Implementation Science

STUDY PROTOCOL

Open Access

AI Methods for Implementation Science (AIM-IS): developing a framework, toolkit, and reporting standard for the responsible use of AI in implementation practice and research

Guillaume Fontaine[1,2,3,4*], Susan Michie[5], Rinad S. Beidas[6], Elvin Geng[7], Christine Fahim[23,24], Byron J. Powell[8], Vivian Welch[9,10], James Thomas[11,12], Jeffery Chan[13], Samira Abbasgholizadeh-Rahimi[14,15,16], France Légaré[17,18], Janna Hastings[19,20], Sylvie D. Lambert[1,21], Justin Presseau[3,9,22], Sharon E. Straus[23,24,25], Ruopeng An[26], Ashrita Saran[27] and Natalie Taylor[13]

Abstract

Background Artificial intelligence (AI), including machine learning, natural language processing, and large language models, may support implementation practice and research in tasks such as evidence synthesis, determinant assessment, strategy selection, monitoring, adaptation, and theory development. However, these applications of AI do not form a single, uniform category. They span a continuum from practice-facing applications that support local implementation work to research- and methods-facing applications that support evidence generation and synthesis. The guidance on how to classify, evaluate, and report these uses of AI remains limited. The AI Methods for Implementation Science (AIM-IS) program aims to develop, validate, and maintain a suite of products to guide the responsible use of AI across implementation practice, implementation research, and bridging use cases.

Methods AIM-IS is a multi-phase, multi-method methodological development program. The unit of analysis is the AI-for-implementation use case: a specific AI capability supporting a defined implementation practice or research task within a workflow, decision point, and governance context. Phase 1 is a living scoping review mapping published AI use cases in implementation science, including how they are evaluated and what risks they raise. Phase 2 is a qualitative interview study with implementation researchers, practitioners, AI experts, community members, and data infrastructure and governance experts to refine use cases and identify feasibility constraints, outcome priorities, and reporting needs. Phase 3 will integrate findings from Phases 1 and 2 to develop the draft AIM-IS products, including a framework, a taxonomy of use cases, guardrails for responsible use, a practical guide, outcome domains, and reporting items. Phase 4 will use an eDelphi process and consensus meeting to refine and finalize these products. Phase 5 will conduct usability testing to improve clarity and ease of use, resulting in the finalized AIM-IS products. AIM-IS is informed by implementation science, sociotechnical systems, equity, and responsible AI frameworks, and includes a living-update approach to support ongoing refinement.

Discussion The AIM-IS program will deliver a suite of products, including a framework, toolkit and reporting standard, to support the specification, governance, evaluation, and reporting of AI in implementation science. Together,

*Correspondence:
Guillaume Fontaine
guil.fontaine@mcgill.ca
Full list of author information is available at the end of the article

these products aim to strengthen transparency, comparability, accountability, and attention to equity in how AI is used by implementation practitioners and researchers over time.

Registration Open Science Framework, March 15, 2026: https://doi.org/10.17605/OSF.IO/BX35K

Keywords Implementation research, Implementation practice, Artificial intelligence, Machine learning, Large language models, Generative AI, Reporting guideline, Methodology

Contributions to the literature

- AI is entering implementation practice and research, but the field lacks shared methods to classify, evaluate, and report such applications.
- AIM-IS treats the AI-for-implementation use case, not the AI model alone, as the key unit of analysis, anchoring each AI capability to its workflow, intended users, decision points, and governance context.
- This protocol describes a multi-phase, living process to develop a framework, toolkit, and reporting standard for responsible AI in implementation science.
- The project combines evidence synthesis, interest-holder input, consensus methods, and usability testing to improve transparency, comparability, equity, and accountability in how AI is used across implementation practice and research.

Background

Health systems and communities continue to face a persistent "know–do" gap: effective interventions, programs, and policies are often adopted slowly, implemented unevenly, and sustained inconsistently [1–3]. Implementation science addresses these problems by studying the methods and strategies that support the uptake, integration, and sustainment of evidence-based interventions in real-world settings [4–6]. Over the past two decades, the field has generated major conceptual and empirical advances, including theories, models, and frameworks to characterize implementation processes, determinants, strategies, and outcomes [7–17]. Yet, implementation practice and research remain resource-intensive and methodologically demanding, especially across complex and under-resourced settings [18–20].

Many implementation practice and research tasks require teams to interpret and act on large volumes of heterogeneous information, including research evidence, local data, workflow constraints, organizational capacity, policy requirements, and lived or living experience [21–26]. This creates a substantial burden for teams and can limit efficiency and cumulative learning across projects and settings. Recent advances in artificial intelligence (AI), including machine learning (ML), natural language processing (NLP), large language models (LLMs), as well as generative and agentic AI systems, create new opportunities to support implementation science [27]. In this protocol, AI refers to computational methods used to support implementation science-related activities and decisions, for example by extracting, synthesizing, structuring, generating, classifying, predicting, or prioritizing information relevant to the implementation and evaluation of evidence-based interventions. This is distinct from AI used to support clinical decision-making, such as diagnostic or prognostic tools [28, 29]. Within implementation science, AI may support tasks such as evidence synthesis, determinant assessment, strategy selection and tailoring, monitoring, adaptation tracking, and theory development [24, 30–33].

AI-for-implementation use cases do not form a single homogeneous category (see Fig. 1 for key terms) [34]. Rather, they span a continuum from practice-facing applications (e.g., AI systems supporting local implementation planning, tailoring, monitoring, or adaptation), to research- and methods-facing applications (e.g., AI systems supporting study design, evidence synthesis, coding/analysis, theory development, or meta-research), to bridging applications that connect operational implementation with cumulative scientific learning [24, 30, 35–39]. These use cases differ in their users, workflows, decision points, governance needs, and risk profiles. As a result, they cannot and should not be classified, evaluated, or reported in the same way [34].

At present, this area remains methodologically underdeveloped. The use cases of AI in the field are emerging rapidly, but they are inconsistently specified, variably evaluated, and often poorly reported [30, 36, 39]. Existing AI guidance has been developed mainly for clinical, preclinical, or translational applications and does not adequately address the distinctive features of implementation practice and research, including workflow dependence, context sensitivity, iterative adaptation, equity implications, and ongoing reliance on human judgement [40–45]. AI also introduces risks that are especially salient in implementation contexts, including algorithmic bias and inequitable performance, privacy and data governance challenges, overreliance, poor workflow fit, model drift, and unclear accountability for monitoring and response [46–52]. Without a shared approach to classification, evaluation, and reporting, and without minimum guardrails for responsible use, cumulative learning across AI-enabled implementation efforts will remain limited [53, 54].

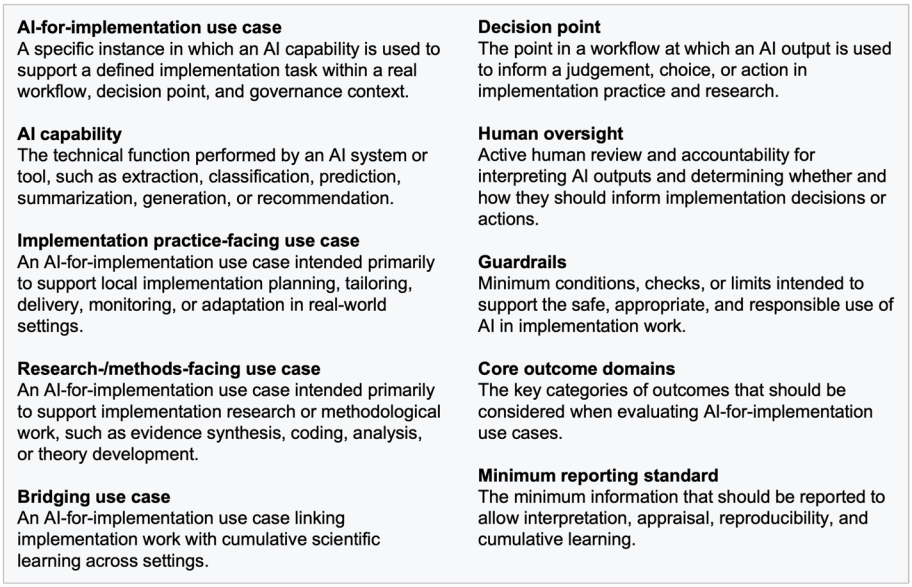

Fig. 1 Key AIM-IS terms

The AI Methods for Implementation Science (AIM-IS) program is designed to address this methodological gap. AIM-IS will develop an evidence- and consensus-informed suite of linked products to support the responsible specification, classification, governance, evaluation, and reporting of AI across the implementation practice–research continuum. The program combines living evidence synthesis, interest-holder input, structured integration, consensus methods, and usability testing to develop practical, methodologically grounded guidance in this rapidly evolving area.

