Study: a practical application of the PRISM-Capabilities model in the HEALing Communities Study (HCS)

HCS was the largest implementation science research effort to date to address fatal overdose deaths in the US. Guided by the PRISM framework [20, 31, 54, 55], HCS was a multisite, community-level, cluster-randomized controlled trial designed to evaluate the effectiveness of the Communities that HEAL (CTH) intervention in reducing opioid-related overdose deaths in highly affected communities [56]. The trial was guided by the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework and PRISM [20]. A total of 67 communities in Kentucky, Massachusetts, New York, and Ohio were randomly assigned to either the intervention arm ($n=34$ communities) or the wait-list control arm ($n=33$ communities), stratified by state. The study was approved by Advarra, an independent research review organization, which served as the single Institutional Review Board. Oversight was provided by a Data and Safety Monitoring Board (DSMB) chartered by the National Institute on Drug Abuse (NIDA) [31, 56, 57].

The CTH intervention unfolded in six phases (Fig. 2), emphasizing community engagement, evidence-based practices (EBPs), and data-driven decision-making by community coalitions who were aided by visualizations made available via community-specific data dashboards [58, 59]. EBPs included increased naloxone distribution, expanded access to medications for opioid use disorder (MOUD), improved MOUD linkage and retention, promotion of safer opioid prescribing and dispensing, and communication campaigns to drive demand for EBPs and reduce stigma toward MOUD and people who use drugs [60].

The HCS utilized a vast amount of data from multiple sources (Table 4). To ensure fidelity to the CTH intervention, researchers implemented rigorous monitoring protocols, including monthly assessments of EBPs delivered in communities. In addition to qualitative data, HCS collected extensive administrative and epidemiological data.. Some study sites used advanced modeling techniques to further refine predictive capabilities, such as SDM to capture the interconnected nature of the opioid crisis and intervention points to inform the deployment of EBPs with community coalitions [61]. This approach allowed for a more holistic view of the system-wide impacts of implemented EBPs. The New York sites also utilized ABM to simulate individual behaviors and interactions within the community to predict how much EBPs needed to increase to achieve the study outcomes [62]. The integration of these diverse data sources and analytical methods created a robust framework for evaluating the effectiveness of the CTH intervention. By combining qualitative insights, quantitative metrics, and advanced modeling techniques, HCS was able to provide a comprehensive assessment of community-level efforts to implement EBPs and reduce opioid overdose deaths.

Potential empirical validation and scenarios for retrospective and real-world applications of the PRISM-capabilities model for AI using HCS

While the PRISM-Capabilities Model for AI was not used during HCS implementation, we offer concrete operationalization of the six components through retrospective and real-time applications of AI. Table 5 details how AI tools could be used to retrospectively analyze existing HCS data and simulates real-time utility in CER to validate the model. Table 1 also maps the six model components to corresponding AI tools and analytic objectives (e.g., sentiment shifts, engagement trajectories, policy simulations). Together, these tables illustrate the model's empirical utility and transition it from theoretical abstraction to a data-driven implementation roadmap.

Retrospective and real-time validation using HCS data

To operationalize the PRISM-Capabilities model for AI in dynamic, CER, we are conducting a multi-pronged, post-hoc analysis using HCS data to evaluate the practical utility of the PRISM-Capabilities model for AI and empirically test each of its six components. We describe potential retrospective analyses and real-time AI tools across all six components to enable responsive implementation, ethical oversight, and iterative adaptation in CER below. Our aim is to assess whether state-of-the-art

