Implementation Science

SYSTEMATIC REVIEW

Open Access

Artificial intelligence in implementation science and practice: a living scoping review

Guillaume Fontaine[1,2,3,4,5,6*], Olivia Di Lalla[1,6], Rachael Laritz[6], Jeremiah Durran[1], Chi Zhang[7], Jeffery Chan[7], Laura Crump[1,3], Alenda Dwiadila Matra Putra[1], Samira Abbasgholizadeh-Rahimi[8,9,10], Ruopeng An[11], Rinad S. Beidas[12], Christine Fahim[13,14], Elvin Geng[15], Ian D. Graham[4,16], Janna Hastings[17,18], Sylvie D. Lambert[1,19], France Légaré[20,21], Susan Michie[22], Byron J. Powell[23], Justin Presseau[4,16,24], Joseph Elias[7], Thomas Rudge[5], Sharon E. Straus[13,14,25], James Thomas[26,27], Vivian Welch[27] and Natalie Taylor[7]

Abstract

Background Artificial intelligence (AI) encompasses computational systems that perform tasks typically requiring human intelligence, including machine learning, generative AI, and agentic applications. The use of AI to support implementation activities is growing, but its applications and evaluation remain poorly characterised. We conducted the first cycle of a living scoping review to identify how AI is being used and evaluated across implementation science and practice.

Methods We followed JBI and Cochrane guidance and reported findings per PRISMA-ScR and PRISMA-LSR. We searched six databases through April 6, 2026, and included sources describing or evaluating AI to support implementation research or practice activities. We excluded adjacent uses such as knowledge synthesis automation. We extracted study characteristics, AI approaches, implementation tasks, evaluation methods, outcomes, and risks, and synthesised findings descriptively.

Results We identified 7,203 records and included 40 sources, 34 (85%) of which were published since 2021. Thirty-one sources described, developed, or evaluated a specific AI system, comprising 22 primary research articles, three protocols, three conference abstracts, and three other sources. The remaining nine conceptual, methodological, framework or review articles discussed potential applications of AI in implementation science. Across all 40 sources, AI was most often used or proposed to support evaluation (25/40), implementation strategy selection and tailoring (23/40), barrier and facilitator assessment (21/40), and implementation monitoring (20/40). Among the 31 sources involving a specific AI system, the use of conventional machine learning was most common (10/31), followed by generative AI (7/31), multicomponent AI systems or studies comparing AI approaches (5/31), and non-generative deep learning (3/31). Implementation-related outcomes were reported or prospectively specified in 25/31 sources, technical performance in 19/31, human-centred outcomes in 12/31, time or efficiency in 10/31, clinical or health system outcomes in 9/31, equity in 3/31, and cost or resource outcomes in 2/31. Risks were discussed in about half of sources, although assessment of harms and environmental or societal consequences was rare.

*Correspondence:
Guillaume Fontaine
guil.fontaine@mcgill.ca

Full list of author information is available at the end of the article

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Conclusions AI is in a nascent stage of supporting implementation science and practice. Reported use is narrow and methodologically underdeveloped, and much routine use is likely unpublished. The field needs rigorous, comparative, prospective, and equity-attentive research to establish whether and how AI improves implementation methods, processes, and outcomes.

Registration Open Science Framework, May 2025: https://doi.org/10.17605/OSF.IO/2Q5DV

Keywords AI, Machine learning, Natural language processing, Large language models, Implementation science, Scoping review, Evaluation, Risks, Equity

Contributions to the literature

- This review provides the first field-wide synthesis of how AI is being used or proposed to support implementation science and practice across the implementation process.
- Applications cluster around evaluation, strategy tailoring, determinant assessment, and monitoring, while adaptation, problem prioritisation, sustainment, and scale-up remain less developed.
- The evidence base is development-heavy, with limited external validation and little assessment of AI system fairness and robustness, resource and cost implications, or equity impacts.
- Although risks were frequently acknowledged, harms and workflow-related, societal, and environmental consequences were rarely assessed empirically.

Background

Artificial intelligence (AI) is increasingly used across health, public health, and related service sectors and is beginning to be applied in implementation science and practice [1, 2]. Although clinical prediction and decision support dominate attention, AI is also being applied to implementation tasks such as analysing contextual data, tailoring implementation strategies, monitoring intervention fidelity, and evaluating implementation [3–7]. AI includes a wide range of approaches, defined in Box 1 [2, 8–11]. In effect, the use of AI in implementation science appears to be outpacing the field's ability to characterise, evaluate, or govern it. Mapping this rapidly evolving landscape is a prerequisite for responsible, equitable, and methodologically robust use.

Implementation science and practice are labor- and information-intensive, and highly sensitive to context [26–31]. They involve diverse settings, partners, and data sources, including literature, guidelines, policy documents, transcripts, workflow records, field notes, audit logs, and health-system data [28–32]. Turning this fragmented information into context-responsive implementation plans still depends heavily on human interpretation and collaboration. AI could support this work by identifying patterns across determinants, strategies, mechanisms, adaptations, and outcomes; accelerating evidence extraction and mixed-methods analysis; informing strategy design and tailoring; and enabling more continuous evaluation and monitoring [28–31]. It could also clarify or reduce overlap across theories, models and frameworks, and support their refinement. In turn, these applications may strengthen cumulative learning and support uptake, scale-up, and sustainment, while helping identify populations or settings that might otherwise be overlooked.