Overall aim and objectives

AIM-IS aims to develop, validate, and maintain a suite of linked products, including a framework, toolkit, and reporting standard, to guide the responsible use of AI across implementation practice, implementation research, and bridging use cases. Across five phases, the program will: (1) map published use cases of AI in implementation science, including their evaluation approaches, outcomes, and risks; (2) explore interest-holder perspectives on priority use cases, feasibility, governance, outcomes, and reporting needs; (3) integrate these findings to develop the draft AIM-IS products, including a framework, toolkit, and reporting standard; (4) finalize these products through a modified eDelphi process and consensus meeting; and (5) test and refine their clarity, usability, workflow fit, and burden to produce AIM-IS v1 and support future updates.

Methods

Scope, unit of analysis and boundary rules

AIM-IS is designed to provide practical guidance for the responsible use of AI across the full implementation practice–research continuum, including bridging use cases (see Fig. 2). Because "AI in health" is a broad and rapidly expanding area, AIM-IS applies boundary rules to define what is in scope, what is out of scope, and what minimum specifications are required for an AI-for-implementation use case to be eligible for AIM-IS classification, evaluation guidance, and reporting.

In scope are the use of AI to support implementation practice and research tasks within real workflows, such as evidence synthesis, determinant assessment, strategy selection or tailoring, material development, monitoring, adaptation tracking, or theory development (see Table 1 for illustrative examples). The unit of analysis is the *AI-for-implementation use case*: a specific AI capability used to support a defined implementation task within a real workflow, decision point, and governance context, with clear outputs, and intended users. These use cases may be practice-facing, research- or methods-facing, or bridging applications that connect implementation with cumulative scientific learning. AIM-IS includes any AI approach, capability or system that materially supports such tasks, including ML, NLP, LLMs, agentic AI, and other computational methods.

AIM-IS is not aimed at developing general guidance for AI in healthcare or public health. Out of scope are uses of AI as the clinical or public health intervention itself, clinical decision support, general health information

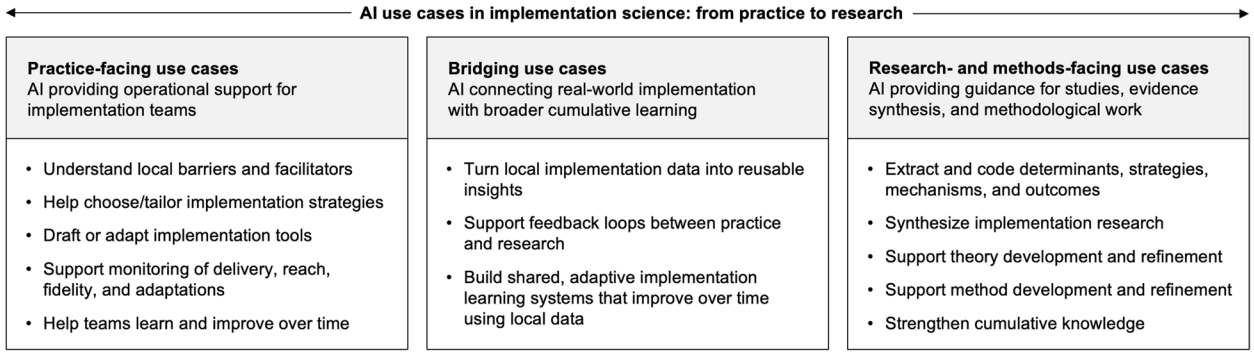

Fig. 2 The use cases of AI in implementation science span a continuum from practice-facing to research- and methods-facing

Table 1 Illustrative AI use cases across the implementation practice–research continuum

Primary orientation	Example use case	Illustrative source
Practice-facing	AI-supported geographic priority targeting to identify where implementation effort should go	Bartholomew et al. [55]
Practice-facing	Data-driven AI tailoring of implementation or facilitation strategies to local barriers, context, or practice characteristics	Moussa et al. [56]
Practice-facing	AI-assisted drafting of implementation materials	Aydin et al. [31]
Practice-facing	AI-enabled implementation monitoring using routine data sources such as transcripts, audio, communication logs, or EHR audit logs	Chen et al. [57]
Bridging	Living implementation support platform powered by different AI technologies combining evidence extraction, contextual analysis, and decision support across multiple KTA steps	Chan et al. [25]
Bridging	Learning-system style use of implementation-generated data and AI to support ongoing adaptation and cumulative improvement over time	Despraz et al. [58]
Research-/methods-facing	AI-assisted implementation evidence synthesis and annotation of implementation-relevant intervention content	Michie et al. [33]
Research-/methods-facing	AI-assisted qualitative coding for implementation needs assessment or determinant analysis	Chin-Purcell et al. [59]
Research-/methods-facing	AI-assisted rapid synthesis of implementation consultation or qualitative data to support implementation science methods development	Swanson et al. [38]

technology implementation, generic productivity uses, and AI applications without a clear implementation science-related function, workflow, or decision point.

Study team and Global Reference Group

The AIM-IS program is led by an interdisciplinary, international team with expertise in implementation science (GF, SM, RSB, EG, CF, BJP, FL, SDL, JP, SES, NT), evidence synthesis and reporting guidance (VW, JT, AS), and AI and data science (JC, SAR, JH, RA). Collectively, the team brings experience in methodological development, consensus processes, interest-holder engagement, and the translation of complex guidance into practical tools. To strengthen global relevance, applicability, and equity across settings, we will establish a Global Reference Group to provide structured feedback at key milestones. Membership will be purposively selected to include perspectives across regions and roles, drawing on the team's networks and partnerships, including engagement with WHO and *The Lancet Commission on Evidence-Based Implementation in Global Health*.

Design overview

AIM-IS is a multi-phase, multi-method methodological development program, informed by established guidance for the development of health research reporting guidelines and other consensus-based methodological guidance [60–62], to produce a linked suite of products across the implementation practice–research continuum: (i) a framework, (ii) a toolkit, and (iii) a reporting standard. The AIM-IS program is structured around five phases that move from evidence mapping and interest-holder input to integration, consensus-based finalization, and usability-focused productization (Fig. 3).

Guiding frameworks

AIM-IS is guided by complementary frameworks that (i) structure the implementation work being supported, (ii) foreground the sociotechnical, equity and governance