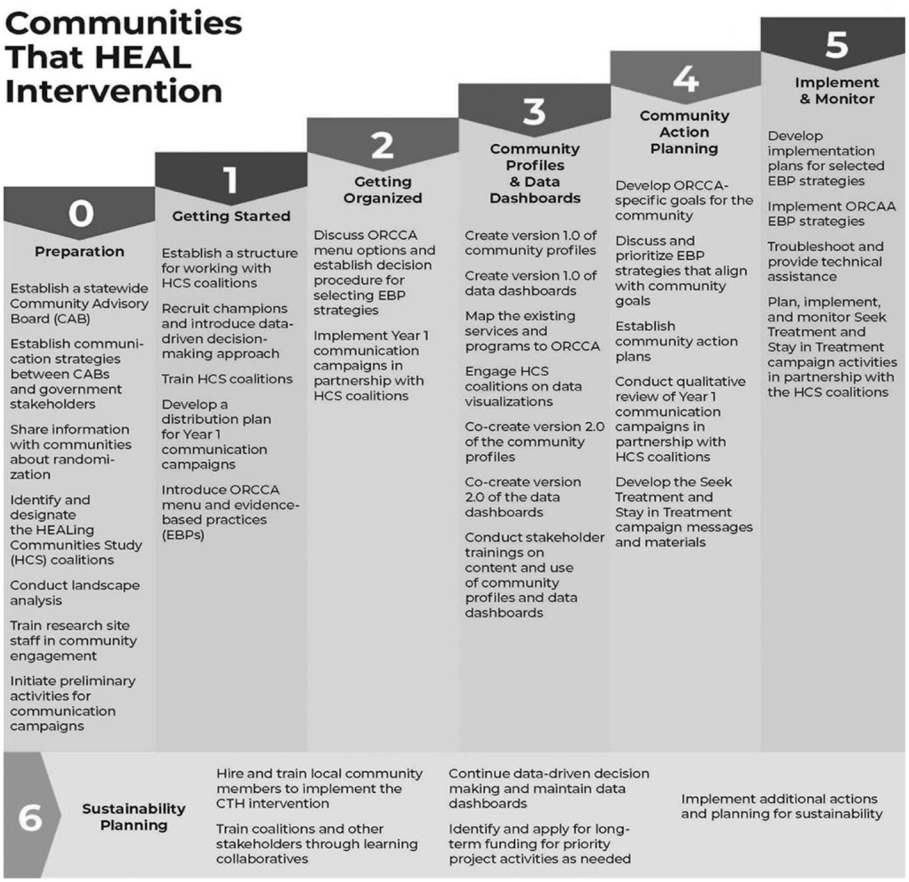

Fig. 2 The phased approach for implementation of the Communities that HEAL (CTH) intervention of the HCS to reduce fatal overdose (Martinez, L.S., et al., *Community engagement to implement evidence-based practices in the HEALing communities study.* Drug and alcohol dependence, 2020. **217**: p. 108,326.)

AI techniques can replicate the barriers, facilitators, and outcomes observed during the original HCS implementation [69]. These analyses leverage both structured and unstructured data, collected from 16 New York State communities (Table 4).

Optimizing engagement

Retrospective

As described in Table 5, NLP methods such as BERTopic and Latent Dirichlet Allocation (LDA) could be applied to coalition transcripts and interviews to uncover evolving themes when engaging community service providers for the deployment of EPBs. Sentiment analysis tools (e.g., Vader, TextBlob, RoBERTa-based models) can track emotional tone related to stigma, optimism, and resistance. Sequential pattern analysis could be used to map the alignment between coalition goals and researcher priorities, cross-referenced with fidelity data and TA logs to assess community coalitions engagement trends. Furthermore, sequential pattern analysis could be used to understand the barriers and facilitators community service providers and other stakeholders faced when identifying and deploying EPBs.

Real-time

Transformer-based NLP models (e.g., RoBERTa) and real-time topic modeling could monitor ongoing transcripts, TA logs, and coalition involvement in the deployment of EBPs. These tools could be used to generate sentiment dashboards, flag disengagement, track participation equity, and detect signs of community fatigue. Automated alerts could be used to support timely facilitation adjustments and re-engagement of underrepresented stakeholders and coalitions members.

Table 4 Sources of data used in the New York State site of the HCS

Data Source	Objective
250 + coalition meeting transcripts and minutes	Captured deliberations on best strategies for EBP deployment as communities implemented the Communities That HEAL intervention
300 + in-depth interviews	Coalition members, community partners, policymakers, PWUD, PWLE
Focus groups data across 16 NY communities	Coalitions' perspectives on challenges related to the barriers to implementing the deployment of evidence-based practices and on systems work
Data on the number and type of EBPs (including stigma reduction)	Challenges and successes of EBP deployment, what communities and populations reached
Fidelity Measures	Attendee information, subcommittee checklist, coalition meeting information and feedback
Opioid-overdose Reduction Continuum of Care Approach (ORCCA) Tracker	Tracks implementation evidence-based practices
Asset Classification	Landscape analysis of participating communities
Community Engagement	Meeting data, coalition meeting information, attendee survey
Training and Technical Assistance Tracker	Tracks meetings used to educate key partners
Surveys	Survey data related to criminal justice, HCS coalition members, HCS community advisory board members, toxicology, communications campaign
Costing	Community advisory board, communications campaign, staff activity, miscellaneous activity costs
U.S. Census Bureau's Single-Race Resident Population Estimates Data and NCHS Bridged-Race Resident Population Estimates Data[a]	County-defined community population denominator, all ages
2014–2018 5-Year ACS Data[b]	Zip code-defined community population denominator, all ages
New York State Department of Health Bureau of Vital Records[c]	Number of opioid overdose deaths among HCS community residents
New York State Hospital Inpatient Billing Claims and Emergency Department Billing Claims Data	Number of nonfatal drug poisoning hospitalizations and ED visits
New York State Department of Health, Office of Health Insurance Programs, Medicaid Claims Data	Number of individuals with opioid dependence or abuse; number of individuals receiving naltrexone; individuals with OUD receiving MOUD; individuals with OUD receiving linkages to care
New York State Prescription Drug Monitoring Program[d]	Number of individuals receiving buprenorphine products that are FDA approved for treatment of OUD
New York State Emergency Medical Services Runs Data	Number of EMS events involving naloxone administration
New York State Office of Drug User Health	Number of naloxone units distributed in community
Syndromic Surveillance Data	Number of opioid-related overdoses treated in the ED
Drug Enforcement Administration Data	Number of providers with DATA 2000 waiver

[a] Centers for Disease Control and Prevention, *CDC Wonder*. https://wonder.cdc.gov/single-race-population.html

[b] Centers for Disease Control and Prevention, *CDC Wonder*. https://wonder.cdc.gov/bridged-race-population.html

[c] NYS Department of Health, *New York State County Opioid Quarterly Reports*

[d] NYS Department of Health, *PMP/I-STOP - Prescription Monitoring Program- Internet System for Tracking Over-Prescribing*

Characteristics of implementers, settings, and recipients

Retrospective

ML classifiers (e.g., Random Forests, XGBoost) could be trained on structured data from readiness assessments, staffing patterns, and coalition characteristics to predict effective EBP implementation. Clustering algorithms (e.g., k-means, DBSCAN) could be used to identify distinct community typologies that may benefit from tailored implementation support.

Real-time

ML classifiers and NLP analytics could be applied to readiness data, interviews, and implementation records could be used to detect contextual misalignments (e.g., low local capacity, cultural misalignment). These tools could be used to guide real-time adaptation by matching interventions to community strengths, flagging implementation risks, and dynamically tailoring TA logs and resource allocation.

Table 5 Empirical operationalization of the PRISM-capabilities model using retrospective and real-time AI applications in the HEALing Communities Study (HCS)

PRISM-Capabilities Component	Retrospective Application	Real-Time Application	AI Tools	Primary Data Source	Goals of AI Use
Optimizing engagement	NLP analysis of 250 + coalition meeting transcripts; topic modeling and sentiment analysis to trace engagement dynamics and correlate with fidelity metrics	Transformer-based models (e.g., RoBERTa) to generate real-time sentiment dashboards and alerts for engagement shifts	NLP (LDA, sentiment analysis, NER); sequential analysis	Meeting transcripts, interviews, TA logs	Enhance inclusivity, monitor morale, detect disengagement early, guide responsive facilitation
Characteristics of implementers, settings, and recipients	ML models predict success from contextual variables (e.g., staffing, readiness); clustering of implementation context types	Predictive tools identify low-capacity, high-need sites; NLP tracks provider feedback on intervention alignment	ML (Random Forest, XGBoost, clustering,etc.); NLP	Readiness assessments, asset maps, qualitative interviews	Stratify support, match interventions to local conditions, optimize rollout sequencing
Equity assessment and risk management	Fairness-aware AI to detect disparities in intervention access and language biases in meeting discourse	Dashboards highlight underserved subgroups or stigmatizing trends; mid-course corrections triggered	Fairness metrics (demographic parity, disparate impact); NLP bias detection	MOUD uptake, naloxone data, coalition transcripts	Institutionalize equity checks, reduce structural bias, respond to community-specific disparities
Implementation and sustainability infrastructure	AI-driven SDM and ABM simulate how fidelity, TA, and timelines impact success	Scenario-planning simulations forecast effects of decisions like delayed rollout or leadership turnover	SDM, ABM, ML forecasting	Fidelity logs, TA records, coalition timelines	Support sustainability planning, simulate intervention pathways, predict fatigue/success
External environment	NLP tracks policy/media shifts and correlates response patterns with EBP trends using temporal models	News and policy tracking inform real-time adaptation; AI generates briefings for coalition planning	NLP, temporal topic modeling, change-point detection	Policy memos, news reports, meeting minutes	Enhance external adaptability, identify implementation disruptors, generate foresight
Ethical assessment and evaluation	LLMs generate ethics summaries; fairness metrics and entity anonymization algorithms detect underrepresentation and de-identification lapses fairness metrics and entity anonymization algorithms detect underrepresentation and de-identification lapses	LLM-based snapshots surface dissatisfaction or privacy gaps; NLP flags language violations	LLMs, NLP bias detection, text anonymization	CAB records, interviews, participant feedback	Raise ethical alerts, preserve trust, enhance transparency and procedural justice

Equity assessment and risk management

Retrospective

NLP tools could be used to surface patterns of underrepresentation, exclusionary language, and implicit bias in coalition discourse. Fairness-aware ML algorithms (e.g., reweighing, adversarial debiasing) could be used to detect disparities in resource distribution and access to EBPs across demographic groups. Findings coulc be triangulated with fidelity scores and intervention outcomes and could be used to validate equity concerns raised by the community coalitions involved in CER.