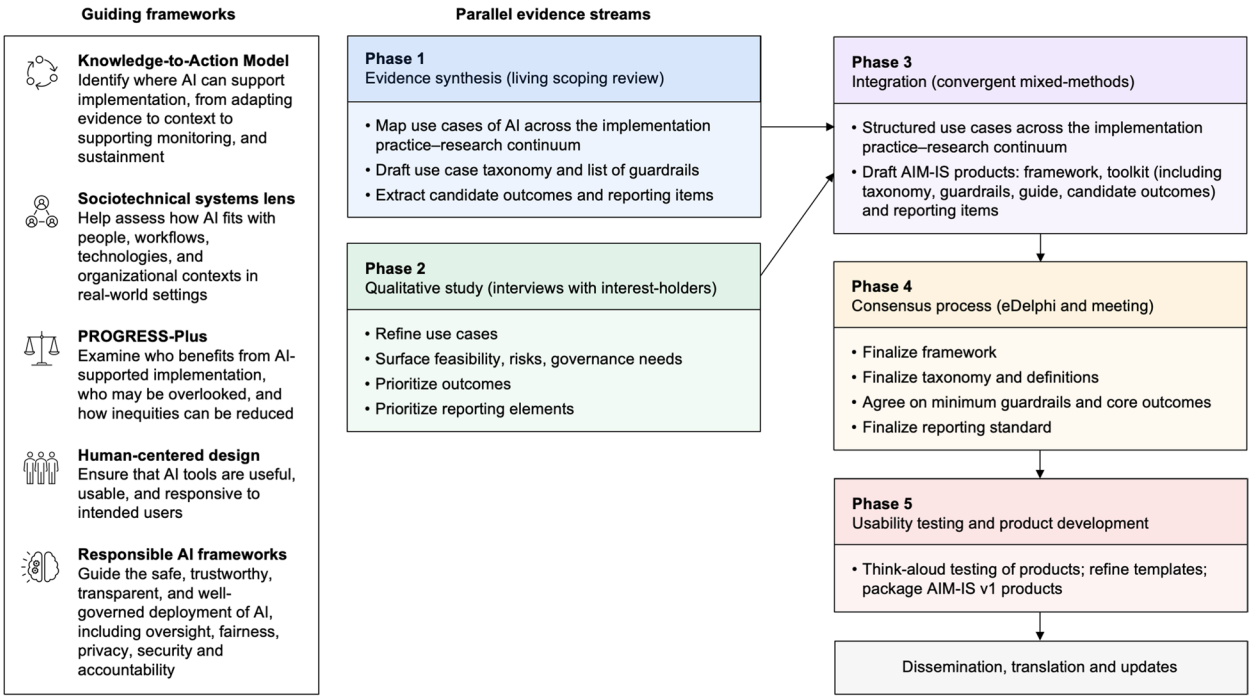

Fig. 3 The AIM-IS multi-phase development program and guiding frameworks

implications of AI-enabled methods, and (iii) ensure that outputs are usable and reportable. First, we will use an implementation science process model, the Knowledge-to-Action (KTA) model [63], to identify where AI may support activities across evidence creation, evidence synthesis and the implementation cycle. Because AI-enabled methods are embedded within workflows and organizations, data collection, analysis, and interpretation will be informed by a Sociotechnical Systems (STS) perspective [64]. STS emphasizes the interactions among people, tasks, technologies, organizational structures, processes, and the external environment, and is well suited to identifying risks that arise when technical performance does not align with the context. This perspective will inform how use cases are specified, what conditions are required for safe use, and how guardrails are operationalized.

We will apply Place, Race/ethnicity, Occupation, Gender, Religion, Education, Socioeconomic status, and Social capital (PROGRESS-Plus) [65] as an equity lens to identify differential feasibility, benefits, and harms across populations and contexts. It will guide extraction of equity-relevant outcomes and reporting elements and ensure AIM-IS considers who may benefit, who may be harmed, and under what conditions inequitable performance may occur. Where evidence is limited, we will identify uncertainty and treat equity-related unknowns as priorities for future evaluation and monitoring [66].

AIM-IS will incorporate principles from human-centered design and human factors [67–69] during product development and usability testing. This includes iteratively refining language, workflow, and burden; testing comprehension and task completion using real scenarios; and prioritizing changes based on frequency, severity, and equity and safety implications. The aim is to produce guidance that is both conceptually rigorous and practically feasible under real-world time and resource constraints.

Finally, AIM-IS will be guided by the NIST AI Risk Management Framework (AI RMF 1.0) [70] and the FUTURE-AI guideline for trustworthy and deployable AI [71]. These will inform governance, risk assessment, bias and privacy controls, documentation, human oversight, and healthcare-specific best practices. AIM-IS will also examine how incentives and commercialization shape design, documentation, and adoption, including issues such as vendor dependence, procurement constraints, data rights, and accountability for downstream harms. Where appropriate, we will draw from the Responsible use of AI in evidence SynthEsis (RAISE) recommendations [72].

Phase 1 — evidence synthesis (living scoping review): mapping the use of AI for implementation practice and research and drafting AIM-IS components

Design and reporting

Objective 1 will be addressed through a living scoping review designed to map and characterize how AI is used and evaluated to support implementation practice and

research. The review will follow JBI guidance for scoping reviews [73] and apply methods for living evidence syntheses [74, 75]. Reporting will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews (PRISMA-ScR) [76] and its extension for living systematic reviews (PRISMA-LSR) [77].

The review was prospectively registered on May 23, 2025 on the Open Science Framework (OSF) [78], and the procedures are described in detail in the published protocol [34], which will be adhered to unless an amendment is required. We present hereafter an abridged version of the review methods.

Eligibility criteria, information sources and search strategy

We will apply Population–Concept–Context criteria as specified in the published scoping review protocol [34]. In brief, we will include evidence describing AI-enabled approaches that support implementation practice and research tasks across healthcare, public health, community-based, and policy settings. Eligible evidence will include empirical studies of any design, applied case reports, implementation-relevant descriptions of deployed tools, relevant reviews, and selected grey literature [34]. Searches will be conducted across the databases and supplementary sources specified in the protocol, with targeted citation searching and grey literature searching to capture emerging work [34].

Study selection and data charting

Records will be managed in Covidence, with dual independent screening at title and abstract and full-text stages and consensus resolution of disagreements. Data will be charted using a piloted extraction form, with duplicate charting for included records. We will extract: (i) study and setting characteristics; (ii) the AI approach and level of deployment maturity; (iii) the implementation practice or research task supported and its position in the implementation lifecycle; (iv) inputs, outputs, intended users, and workflow integration; (v) oversight and governance features; (vi) evaluation approaches and outcome domains, including implementation, equity, and safety and harms where reported; and (vii) reported risks, failure modes, and mitigation strategies.

Synthesis and Phase 1 outputs

We will conduct a descriptive synthesis of the evidence to summarize how AI is currently being used to support implementation practice and research, how these applications are being evaluated, what outcomes are reported, and where important gaps remain. We will also conduct an interpretive synthesis across use cases to identify common assumptions, key gaps, possible mechanisms through which AI may influence implementation processes, and the conditions under which these approaches may be ineffective or harmful. This will generate priority questions and hypotheses to carry forward into Phases 2 and 3.

Phase 1 will produce evidence-informed draft components to carry forward into later phases, including: (a) a preliminary taxonomy of AI-for-implementation use cases with examples; (b) an evidence map of application areas, maturity, and evaluation approaches; (c) a preliminary set of guardrails for responsible use; and (d) a preliminary set of evaluation domains and candidate indicators. Phase 1 will also identify priority evidence gaps to address in later phases, particularly around real-world evaluation, comparative value, equity, governance, and monitoring over time.

Phase 2 — interest-holder interviews: refining priority use cases, constraints, risks, and outcome priorities

Design and reporting

Objective 2 will be addressed through a qualitative interview study conducted in parallel with Phase 1. The purpose of this phase is to examine how relevant interest-holders view the feasibility, potential benefits, and risks of using AI across implementation practice and research. It is intended to surface opportunities, constraints, and ethical or practical concerns that can be considered alongside the evidence synthesis during integration and consensus-building. The study will be reported in accordance with the Consolidated criteria for Reporting Qualitative research (COREQ) [79].

Sampling and recruitment

Participants will be drawn from five groups: (1) implementation researchers, (2) implementation practitioners, (3) community members, (4) AI experts, and (5) data infrastructure and governance experts, such as specialists in IT, privacy, security, data access, trusted research environments, and research technology. The fifth group is included because responsible AI-enabled implementation work often depends on whether health-system data can be accessed, linked, analysed, and governed within secure and policy-compliant environments.

Implementation researchers will have at least three years of active experience in the field. Implementation practitioners will have at least three years of experience applying evidence-based practices in clinical, organizational, community, or public health settings. Community members will be recruited on the basis of relevant lived, living, or community-based experience, with no minimum time requirement. AI experts will have at least three years of experience in AI development, evaluation, governance, regulation, or policy, preferably in health-related

contexts. Data infrastructure and governance experts will have at least three years of experience supporting or overseeing secure data environments, privacy, interoperability, data access, or digital research infrastructure.

We aim to conduct approximately 30–40 interviews overall, with target ranges of 10–12 implementation researchers, 6–8 implementation practitioners, 5–7 community members, 4–6 AI experts, and 4–6 data infrastructure and governance experts. Final numbers may be adjusted iteratively based on information power and the need for sufficient depth and diversity across groups [80]. Interviews will be conducted remotely via Microsoft Teams.

We will use purposive, maximum-variation sampling to recruit participants with relevant expertise while ensuring diversity in role, sector, geography, career stage, and implementation context, including practice-facing, research-facing, and bridging AI use cases where possible [81]. We will also aim to recruit participants from different countries and country income settings to capture variation in health-system infrastructure, governance, and implementation conditions. Snowball sampling will be used strategically to identify under-represented perspectives, particularly community voices and participants outside major academic centres. Recruitment will draw on professional networks, collaborating institutions, academic and practice partnerships, and community-facing channels. Sample composition will be monitored throughout and outreach adjusted as needed.

Data collection procedures

Data will be collected through semi-structured interviews lasting approximately 45 to 60 min, conducted via Microsoft Teams by trained study team members. With consent, interviews will be audio-recorded and transcribed using automated transcription, followed by review and cleaning to improve accuracy. Separate interview guides will be developed for each participant group.