Real-time

Equity dashboards integrate MOUD uptake, naloxone distribution, and demographic data could be used to reveal disparities by race, geography, or implementation wave. NLP techniques analyze meeting notes and community records could be used to identify stigmatizing narratives. When patterns of inequity are detected, implementers could initiate mid-course corrections, such as revising outreach strategies or materials to better serve research participants from marginalized groups.

Implementation and sustainability infrastructure

Retrospective

SDM and ABM could be used to simulate how staffing ratios, TA intensity, and rollout timing influence implementation fidelity and sustainability. AI could also be used to identify fidelity gaps and quality concerns by analyzing meeting transcripts and TA documentation across implementation phases.

Real-time

Real-time SDM and ABM simulations could incorporate live data streams including fidelity logs, TA logs, and implementation milestones to identify how much (e.g., the 'dose') EBPs are needed to achieve study outcomes and forecast risks such as leadership turnover or rollout delays. These projections could support adaptive workflow planning, burnout mitigation, and long-term sustainability optimization.

External environment

Retrospective

NLP could be applied to policy documents, news articles, and other public-facing sources to assess topic framing, sentiment shifts, and media narratives over time. These analyses could be linked to external discourse trends and implementation outcomes, providing insight into how policy and media shaped local decision-making.

Real-time

Temporal NLP models and change-point detection algorithms could be used to continuously monitor media coverage and policy developments. Weekly AI-generated summaries could alert coalitions to external disruptions (e.g., new legislation, health crises), helping them adapt strategies and align messaging in real-time to maintain community relevance.

Ethical assessment and evaluation

Retrospective

Large Language Models (LLMs) such as GPT-4 could be used to synthesize ethical themes emerging from coalition discussions and community feedback. Privacy-preserving NLP tools (e.g., differential privacy, semantic-preserving redaction) could be used to detect sensitive content, underreported harms, and power imbalances. These outputs could be benchmarked against manual ethical assessments to evaluate completeness and fidelity.

Real-time

LLMs and fairness metrics (e.g., demographic parity, disparate impact) could be used to analyze interviews, feedback forms, and Community Advisory Board (CAB) records in real-time. NLP-based bias and anonymization tools could be used to flag ethical risks such as group underrepresentation or procedural injustice triggering alerts for implementers to address concerns, strengthen trust, and reinforce ethical safeguards during implementation.

Methodological limitations for retrospective application

While the PRISM-Capabilities model offers a valuable framework for evaluating AI-enabled CER, retrospective application introduces inherent limitations such as causal inference, adaptive learning, and real-time decision-making. Methodological limitations in the retrospective application of the PRISM-Capabilities model for AI to HCS includes reliance on proxy variables, temporal mismatches, static modeling assumptions, and potential unmeasured confounding. To address these, we propose mitigation strategies mapped to the components of the PRISM-Capabilities model. Mitigation strategies can include triangulating data from interviews, TA logs, and coalition meeting transcripts to reconstruct site-specific timelines and align findings with community context (Characteristics of Implementers & Recipients). Proxy variables may be validated via human coding and transformer-based NLP models (e.g., BERT) for sentiment analysis, enhancing construct validity and ethical coherence (Characteristics of Implementers & Recipients; Ethical Assessment and Evaluation). Speaker diarization

tools like pyannote-audio could segment recordings by speaker, clarifying engagement patterns and decision roles (Optimizing Engagement). Linking external datasets and completing sensitivity analyses could addressing unmeasured confounding (External Environment, Equity Assessment & Risk Management). Equity gaps may be surfaced through audit trails, subgroup analysis, and community-partner-led feedback loops (Equity Assessment & Risk Management; Ethical Assessment and Evaluation). We will assess the model's effectiveness using criteria such as alignment with historical records, AI prediction accuracy (e.g., precision-recall), stakeholder validation of findings, and actionable insights that improve future implementation outcomes. Collectively, these strategies enhance transparency, support validation, and offer practical insights for improving community-engaged implementation science through responsible, retrospective AI integration.

Addressing the potential technical limitations of AI

To address core limitations of AI, the PRISM-Capabilities model for AI proposes multiple safeguards. To ensure ethical and effective AI integration in CER, the PRISM-Capabilities model addresses hallucinations, explainability, automation trade-offs, and algorithmic bias. RAG methods, paired with human review, are recommended to reduce hallucinations in tasks such as transcript summarization [63, 64]. Explainability is enhanced through SHAP and LIME, which allows researchers and community partners to visualize and examine how model predictions are influenced by specific input features [65]. To mitigate algorithmic bias, we propose dataset diversification, subgroup performance evaluation, and participatory bias audits in collaboration with community stakeholders [34, 66]. For policy-related insights, simulation models are co-developed with domain experts to ensure contextual validity [61]. Finally, the integration of advanced anonymization techniques, including context-aware NLP tools, differential privacy [67], and enterprise-level data security frameworks, can safeguard sensitive information. Collectively, these strategies ensure that AI use within the PRISM-Capabilities model remains rigorous, transparent, and aligned with community values.

Discussion

The PRISM-Capabilities Model for AI introduces a novel, empirically grounded framework for integrating AI into CER by reimagining AI not simply as a technical enhancement, but as an ethically governed, co-designed intervention. This model reconceptualizes AI implementation through a human-centered, capabilities-based lens, filling critical gaps in prevailing frameworks like PRISM [26] and RE-AIM [20]. The PRISM-Capabilities model for AI centers participatory decision-making, interpretive authority, and the ability to contest algorithmic outputs as fundamental measures of success, consistent with the Capabilities Approach [22, 27]. These freedoms are operationalized using AI techniques like NLP for community sentiment analysis, ML for predictive modeling, and topic modeling (e.g., LDA) to uncover local priorities in unstructured text. These applications are currently being deployed in retrospective analyses of the HCS intervention to validate alignment with community-defined outcomes and ethical benchmarks [31, 56]. The model's novelty lies in integrating AI ethics and participatory governance across all six components of the PRISM-Capabilities model. For example, NLP is used to assess trust and inclusivity in engagement efforts, while ML forecasts the influence of policy environments. Each use case is co-designed with stakeholders and subject to iterative human-AI deliberation cycles, reinforcing community oversight and decision-making [33, 68]. In sum, the PRISM-Capabilities Model for AI presents a replicable and actionable structure for the use of AI in CER, balancing methodological rigor with ethical responsibilities. It offers both a theoretical and empirical advancement, meeting urgent calls for AI systems that enhance rather than constrain human agency and community sovereignty in implementation science [19].

Conclusion

The PRISM-Capabilities model for AI accounts for the complexities of real-world social science problems and explicitly positions AI tools at bottlenecks faced by conventional research approaches. Central to this are human-centered principles that prioritize human-AI collaboration, allowing for the operationalization of responsible AI practices. As this marks the initial version of our framework, we acknowledge that continuous refinement and validation of the model retrospectively and prospectively using real-world CER are essential. Specifically, there is a need to examine generalizability and adaptability across diverse socio-cultural, economic, and geographic contexts. Leveraging data from existing implementation science research presents a valuable opportunity to validate the model's effectiveness. The extensive dataset generated by HCS offers an exceptional resource for achieving these goals. To further advance the model's development and application, there is a need to develop and validate standardized metrics for evaluating the model's performance in other complex public health challenges, such as opioid overdose, HIV, and chronic disease epidemics, in other diverse settings. By implementing these validation and refinement processes, the PRISM-Capabilities model for AI has the potential to significantly advance community-engaged implementation

science. As with any emerging technology, actual results and applications may differ, and ongoing evaluation is necessary to ensure that accuracy, efficacy, and ethical considerations are maintained. While the PRISM-Capabilities model for AI has significant potential to inform the use of AI in CER, further research is needed to validate and refine the model.

Authors' information

Not applicable.

Abbreviations

ABM	Agent-Based Modeling
AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
CER	Community-Engaged Research
CAB	Community Advisory Board
EBP	Evidence-Based Practice
GPT	Generative Pre-Trained Transformer
GLM	Generalized Language Model
GNN	Graph Neural Network
HCD	Human-Centered Design
HCS	HEALing Communities Study
LDA	Latent Dirichlet Allocation
LIME	Local Interpretable Model-agnostic Explanations
LLM	Large Language Model
LSTM	Long Short-Term Memory
ML	Machine Learning
MOUD	Medications For Opioid Use Disorder
NER	Named Entity Recognition
NLP	natural language processing
PRISM	Practical, Robust Implementation and Sustainability Model
RE-AIM	Reach, Effectiveness, Adoption, Implementation, Maintenance
RoBERTa	Robustly Optimized BERT Approach
SDM	Systems Dynamics Modeling
SHAP	SHapley Additive exPlanations
TA	Technical Assistance
XAI	Explainable Artificial Intelligence

Acknowledgements

We wish to acknowledge the participation of the HEALing Communities Study communities, community coalitions, community partner organizations and agencies, and Community Advisory Boards and state government officials who partnered with us on this study. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Substance Abuse and Mental Health Services Administration or the NIH HEAL Initiative. Dr. Chandler participated in this manuscript per National Institutes of Health policy in accordance with her role as a Science Officer.