For implementation researchers and practitioners, interviews will explore how AI may support or complicate tasks across KTA stages, including evidence synthesis, knowledge tool development, problem identification, contextual assessment and adaptation, barrier and enabler assessment, strategy selection and tailoring, monitoring, evaluation, and sustainment. For community members, interviews will focus on relevance, acceptability, trustworthiness, accessibility, perceived benefits and harms, and whether AI-enabled implementation approaches may affect communities differently. For AI experts and data infrastructure and governance experts, interviews will examine technical feasibility, system integration, validation, governance, scalability, privacy, interoperability, and regulatory or policy constraints.

Across all groups, interviews will address privacy, bias, transparency, accountability, trust, and differential impacts across populations, settings, countries, and resource contexts. Because the study aims to include participants from different countries and income settings, interviews will also attend to how infrastructure, governance capacity, data availability, and implementation conditions shape feasibility. Participants may be contacted for brief follow-up clarification or member checking, as specified in the protocol.

Where feasible and acceptable, we will also invite participating researchers, practitioners, or teams to share de-identified examples of AI use, such as prompts, outputs, and brief reflections on how outputs informed decisions. This optional real-world evidence stream will provide additional insight into how AI is being used in practice and where risks or misunderstandings arise.

Synthesis and Phase 2 outputs

Interview transcripts will be de-identified, stored on secure access-controlled institutional servers, and analysed in NVivo 15 using inductive thematic analysis [82]. Findings will then be interpreted through a sociotechnical systems lens [64] to examine how AI use is shaped by interactions among people, tasks, technologies, organizational processes, and governance contexts. The analysis will identify cross-cutting themes related to feasibility, workflow fit, governance, trust, and equity, as well as group-specific perspectives and areas of convergence or divergence. Unresolved tensions or contested issues will be documented and carried forward to later phases.

Phase 2 will generate stakeholder-informed outputs for Phase 3 integration and Phase 4 consensus-building, including: (a) refined priority AI-for-implementation use cases, with clearer definitions of what each use case does, who uses it, and where it applies; (b) a practical map of feasibility constraints across data, infrastructure, workforce capacity, governance, and workflow integration; (c) preliminary guardrails for responsible use; and (d) prioritized evaluation domains and outcomes.

Phase 3 — integration: draft AIM-IS framework, toolkit and reporting standard

Design and reporting

Objective 3 will be addressed through a convergent mixed-methods integration phase [83, 84] to translate what is reported in the literature and what is judged feasible, acceptable, and responsible in real-world settings into draft AIM-IS products. Integration will occur within and across use cases, examining convergence and divergence in intended function, workflow fit, governance requirements, minimum conditions for use, outcome priorities, and reporting needs, while also

identifying broader patterns relevant to classification, guardrails, evaluation, and accountability. This work will be informed by the guiding frameworks described previously, including the KTA model, a sociotechnical systems perspective, PROGRESS-Plus, and responsible AI frameworks.

Inputs, integration rules and outputs

Phase 3 will integrate two complementary sources of evidence: Phase 1, which maps AI-for-implementation use cases, their evaluation, risks, and reporting gaps; and Phase 2, which contributes refined use-case definitions, feasibility constraints, preliminary guardrails, priority outcomes, and unresolved tensions. The unit of integration will be the AI-for-implementation use case. For each use case, we will examine where the literature and interest-holder perspectives converge, complement one another, or remain in tension, with the aim of producing draft content that is evidence-informed, feasible, and explicit about uncertainty. Items that remain unresolved or substantially context-dependent will be flagged and brought forward as decision points for the Phase 4 for consensus.

Phase 3 will generate draft AIM-IS products, including a draft framework, taxonomy, guardrails checklist, guide outline, candidate outcome domains and indicators, and candidate reporting items. Figure 4 and Table 2 present the AIM-IS products, their purpose, intended users, and intended use.

Integration procedures

Phase 3 integration will be conducted by a multidisciplinary working group with expertise in implementation science, evidence synthesis, and AI/data science. Integration will use joint displays [85] as the primary method to align Phase 1 evidence with Phase 2 interest-holder findings for each AI-for-implementation use case (see Appendix 1). Joint displays are tables that will support transparent comparison across the two evidence streams and document how draft conclusions are developed regarding use-case definitions, workflow fit, governance requirements, guardrails, outcome domains, and reporting needs [85]. Two trained analysts will independently complete each joint display using a standardized template, before reconciling their interpretations through structured discussion. The team will calibrate the process on a small number of use cases to refine definitions, category boundaries, and template fields. Any unresolved disagreements will be adjudicated by a senior team member. A versioned decision log will document major refinements and the rationale for integrated decisions. To strengthen consistency and transparency, a senior reviewer will periodically review a sample of completed joint displays and decision logs to identify drift or recurring ambiguities.

Synthesis and Phase 3 outputs

Phase 3 will synthesize integrated findings into draft versions of all AIM-IS products. Cross-cutting uncertainties, contested issues, and context-sensitive elements identified during integration will be documented explicitly and carried forward as decision points for the Phase 4 consensus process. We will also maintain an audit trail linking integrated conclusions to the draft AIM-IS products,

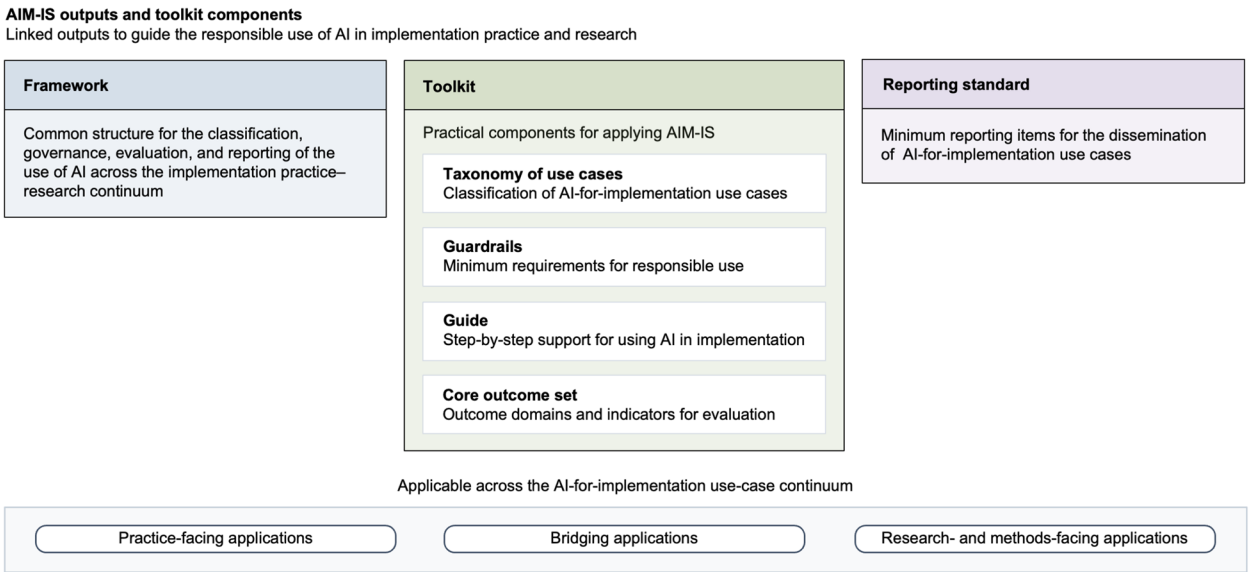

Fig. 4 Planned outputs of the AIM-IS program

Table 2 Planned AIM-IS products, purpose, primary users, and intended use

AIM-IS product		Primary purpose	Primary users	Main content	Intended use
AIM-IS framework		Provide an overall conceptual structure for AI-for-implementation practice and research	Researchers, implementers, decision-makers	Core concepts, relationships among use cases, guardrails, outcomes, and accountability	To provide users with a conceptual structure for specifying, governing, evaluating, and reporting AI-for-implementation practice and research applications
AIM-IS toolkit	*Taxonomy of AI-for-implementation use cases*	Classify the applications of AI across the implementation practice-research continuum	Researchers, practitioners, reviewers	Use-case categories, definitions, examples, boundaries	To support consistent classification and clearer comparison across studies and applications
	Guardrails checklist	Identify the minimum conditions for the responsible use of AI in implementation science	Implementation teams, organizations, governance leads	Preconditions, risks, oversight needs, do-not-use triggers where relevant	To guide decisions about whether and under what conditions an AI use case should be deployed
	Practical guide	Support the stepwise application of AIM-IS products in real workflows	Practitioners, implementation teams, applied researchers	Workflow steps, decision points, worked examples, prompts	To help users apply AIM-IS products in planning, implementation, monitoring, and ongoing adaptation
	Core outcome domains and candidate indicators	Define what should be evaluated across the AI-for-implementation use cases	Researchers, evaluators, funders	Candidate implementation, equity, safety, usability, and governance domains and indicators	To improve the consistency and relevance of evaluation across use cases
AIM-IS reporting standard		Improve the transparency and completeness of reporting of AI in implementation science	Authors, reviewers, editors	Reporting items and explanatory notes	To support transparent interpretation, reproducibility, and cumulative learning

so that Phase 4 panelists and future users can trace how draft recommendations were derived.