Authors' contributions

All authors read and approved the final manuscript. NE: substantial contributions to the conception; design of the work; data acquisition or analysis or interpretation of data; drafted the work and substantively revised it. JD: substantial contributions to the conception; design of the work; data acquisition or analysis or interpretation of data; drafted the work and substantively revised it. TIM: substantial contributions to the conception; design of the work; data acquisition or analysis or interpretation of data; drafted the work and substantively revised it MA: drafted the work or substantively revised it. EW: Reviewed early drafts and provided substantial feedback to inform conception. LG: Reviewed early drafts and provided substantial feedback to inform conception. SW: Reviewed early drafts and provided substantial feedback to inform conception. RC: Reviewed early drafts and provided substantial feedback to inform conception. TH: Reviewed early drafts and provided substantial feedback to inform conception. VF: Reviewed early drafts and provided substantial feedback to inform conception. AC: Reviewed early drafts and provided substantial feedback to inform conception. DGE: Reviewed early drafts and provided substantial feedback to inform conception. KK: Reviewed early drafts and provided substantial feedback to inform conception. SNB: Reviewed early drafts and provided substantial feedback to inform conception. RB: Reviewed early drafts and provided substantial feedback to inform conception. SM: Reviewed early drafts and provided substantial feedback to inform conception. EA: substantial contributions to the conception;design of the work; data acquisition or analysis or interpretation of data; drafted the work and substantively revised it. MC: Reviewed early drafts and provided substantial feedback to inform conception. PD: Reviewed early drafts and provided substantial feedback to inform conception. DL: Reviewed early drafts and provided substantial feedback to inform conception. NS: Reviewed early drafts and provided substantial feedback to inform conception. DF: Reviewed early drafts and provided substantial feedback to inform conception. TH: Reviewed early drafts and provided substantial feedback to inform conception. TZ: substantial contributions to the conception; design of the work; data acquisition or analysis or interpretation of data; drafted the paper and substantively revised it.

Funding

This research was supported by the National Institutes of Health (NIH) and the Substance Abuse and Mental Health Services Administration through the NIH HEAL (Helping to End Addiction Long-term℠) Initiative under award numbers UM1DA049394, UM1DA049406, UM1DA049412, UM1DA049415, UM1DA049417 (ClinicalTrials.gov Identifier: NCT04111939).
National Institute of Health (US),UM1DA049406,Nabila El-Bassel,UM1DA049412,Nabila El-Bassel,UM1DA049415,Nabila El-Bassel,UM1DA049417,Nabila El-Bassel

Data availability

Not Applicable.

Declarations

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Social Work, Columbia University, New York City, United States. [2]Information Technology, Columbia University, New York City, United States. [3]School of Public Health, University of North Texas Health Science Center, Fort Worth, United States. [4]National Institute on Drug Abuse, North Bethesda, Maryland, United States. [5]Department of Psychiatry, Columbia University Irving Medical Center, New York City, United States. [6]Department of Epidemiology, Columbia Mailman School of Public Health, New York City, United States. [7]University of Miami, Department of Public Health Sciences, Biostatistics, Miami, Florida, United States. [8]Department of Computer Science, Barnard College, New York City, United States. [9]Department of Statistics, Columbia University, New York City, United States. [10]Albert Einstein College of Medicine, Bronx, New York City, United States. [11]School of Public Health, City University of New York, New York City, United States.

Received: 18 February 2025 Accepted: 9 July 2025
Published online: 07 August 2025

References

1. Grossmann I, et al. AI and the transformation of social science research. Science. 2023;380(6650):1108–9.
2. Farahani MS. Applications of artificial intelligence in social science issues: a case study on predicting population change. J Knowl Econ. 2024;15(1):3266–96.
3. Liu, B., Pan, S. J., & Bing, L. Sentiment Analysis in the Era of Large Language Models: A Reality Check. in Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). 2024.

4. Lin, H. and Y. Zhang, The Risks of Using Large Language Models for Text Annotation in Social Science Research. arXiv preprint arXiv:2503.22040, 2025.
5. Raub M. Bots, bias and big data: artificial intelligence, algorithmic bias and disparate impact liability in hiring practices. Ark Law Rev. 2018;71:529.
6. Parthasarathy, S. and J. Katzman, Bringing communities in, achieving AI for all. Issues in Science and Technology, 2024. 10.
7. Pasquale, F., The Second Wave of Algorithmic Accountability. 2019: The Law and Political Economy (LPE) Project
8. Bondi, E., et al. Envisioning communities: a participatory approach towards AI for social good. in Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. 2021.
9. Straub VJ, Burton JW. Participatory approaches should be used to address the ethics of social media experiments. Commun Psychol. 2025;3(1):28.
10. Ghai B, Mueller K. D-bias: a causality-based human-in-the-loop system for tackling algorithmic bias. IEEE Trans Vis Comput Graph. 2022;29(1):473–82.
11. Kuo, R.Y.L., et al., Stakeholder perspectives towards diagnostic artificial intelligence: a co-produced qualitative evidence synthesis. EClinicalMedicine, 2024. 71.
12. Rambach T, et al. Challenges and facilitation approaches for the participatory design of AI-based clinical decision support systems: protocol for a scoping review. JMIR Res Protoc. 2024;13(1): e58185.
13. Stogiannos N, et al. A multidisciplinary team and multiagency approach for AI implementation: a commentary for medical imaging and radiotherapy key stakeholders. J Med Imaging Radiat Sci. 2024;55(4): 101717.
14. Pellegrini G, Lovati C. Stakeholders' engagement for improved health outcomes: a research brief to design a tool for better communication and participation. Front Public Health. 2025;13:1536753.
15. Hsu, Y.-C., et al., Empowering local communities using artificial intelligence. Patterns, 2022. 3(3).
16. El-Bassel N, et al. Using community engagement to implement evidence-based practices for opioid use disorder: a data-driven paradigm & systems science approach. Drug Alcohol Depend. 2021;222: 108675.
17. Trinkley KE, et al. Leveraging artificial intelligence to advance implementation science: potential opportunities and cautions. Implement Sci. 2024;19(1):17.
18. Suva, M. and G. Bhatia, Artificial Intelligence in Addiction: Challenges and Opportunities. Indian J Psychol Med, 2024: p. 02537176241274148.
19. Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. Implement Sci. 2024;19(1):27.
20. Glasgow RE, et al. Re-aim planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health. 2019;7:64.
21. Robeyns, I., The capability approach in practice. Journal of political philosophy, 2006. 14(3).
22. Robeyns I. The capability approach: a theoretical survey. J Hum Dev. 2005;6(1):93–117.
23. Oetzel JG, et al. Exploring theoretical mechanisms of community-engaged research: a multilevel cross-sectional national study of structural and relational practices in community-academic partnerships. Int J Equity Health. 2022;21(1): 59.
24. Israel BA, et al. Community-based participatory research: lessons learned from the centers for children's environmental health and disease prevention research. Environ Health Perspect. 2005;113(10):1463–71.
25. Wallerstein, N., Commentary on community-based participatory research and community engaged research in health for journal of participatory research methods. Journal of Participatory Research Methods, 2020. 1(1).
26. Feldstein AC, Glasgow RE. A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. Jt Comm J Qual Patient Saf. 2008;34(4):228–43.
27. Nussbaum, M.C., Creating capabilities: The human development approach. 2011: Harvard University Press.
28. Sen, A., Commodities and Capabilities. 1999: Oxford University Press.
29. Duboz R, et al. Systems thinking in practice: participatory modeling as a foundation for integrated approaches to health. Front Vet Sci. 2018;5:303.
30. Arnold RD, Wade JP. A definition of systems thinking: a systems approach. Procedia Comput Sci. 2015;44:669–78.
31. Chandler RK, et al. Addressing opioid overdose deaths: the vision for the HEALing communities study. Drug Alcohol Depend. 2020;217: 108329.
32. Graves, M. and E. Ratti, A Capability Approach to Ethical Development and Internal Auditing of Ai Technology. Available at SSRN 4985113.
33. Floridi L, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Mind Mach. 2018;28:689–707.
34. Dignum, V., Responsible artificial intelligence: how to develop and use AI in a responsible way. Vol. 2156. 2019: Springer.
35. Senge, P., Peter Senge and the learning organization. Dimension, 1990. 14.
36. Huang TT-K, et al. Leveraging systems science and design thinking to advance implementation science: moving toward a solution-oriented paradigm. Front Public Health. 2024;12: 1368050.
37. Salih AM, et al. A perspective on explainable artificial intelligence methods: SHAP and LIME. Adv Intell Syst. 2025;7(1):2400304.
38. McGovern A, et al. Using artificial intelligence to improve real-time decision-making for high-impact weather. Bull Am Meteor Soc. 2017;98(10):2073–90.
39. Cumberlands, U.o.t. The Use of AI in Real-time Data Analysis and Decision-making. 2023; Available from: https://www.ucumberlands.edu/blog/use-ai-real-time-data-analysis-and-decision-making.
40. Jakimowicz, K. How AI & Data Science Can Foster More Equitable Distribution of Health Resources During COVID-19. 2020; Available from: https://datasmart.hks.harvard.edu/news/article/how-ai-data-science-can-foster-more-equitable-distribution-health-resources-during.
41. Takyar, A. AI for trend analysis: Use cases, benefits, technologies, architecture, implementation and development. Available from: https://www.leewayhertz.com/ai-in-trend-analysis/.
42. Rabin BA, et al. A citation analysis and scoping systematic review of the operationalization of the Practical, Robust Implementation and Sustainability Model (PRISM). Implement Sci. 2022;17(1):62.
43. Ordu M, et al. A novel healthcare resource allocation decision support tool: a forecasting-simulation-optimization approach. J Oper Res Soc. 2021;72(3):485–500.
44. Ye J, et al. Community-based participatory research and system dynamics modeling for improving retention in hypertension care. JAMA Netw Open. 2024;7(8):e2430213–e2430213.
45. Vermeer, W.H., et al., High-fidelity agent-based modeling to support prevention decision-making: An open science approach. Prevention Science, 2022: p. 1–12.
46. McCreight MS, et al. Using the practical, robust implementation and sustainability model (PRISM) to qualitatively assess multilevel contextual factors to help plan, implement, evaluate, and disseminate health services programs. Transl Behav Med. 2019;9(6):1002–11.
47. Fadiel A, et al. Utilizing geospatial artificial intelligence to map cancer disparities across health regions. Sci Rep. 2024;14(1):7693.
48. Jean N, et al. Combining satellite imagery and machine learning to predict poverty. Science. 2016;353(6301):790–4.
49. Jančařík, A. and O. Dušek. The Problem of AI Hallucination and How to Solve It. in Proceedings of The 23rd European Conference on e-Learning. 2024. Academic Conferences International.
50. Shao, A., Beyond Misinformation: A Conceptual Framework for Studying AI Hallucinations in (Science) Communication. arXiv preprint arXiv:2504.13777, 2025.
51. Chen, J., et al., Benchmarking large language models in retrieval-augmented generation. arXiv. arXiv preprint arXiv:2309.01431, 2023.
52. Li, A., et al., Mitigating Hallucinations in Large Language Models: A Comparative Study of RAG-enhanced vs. Human-Generated Medical Templates. medRxiv, 2024: p. 2024.09. 27.24314506.
53. Gupta S. Ethical issues in designing internet-based research: recommendations for good practice. J Res Pract. 2017;13(2): D1.
54. Knudsen HK, et al. Model and approach for assessing implementation context and fidelity in the HEALing communities study. Drug Alcohol Depend. 2020;217: 108330.
55. El-Bassel N, et al. Introduction to the special issue on the HEALing communities study. Drug Alcohol Depend. 2020;217: 108327.
56. Consortium, H.C.S., Community-based cluster-randomized trial to reduce opioid overdose deaths. New England Journal of Medicine, 2024. 391(11): p. 989–1001.