Phase 4 — eDelphi and consensus meeting: finalizing the AIM-IS products

Design and reporting

Objective 4 will be addressed through an eDelphi process with two online survey rounds followed by a facilitated virtual consensus meeting. The eDelphi will be designed and reported in accordance with CREDES recommendations [61], and the core outcome set component will be informed by COMET-aligned standards [86, 87]. The purpose is to move from the draft AIM-IS products developed in Phase 3 to a pre-final set of AIM-IS products that are clear, feasible, and usable across diverse settings. Deliberation will be informed by the guiding frameworks.

Panel composition and recruitment

We will recruit a purposive panel of approximately 30 to 40 participants reflecting the main stakeholder groups represented in earlier phases, including implementation researchers, implementation practitioners, community members or representatives, AI experts, and data infrastructure and governance experts. Additional expertise in ethics, privacy, legal or regulatory issues, and health-system governance will be included as appropriate.

We will seek diversity across countries, country income levels, sectors, and implementation contexts, with particular attention to settings where infrastructure, governance, workforce, or equity constraints may shape feasibility. Recruitment will draw on the AIM-IS team's networks, collaborating institutions, partner organizations, and the Global Reference Group. Panel composition will be monitored and adjusted as needed to strengthen representation across role, geography, sector, and lived or community experience. Ratings will be anonymous to other panelists to reduce power dynamics and encourage candid input.

eDelphi Round 1 procedures

Round 1 will be an online survey based on the draft AIM-IS package produced in Phase 3. Panelists will rate draft content using structured rating domains tailored to item type. Taxonomy items will be rated mainly for clarity and appropriateness, whereas guardrails, candidate outcome domains, and reporting items will be rated mainly for importance and feasibility. Where relevant, items will also be rated for clarity of wording. The survey will be administered using REDCap, with a 1–9 scale in

which higher scores indicate greater importance, feasibility, clarity, or appropriateness. Free-text fields will allow panelists to suggest revisions, identify missing items, flag unintended consequences, and note where guidance may need to differ across contexts. Round 1 will include the draft framework, taxonomy, guardrails, and guide structure, as well as the candidate outcome domains and indicators, and candidate reporting items.

eDelphi Round 2 procedures

Round 2 will re-present items that did not reach consensus in Round 1, along with items substantially revised in response to panel feedback. The aim of Round 2 will be to refine wording, reduce ambiguity, and clarify whether items should be classified as minimum, recommended, context-specific, or excluded. For each retained item, panelists will receive an anonymized summary of Round 1 results, including rating distributions and a concise summary of the main reasons for agreement or disagreement. Panelists will then re-rate items using the same scales as in Round 1. Participation and attrition across rounds will be tracked and reported.

Consensus meeting and decision finalization

After Round 2, we will hold a virtual consensus meeting to resolve a small number of high-priority unresolved items, focusing on items below the consensus threshold, items with important equity, governance, or safety implications, and items where universal versus context-specific guidance remains uncertain. Voting will be restricted to Delphi panelists. The meeting will include a brief review of the relevant Phase 3 evidence and Round 1–2 results, followed by structured discussion and, where needed, anonymous electronic voting on revised wording. Attendance, key discussions, wording changes, voting results, rationales, and relevant minority viewpoints will be documented.

Analysis and consensus thresholds

For each item, we will summarize ratings using descriptive statistics and apply pre-specified consensus rules. An item will be considered to have reached consensus for inclusion as a minimum requirement if at least 70% of panelists rate it in the upper range (7–9) and no more than 15% rate it in the lower range (1–3) on the relevant rating domain. For items rated on both importance and feasibility, both dimensions will inform classification. Items judged important but not sufficiently feasible across settings may be classified as recommended or context-specific rather than minimum requirements. Items that remain highly context-dependent, particularly across different countries, country income levels, or resource conditions, may also be classified as context-specific. Items consistently rated as low priority or infeasible will be excluded unless discussion in the consensus meeting supports retention.

Free-text comments will be analyzed using rapid content analysis to identify common proposed revisions, recurring concerns, and issues related to equity, governance, safety, or burden. We will also examine rating stability between rounds and explore whether response patterns differ meaningfully across interest-holder groups or contexts.

Phase 4 outputs

Phase 4 will produce a consensus-informed AIM-IS package for Phase 5 usability testing. Items classified as recommended or context-specific will be explicitly labelled, and unresolved issues will be retained as known tensions to inform Phase 5 and future updates.

Phase 5 — usability testing and product development: packaging the AIM-IS v1 products for real-world use

Design and purpose

Objective 5 will be addressed through formative usability testing [88, 89] of the Phase 4 consensus-informed AIM-IS beta package. The purpose of this phase is not to revisit the substantive minimum content agreed through the eDelphi and consensus meeting, but to assess whether end-users can understand and apply the AIM-IS products accurately and efficiently in realistic conditions. Phase 5 will focus on usability, clarity, workflow fit, and burden, and will inform final refinements.

AIM-IS pre-final package to be tested

The pre-final materials developed in Phase 4 to be tested will include: (a) the AIM-IS framework; (b) the taxonomy of AI-for-implementation use cases; (c) the guardrails checklist; (d) the practical guide; (e) the core outcome set; and (f) the reporting standard.

Participants and sampling

Usability testing will involve approximately 6 to 10 implementation practice or research teams, recruited purposively to capture variation in country/setting, implementation role, AI experience, and resource context. We will seek representation across health system, public health, and community-based settings, and across practice- and research-facing implementation contexts. To strengthen relevance beyond highly resourced environments, we will include settings where infrastructure, governance, workforce, or equity constraints may affect the usability of AIM-IS.

Usability testing procedures

Each participating team will be asked to apply the AIM-IS pre-final package to one realistic AI-for-implementation

scenario drawn from their own work where feasible, or to a standardized scenario where needed for comparability. Sessions will be conducted remotely using a structured facilitation guide. Participants will be asked to: (i) identify and classify the use case using the taxonomy; (ii) apply the guardrails checklist; (iii) select relevant core outcome domains and candidate indicators; and (iv) complete the reporting checklist and selected template fields for that use case.

Sessions will incorporate brief think-aloud prompts to identify where users hesitate, misunderstand terms, make assumptions, or encounter problems in the workflow. Each session will conclude with a short debrief focused on clarity, burden, feasibility, and any safety, equity, or governance concerns raised by the way the materials are structured or worded. Brief post-session ratings will also be collected to assess key usability dimensions such as clarity, confidence, and perceived burden.

Outcomes and measures

Phase 5 will focus on task completion, correctness of application, clarity of language, ease of navigation, perceived burden, workflow fit, and confidence in use. We will also examine whether any features of the materials create avoidable risks, such as ambiguity around decision points, guardrails, accountability, or reporting expectations. Data sources will include structured facilitator notes, brief user ratings, and qualitative feedback from think-aloud and debrief discussions.

Analysis and iteration plan

Usability data will be synthesized rapidly using a structured matrix capturing what worked, what created difficulty, and what requires revision. Problems will be prioritized by frequency, severity, and risk, with particular attention to issues that block use, create confusion at key decision points, increase burden unnecessarily, or raise plausible equity, privacy, governance, or overreliance concerns. Revisions will focus on wording, ordering, navigation, cross-referencing, decision flow, and template design rather than revisiting the substantive minimum content established in Phase 4. We anticipate one to two iterative refinement cycles, with all changes documented in a versioned change log linked to the usability findings.

Phase 5 outputs

Phase 5 will produce the final AIM-IS v1 products, comprising the framework, toolkit, and reporting standard. This will also include a polished framework figure, refined toolkit materials, reporting templates, and a documented change history from the Phase 4 consensus package to the finalized products.

Timeline and milestones

AIM-IS will be conducted over approximately 36 months in five phases. Phases 1 and 2 began in May 2025 and are being conducted in parallel, with completion expected by August 2026. Phase 3 (August 2026–May 2027) will integrate these findings to develop AIM-IS. Phase 4 (June–October 2027) will refine and finalize the AIM-IS components through a modified eDelphi and consensus meeting. Phase 5 (November 2027–April 2028) will use usability testing to produce AIM-IS v1.

Discussion

AI is beginning to reshape implementation science, but its importance may lie less in automating isolated tasks than in changing how implementation and research teams generate, combine, and act on knowledge across the lifecycle of implementation [24, 30, 35–39]. In this sense, AI is not simply another tool added to implementation practice and research; it may become part of the knowledge infrastructure through which teams interpret evidence, diagnose problems, select strategies, monitor progress, and learn over time. That shift creates opportunity, but also a methodological problem: without a shared way to specify what an AI application is doing, for whom, at what decision point, and under what conditions, the field risks accumulating scattered demonstrations of technical promise without building a cumulative science of use. AIM-IS is designed to respond to that problem by treating the AI-for-implementation use case, rather than the AI model alone, as the unit of analysis. This focus directs evaluation to the question that matters most in implementation science: whether an AI-enabled capability adds value to a defined implementation decision, for specific users, in a particular workflow, under real organizational and governance conditions.

This perspective also helps explain why existing AI guidance is not enough. Most available frameworks were developed for clinical AI interventions, prediction models, or early-stage decision-support tools, where the central question is often whether a model performs accurately and safely [40–42]. In implementation practice and research, however, the more consequential question is whether an AI-enabled approach improves judgement and action without distorting local priorities, obscuring accountability, or widening inequities. Put differently, sociotechnical fit is imperative: outputs need to be interpretable, actionable, governable, and acceptable in settings where implementation decisions are negotiated under resource, time and organizational constraints. For that reason, AIM-IS combines implementation science concepts with sociotechnical systems thinking, equity lenses, responsible AI frameworks, and human-centered design.

This program should also be interpreted in light of several challenges. The evidence base is likely to remain heterogeneous, incompletely reported, and skewed

toward highly resourced settings and more commercially prominent or widely published AI technologies. Terminology, AI model capabilities, and regulatory and governance expectations are also evolving quickly. Commercial incentives and market dynamics will shape both the availability of evidence and the direction of AI tool development. AIM-IS cannot eliminate these uncertainties. Its contribution is instead to make uncertainty more visible, distinguish minimum standards from context-specific guidance, and provide a structure that can be updated over time through a living update process.

AIM-IS products will be disseminated through peer-reviewed publications, a project website, and an OSF hub hosting worked examples, versioned materials, and, where feasible, non-sensitive methodological resources. Dissemination will include a coordinated online launch and a multi-channel strategy targeting researchers and implementers through targeted email outreach, blog and newsletter pieces, brief video walkthroughs, podcast appearances, social media threads, and presentations or workshops in key venues, leveraging established communication channels and team networks. User feedback and pragmatic indicators of reach and use, such as website analytics, downloads, webinar attendance, citations, and training requests, will inform future versioned updates.

Conclusion

AIM-IS will deliver a framework, toolkit, and reporting standard to guide responsible use of AI in implementation practice and research. Conducted over 36 months with an international, interdisciplinary team and a commitment to living updates, the project aims to produce guidance that is evidence-grounded, globally relevant, and built for real-world use across diverse implementation contexts. Overall, AIM-IS seeks to provide a practical and methodologically grounded approach to specifying, governing, evaluating, and reporting AI-enabled implementation practice and research. It will help the field move from scattered proof-of-concept applications toward a more cumulative, transparent, and accountable evidence base for AI in implementation science.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13012-026-01503-5.

Supplementary Material 1.

Acknowledgements
Not applicable.

Authors' contributions
GF conceived the AIM-IS program and led the development of the protocol and manuscript. SM, RSB, EG, CF, BJP, VW, JT, JC, SAR, FL, JH, SDL, JP, SES, RA, AS, and NT contributed to study design and critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Funding
GF's salary is funded partly by a Fonds de recherche du Québec - Santé (FRQ-S) Junior 1 Research Scholar Award. The AIM-IS program is supported by the Canadian Institutes of Health Research (CIHR; Phase 2, CGS-M Award), operational funding from the Methods Think Tank of the CIHR Pan-Canadian HIV/AIDS and STBBI Clinical Trials Research Network (CTN +), and start-up funding from McGill University's Faculty of Medicine and Health Sciences. SAR holds a Canada Research Chair (Tier II) in Artificial Intelligence and Advanced Digital Primary Health Care, funded by CIHR. FL holds a Canada Research Chair (Tier 1) in Shared Decision-Making and Knowledge Mobilizatio, funded by CIHR. SDL holds a Canada Research Chair (Tier II) in Self-Management, funded by CIHR. SES holds a Canada Research Chair (Tier 1) in Knowledge Translation. VW holds a CIHR Applied Public Health Chair.

Data availability
Non-sensitive methodological materials and versioned AIM-IS outputs will be made publicly available through the OSF project page. Underlying qualitative data will not be shared where confidentiality or consent constraints preclude this; aggregate summaries, analytic frameworks, and de-identified exemplar outputs will be shared where possible.

Declarations

Ethics approval and consent to participate
Phase 1 synthesizes publicly available literature and does not require research ethics approval. Phase 2 was approved by the McGill University Faculty of Medicine and Health Sciences Institutional Review Board (approval #25–05-032; July 7, 2025), and all interview participants will provide informed consent. Phases 3, 4 and 5 will be conducted in accordance with applicable ethics requirements, including amendments to the initial application as required. eDelphi participation will be voluntary and confidential, with ratings anonymous to other panelists. Usability testing participants will provide informed consent and may withdraw at any time without consequence.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Ingram School of Nursing, Faculty of Medicine and Health Sciences, McGill University, 680 Sherbrooke West, #1812, Montréal, QC H3A 2M7, Canada. [2]Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Montréal, QC, Canada. [3]Methodological and Implementation Research Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada. [4]Kirby Institute, UNSW Sydney, Sydney, NSW, Australia. [5]Centre for Behaviour Change, University College London, London, UK. [6]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. [7]Division of Infectious Diseases, School of Medicine, Washington University in St. Louis, St. Louis, MO, USA. [8]Brown School, Washington University in St. Louis, St. Louis, MO, USA. [9]School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada. [10]Bruyère Health Research Institute, Ottawa, ON, Canada. [11]EPPI Centre, Social Research Institute, University College London, London, UK. [12]Institute of Education, University College London, London, UK. [13]School of Population Health, UNSW Sydney, Sydney, NSW, Australia. [14]Department of Family Medicine, McGill University, Montreal, QC, Canada. [15]Mila – Quebec Artificial Intelligence Institute, Montréal, QC, Canada. [16]Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, QC, Canada. [17]Department of Family and Emergency Medicine, Université Laval, Québec City, QC, Canada. [18]VITAM Research Centre for Sustainable Health, Québec City, QC, Canada. [19]Idiap Research Institute, Martigny, Switzerland. [20]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [21]St. Mary's Research Centre, CIUSSS West-Montreal, Montréal, QC, Canada. [22]School of Psychology, University of Ottawa, Ottawa, ON, Canada. [23]Li Ka Shing Knowledge Institute, Unity Health Toronto, Toronto, ON, Canada. [24]Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada. [25]Department of Medicine, University of Toronto, Toronto, ON, Canada. [26]Silver School of Social Work, New York University, New York, NY, USA. [27]Global Development Network, New Delhi, India.