57. Walsh SL, et al. The HEALing (helping to end addiction long-term SM) communities study: protocol for a cluster randomized trial at the community level to reduce opioid overdose deaths through implementation of an integrated set of evidence-based practices. Drug Alcohol Depend. 2020;217: 108335.
58. Fareed N, et al. Lessons learned from developing dashboards to support decision-making for community opioid response by community stakeholders: mixed methods and multisite study. JMIR Hum Factors. 2024;11: e51525.
59. Wu E, et al. Community dashboards to support data-informed decision-making in the HEALing communities study. Drug Alcohol Depend. 2020;217: 108331.
60. Martinez LS, et al. Community engagement to implement evidence-based practices in the HEALing communities study. Drug Alcohol Depend. 2020;217: 108326.
61. Sabounchi NS, et al. Qualitative system dynamics modeling to support community planning in opioid overdose prevention. Res Soc Work Pract. 2023;33(3):282–95.
62. Cerdá M, et al. Simulating the simultaneous impact of medication for opioid use disorder and naloxone on opioid overdose death in eight New York counties. Epidemiology. 2024;35(3):418–29.
63. Ji Z, et al. Survey of hallucination in natural language generation. ACM Comput Surv. 2023;55(12):1–38.
64. Lewis P, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. Adv Neural Inf Process Syst. 2020;33:9459–74.
65. Lundberg, S.M. and S.-I. Lee, A unified approach to interpreting model predictions. Advances in neural information processing systems, 2017. 30.
66. Raji, I.D., et al. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. in Proceedings of the 2020 conference on fairness, accountability, and transparency. 2020.
67. Dwork, C. and A. Roth, The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science, 2014. 9(3–4): p. 211–407.
68. Badal K, Lee CM, Esserman LJ. Guiding principles for the responsible development of artificial intelligence tools for healthcare. Commun Med. 2023;3(1):47.
69. El-Bassel N, David JL, Aragundi E, Walters ST, Wu E, Gilbert L, Chandler R, Hunt T, Frye V, Campbell AN, Goddard-Erich DA. Artificial Intelligence and Stigma in Addiction Research: Insights From the HEALing Communities Study Coalition Meetings. J Addict Med. 2024:10–97.

Publisher's Note