Received: 15 March 2026 Accepted: 2 April 2026
Published online: 14 April 2026

References

1. Glasziou P, Straus S, Brownlee S, et al. Evidence for underuse of effective medical services around the world. Lancet. 2017;390(10090):169–77. https://doi.org/10.1016/S0140-6736(16)30946-1.
2. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, Nagpal S, Saini V, Srivastava D, Chalmers K, Korenstein D. Evidence for overuse of medical services around the world. Lancet. 2017:156-168. https://doi.org/10.1016/S0140-6736(16)32585-5
3. Squires JE, Cho-Young D, Aloisio LD, et al. Inappropriate use of clinical practices in Canada: a systematic review. CMAJ. 2022;194(8):E279–96. https://doi.org/10.1503/cmaj.211416.
4. Eccles MP, Mittman BS. Welcome to implementation science. Implement Sci. 2006;1(1). https://doi.org/10.1186/1748-5908-1-1
5. Fontaine G, Taylor N, Bruneau J, et al. The urgent need for implementation science to achieve hepatitis C elimination. Lancet Gastroenterol Hepatol. 2025. https://doi.org/10.1016/S2468-1253(25)00050-0.
6. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci. 2015;10(1):53. https://doi.org/10.1186/s13012-015-0242-0.
7. Sergerie-Richard S, Goulet MH, Dumais A, Hinse C, Fontaine G. De-implementation to reduce coercive practices in mental health care. Lancet Psychiatry. 2024;11(7):498–9. https://doi.org/10.1016/S2215-0366(24)00144-5.
8. Mustanski B, Smith JD, Keiser B, Li DH, Benbow N. Supporting the growth of domestic HIV implementation research in the United States through coordination, consultation, and collaboration: how we got here and where we are headed. J Acquir Immune Defic Syndr. 2022;90(S1):S1–8. https://doi.org/10.1097/QAI.0000000000002959.
9. Oh AY, Emmons KM, Brownson RC, et al. Speeding implementation in cancer: The national cancer institute's implementation science centers in cancer control. J Natl Cancer Inst. 2023;115(2):131–8. https://doi.org/10.1093/jnci/djac198.
10. Donohue JF, Elborn JS, Lansberg P, et al. Bridging the "Know-Do" gaps in five non-communicable diseases using a common framework driven by implementation science. J Healthc Leadersh. 2023;15:103–19. https://doi.org/10.2147/JHL.S394088.
11. Sartor C, Hussian M. Mental health implementation science: integrating lived experience expertise. Lancet Psychiatry. 2024;11(5):321–2. https://doi.org/10.1016/S2215-0366(24)00092-0.
12. Fontaine G, Mooney M, Porat-Dahlerbruch J, Cahir K, Ellen M, Spinewine A, et al. Advancing the selection of implementation science theories, models, and frameworks: A scoping review and the development of the SELECT-IT meta-framework. Implement Sci. 2025. https://doi.org/10.1186/s13012-025-01436-5.
13. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. Implement Sci. 2022;17(1):75. https://doi.org/10.1186/s13012-022-01245-0.
14. Damschroder LJ, Reardon CM, Opra Widerquist MA, Lowery J. Conceptualizing outcomes for use with the consolidated framework for implementation research (CFIR): the CFIR outcomes addendum. Implement Sci. 2022;17(1):7. https://doi.org/10.1186/s13012-021-01181-5.
15. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011;38(2):65–76. https://doi.org/10.1007/s10488-010-0319-7.
16. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci. 2013;8(1):139. https://doi.org/10.1186/1748-5908-8-139.
17. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci. 2015;10(1):21. https://doi.org/10.1186/s13012-015-0209-1.
18. Rapport F, Smith J, Hutchinson K, et al. Too much theory and not enough practice? The challenge of implementation science application in healthcare practice. J Eval Clin Pract. 2022;28(6):991–1002. https://doi.org/10.1111/jep.13600.
19. Morrow A, Tyedmers E, Debono D, et al. Trials and tribulations: a qualitative exploration of researcher perspectives on navigating the challenges of health system implementation research. BMJ Open. 2025;15(1):e087926. https://doi.org/10.1136/bmjopen-2024-087926.
20. Squires JE, Graham ID, Santos WJ, Hutchinson AM, Team I. The Implementation in Context (ICON) Framework: A meta-framework of context domains, attributes and features in healthcare. Health Res Policy Syst. 2023;21(1):81. https://doi.org/10.1186/s12961-023-01028-z
21. Beidas RS, Dorsey S, Lewis CC, et al. Promises and pitfalls in implementation science from the perspective of US-based researchers: learning from a pre-mortem. Implement Sci. 2022;17(1):55. https://doi.org/10.1186/s13012-022-01226-3.
22. Lewis CC, Frank HE, Cruden G, et al. A research agenda to advance the study of implementation mechanisms. Implement Sci Commun. 2024;5(1):98. https://doi.org/10.1186/s43058-024-00633-5.
23. Waltz TJ, Powell BJ, Fernandez ME, Abadie B, Damschroder LJ. Choosing implementation strategies to address contextual barriers: diversity in recommendations and future directions. Implement Sci. 2019;14(1):42. https://doi.org/10.1186/s13012-019-0892-4.
24. Chan J, Tyedmers E, Gonzalez M, Mckay S, Liang S, Tiernan G, Mazariego C, Elias J, Torres Robles A, Taylor N. Harnessing data science and artificial intelligence to advance implementation research and practice. JBI Evid Implement. 2025.
25. Chan J, Hastings J, Tiernan G, Paterson J, Lovell N, Betbeder-Matibet L, Tung P, Pink S, Lanzeni D, Donovan T, Milat A, Jorm L, Mazariego C, Zhang C, Fontaine G, Best S, Kennedy G, Michie S, Lin F, Taylor N. Combining implementation and data sciences to accelerate evidence integration into healthcare – ImpleMATE. F1000Research. 2025;14(1184). https://doi.org/10.12688/f1000research.169999.1
26. Zhang C, Chan J, Tiernan G, McKay S, Lin F, Taylor N. Using large language model to identify and extract implementation science concepts for accelerating implementation success. 4th Swiss Implementation Science Conference. 2025.
27. Rashid AB, Kausik MDAK. AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. Hybrid Adv. 2024;7. https://doi.org/10.1016/j.hybadv.2024.100277
28. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med. 2023;388(13):1201–8. https://doi.org/10.1056/NEJMra2302038.
29. Fontaine G, Cossette S, Maheu-Cadotte MA, et al. Efficacy of adaptive e-learning for health professionals and students: a systematic review and meta-analysis. BMJ Open. 2019;9(8):e025252. https://doi.org/10.1136/bmjopen-2018-025252.
30. Trinkley KE, An R, Maw AM, Glasgow RE, Brownson RC. Leveraging artificial intelligence to advance implementation science: potential opportunities and cautions. Implement Sci. 2024;19(1):17. https://doi.org/10.1186/s13012-024-01346-y.
31. Aydin S, Karabacak M, Vlachos V, Margetis K. Large language models in patient education: a scoping review of applications in medicine. Front Med (Lausanne). 2024;11:1477898. https://doi.org/10.3389/fmed.2024.1477898.
32. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med. 2023;388(13):1233–9. https://doi.org/10.1056/NEJMsr2214184.
33. Michie S, Thomas J, Johnston M, et al. The human behaviour-change project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. Implement Sci. 2017;12(1):121. https://doi.org/10.1186/s13012-017-0641-5.
34. Fontaine G, Di Lalla O, Michie S, et al. Tracking the evolving role of artificial intelligence in implementation science: protocol for a living scoping review of applications, evaluation approaches and outcomes. F1000Research. 2026. https://doi.org/10.17605/OSF.IO/2Q5DV.
35. Goedken CC, Hook D, Hill D, Holman J, Morgan D, Brinkley K, Carey E, Branch-Elliman W. Utilizing Artificial Intelligence (AI) rapid qualitative methods for implementation of Veterans Health Administration National AI Institute initiatives. presented at: 18th Annual Conference on the Science of Dissemination and Implementation (D&I); 2025; Maryland, United States.
36. King AJ, Kahn JM. The role of data science in closing the implementation gap. Crit Care Clin. 2023;39(4):701–16. https://doi.org/10.1016/j.ccc.2023.03.005.

37. Mac Aonghusa P, Michie S. Artificial intelligence and behavioral science through the looking glass: challenges for real-world application. Ann Behav Med. 2020;54(12):942–7. https://doi.org/10.1093/abm/kaaa095.
38. Swanson KN, Stadnick NA, Aarons GA, et al. Beyond the surface of capacity building: a mixed-methods study of the core functions and forms of dissemination and implementation science consultations. Implement Sci Commun. 2025;6(1):87. https://doi.org/10.1186/s43058-025-00775-0.
39. Younas A, Reynolds SS. Leveraging artificial intelligence for expediting implementation efforts. Creat Nurs. 2024;30(2):111–7. https://doi.org/10.1177/10784535241239059.
40. Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 2020;26(9):1351–63. https://doi.org/10.1038/s41591-020-1037-7.
41. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26(9):1364–74. https://doi.org/10.1038/s41591-020-1034-x.
42. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924–33. https://doi.org/10.1038/s41591-022-01772-9.
43. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. https://doi.org/10.1136/bmj-2023-078378.
44. Kolbinger FR, Veldhuizen GP, Zhu J, Truhn D, Kather JN. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. Commun Med (Lond). 2024;4(1):71. https://doi.org/10.1038/s43856-024-00492-0.
45. Pinnock H, Barwick M, Carpenter CR, et al. Standards for reporting implementation studies (StaRI) statement. BMJ. 2017;356:i6795. https://doi.org/10.1136/bmj.i6795.
46. Morley J, Murphy L, Mishra A, Joshi I, Karpathakis K. Governing data and artificial intelligence for health care: developing an international understanding. JMIR Form Res. 2022;6(1):e31623. https://doi.org/10.2196/31623.
47. Chustecki M. Benefits and risks of AI in health care: narrative review. Interact J Med Res. 2024;13:e53616. https://doi.org/10.2196/53616.
48. Dankwa-Mullan I. Health equity and ethical considerations in using artificial intelligence in public health and medicine. Prev Chronic Dis. 2024;21:E64. https://doi.org/10.5888/pcd21.240245.
49. Goktas P, Grzybowski A. Shaping the Future of Healthcare: Ethical Clinical Challenges and Pathways to Trustworthy AI. J Clin Med. 2025;14(5). https://doi.org/10.3390/jcm14051605
50. 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. https://doi.org/10.1126/science.aax2342.
51. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11):e745–50. https://doi.org/10.1016/S2589-7500(21)00208-9.
52. Angus DC, Khera R, Lieu T, et al. AI, Health, and Health Care Today and Tomorrow: The JAMA Summit Report on Artificial Intelligence. JAMA. 2025;334(18):1650–64. https://doi.org/10.1001/jama.2025.18490.
53. Beam AL, Manrai AK, Ghassemi M. Challenges to the reproducibility of machine learning models in health care. JAMA. 2020;323(4):305–6. https://doi.org/10.1001/jama.2019.20866.
54. McDermott MBA, Wang S, Marinsek N, Ranganath R, Foschini L, Ghassemi M. Reproducibility in machine learning for health research: Still a ways to go. Sci Transl Med. 2021;13(586). https://doi.org/10.1126/scitranslmed.abb1655
55. Bartholomew TS, Tookes HE, Spencer EC, Feaster DJ. Application of machine learning algorithms for localized syringe services program policy implementation - Florida, 2017. Ann Med. 2022;54(1):2137–50. https://doi.org/10.1080/07853890.2022.2105391.
56. Moussa L, Benrimoj S, Musial K, Kocbek S, Garcia-Cardenas V. Data-driven approach for tailoring facilitation strategies to overcome implementation barriers in community pharmacy. Implement Sci. 2021;16(1):73. https://doi.org/10.1186/s13012-021-01138-8.
57. Chen J, Cutrona SL, Dharod A, et al. Electronic health record activity changes around new decision support implementation: monitoring using audit logs and topic modeling. JAMIA Open. 2025;8(4):ooaf050. https://doi.org/10.1093/jamiaopen/ooaf050.
58. Despraz J, Matusiak R, Nektarijevic S, et al. An artificial intelligence-powered learning health system to improve sepsis detection and quality of care: a before-and-after study. NPJ Digit Med. 2026;9(1):106. https://doi.org/10.1038/s41746-025-02180-2.
59. Chin-Purcell L, Chokron-Garneau H, Rosenberg-Carlson E, McGovern M. Integrating large language models into qualitative methods in implementation science: A proof-of-concept study. presented at: 18th Annual Conference on the Science of Dissemination and Implementation (D&I); 2025; Maryland, United States.
60. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med. 2010;7(2):e1000217. https://doi.org/10.1371/journal.pmed.1000217.
61. Junger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on conducting and REporting DElphi studies (CREDES) in palliative care: recommendations based on a methodological systematic review. Palliat Med. 2017;31(8):684–706. https://doi.org/10.1177/0269216317690685.
62. Gattrell WT, Logullo P, van Zuuren EJ, et al. ACCORD (ACcurate COnsensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. PLoS Med. 2024;21(1):e1004326. https://doi.org/10.1371/journal.pmed.1004326.
63. Graham ID, Logan J, Harrison MB, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof. 2006;26(1):13–24. https://doi.org/10.1002/chp.47.
64. Salwei ME, Carayon P. A sociotechnical systems framework for the application of artificial intelligence in health care delivery. J Cogn Eng Decis Mak. 2022;16(4):194–206. https://doi.org/10.1177/15553434221097357.
65. O'Neill J, Tabish H, Welch V, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. J Clin Epidemiol. 2014;67(1):56–64. https://doi.org/10.1016/j.jclinepi.2013.08.005.
66. Rid A, Wendler D. A framework for risk-benefit evaluations in biomedical research. Kennedy Inst Ethics J. 2011;21(2):141–79. https://doi.org/10.1353/ken.2011.0007.
67. International Organization for Standardization. ISO 9241 – 210 Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. 2022. https://www.iso.org/standard/77520.html
68. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. Int J Qual Health Care. 2021;33(Supplement_1):37–44. https://doi.org/10.1093/intqhc/mzaa127.
69. Dul J, Bruder R, Buckle P, et al. A strategy for human factors/ergonomics: developing the discipline and profession. Ergonomics. 2012;55(4):377–95. https://doi.org/10.1080/00140139.2012.661087.
70. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). 2023. https://doi.org/10.6028/NIST.AI.100-1
71. Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. 2025;388:e081554. https://doi.org/10.1136/bmj-2024-081554.
72. Thomas J, Flemyng E, Noel-Storr A. Responsible use of artificial intelligence in evidence synthesis (RAISE): Recommendations and guidance Open Science Framework. 2025. https://doi.org/10.17605/OSF.IO/FWAUD
73. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: Scoping reviews (2020 version). In: Aromataris E, Munn Z, eds. Joanna Briggs Institute (JBI) manual for evidence synthesis. JBI; 2020.
74. Cochrane. Guidance for the production and publication of Cochrane living systematic reviews: Cochrane Reviews in living mode. 2019. https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/201912_LSR_Revised_Guidance.pdf
75. Elliott JH, Turner T, Clavisi O, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med. 2014;11(2):e1001603. https://doi.org/10.1371/journal.pmed.1001603.
76. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467–73. https://doi.org/10.7326/M18-0850.
77. Akl EA, Khabsa J, Iannizzi C, et al. Extension of the PRISMA 2020 statement for living systematic reviews (PRISMA-LSR): checklist and explanation. BMJ. 2024;387:e079183. https://doi.org/10.1136/bmj-2024-079183.
78. Fontaine G, Di Lalla O, Michie S, Powell BJ, Welch V, Thomas J, … Taylor N. Tracking the evolving role of artificial intelligence in implementation science: protocol for a living scoping review of applications, evaluation methods, outcomes, and research gaps. Open Science Framework [Prospective registration]. 2025. https://doi.org/10.17605/OSF.IO/2Q5DV

79. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349–57. https://doi.org/10.1093/intqhc/mzm042.
80. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res. 2016;26(13):1753–60. https://doi.org/10.1177/1049732315617444.
81. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health. 2015;42(5):533–44. https://doi.org/10.1007/s10488-013-0528-y.
82. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2008;3(2):77–101. https://doi.org/10.1191/1478088706qp063oa.
83. Palinkas LA, Aarons GA, Horwitz S, Chamberlain P, Hurlburt M, Landsverk J. Mixed method designs in implementation research. Adm Policy Ment Health. 2011;38(1):44–53. https://doi.org/10.1007/s10488-010-0314-z.
84. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res. 2013;48(6 Pt 2):2134–56. https://doi.org/10.1111/1475-6773.12117.
85. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Ann Fam Med. 2015;13(6):554–61. https://doi.org/10.1370/afm.1865.
86. Kirkham JJ, Davis K, Altman DG, et al. Core outcome set-STAndards for development: the COS-STAD recommendations. PLoS Med. 2017;14(11):e1002447. https://doi.org/10.1371/journal.pmed.1002447.
87. Kirkham JJ, Gorst S, Altman DG, et al. Core outcome set-STAndards for reporting: the COS-STAR statement. PLoS Med. 2016;13(10):e1002148. https://doi.org/10.1371/journal.pmed.1002148.
88. Dopp AR, Parisi KE, Munson SA, Lyon AR. A glossary of user-centered design strategies for implementation experts. Transl Behav Med. 2019;9(6):1057–64. https://doi.org/10.1093/tbm/iby119.
89. Dopp AR, Parisi KE, Munson SA, Lyon AR. Aligning implementation and user-centered design strategies to enhance the impact of health services: results from a concept mapping study. Implement Sci Commun. 2020;1:17. https://doi.org/10.1186/s43058-020-00020-w.

Publisher's Note