AI can affect who is prioritised, how context is interpreted, whose judgement informs decisions, and whether implementation remains relational and locally accountable. Biased or incomplete data can misdirect priority setting or strategy selection; inaccurate generative summaries can distort contextual interpretation; automation can weaken partner oversight; and tools that perform well in research settings may fail in dynamic service workflows [33, 34]. These issues intersect directly with core implementation concerns such as feasibility, acceptability, appropriateness, fidelity, sustainment, and equity, and may also affect broader Quintuple Aim priorities such as efficiency and workforce well-being [34–40]. Additional concerns include privacy, data governance and sovereignty, limited transparency, automation bias, overreliance, model drift, workflow disruption, and environmental cost [39, 40].

The evidence base remains fragmented, leaving uncertainty about where AI is being used across the implementation process, how these applications are evaluated, and what outcomes, risks, and research gaps have been reported. This scoping review is conducted as Phase 1 of the AI Methods for Implementation Science (AIM-IS) program, which will integrate the findings of this review with stakeholder consultation, consensus methods, and usability testing to develop a framework, toolkit, and reporting standard for the responsible use of AI in implementation practice and research [41].

We conducted the first cycle of a living scoping review to (1) identify how AI is being used and evaluated to support implementation science and practice; and (2) synthesise research gaps and opportunities to advance the responsible, equitable, and methodologically robust use of AI in the field.

Box 1 Core AI concepts relevant to implementation science and practice

Artificial intelligence (AI)
Computational systems designed to perform tasks that typically require human intelligence, such as prediction, classification, recommendation, language processing, content generation, or decision support. AI includes earlier rules-based systems as well as contemporary data-driven and generative approaches [12–14].

Machine learning (ML)
A subfield of AI in which systems learn patterns from data to make predictions, classifications, recommendations, or other outputs. ML includes supervised, unsupervised, self-supervised, and reinforcement learning approaches [16].

Traditional ML
A subclass of ML that relies on structured or manually engineered features and algorithms such as decision trees, random forests, support vector machines, gradient boosting, or penalised regression [19].

Deep learning
A subfield of ML based on multilayer neural networks that learn complex patterns directly from data. Deep learning underpins many contemporary AI systems, including large language and multimodal models [21, 22].

Natural language processing (NLP)
A subfield of AI, increasingly implemented through LLMs, concerned with the computational analysis and generation of human language [24].

Foundation model
A large model pretrained on broad, diverse data that can be adapted to many downstream tasks. Foundation models are the basis of many contemporary language, image, audio, and multimodal AI systems [15].

Large language model (LLM)
A type of foundation model trained on large text corpora to process and generate language. LLMs can support tasks such as summarisation, drafting, coding, question answering, and decision support, but may also produce inaccurate or unverifiable outputs [17, 18].

Multimodal AI model
An AI model trained or adapted to process and integrate multiple data modalities, such as text, images, audio, and video, enabling cross-modal reasoning and outputs grounded in more than text [20].

Generative AI
AI systems that generate new content, such as text, images, audio, video, code, summaries, recommendations, or synthetic data, based on patterns learned from training data [23].

Agentic AI
Goal-directed AI systems that can plan and carry out sequences of actions, often using tools, retrieving information, monitoring intermediate outputs, and adapting subsequent steps [25].

Methods

Review design and reporting

This study is a living scoping review; the present paper reports the initial findings from the first full search-and-synthesis cycle of the review. Searches and evidence syntheses will be updated every six to twelve months, with the frequency of updates guided by the pace of literature growth, the rate of change in AI technologies, and available capacity and funding. Each update will be published as a separate article and versioned, with transparent documentation of newly included records, changes in classification, and emerging themes.

The review was designed and conducted in accordance with the JBI methodology for scoping reviews [42] and aligned with Cochrane guidance for living evidence syntheses [43]. The review protocol was prospectively registered and published [44]. The present manuscript is consistent with the protocol. Reporting followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) [45], and we applied the PRISMA extension for living systematic reviews (PRISMA-LSR) [46] to transparently report the living review component. Completed checklists are provided as Supplementary Files 1 and 2.

To support conceptual coherence and practical relevance, we use the Knowledge-to-Action (KTA) Model [47] as the primary organising framework for this review. KTA is a process model distinguishing the knowledge creation funnel (knowledge inquiry, knowledge synthesis, and the development of knowledge products and tools) from the action (implementation) cycle (problem identification; identification, review, and selection of knowledge; adaptation of knowledge to local context; assessment of barriers and facilitators; selection, tailoring, and implementation of interventions; monitoring of knowledge use; evaluation of outcomes; and sustainment of knowledge use) [47]. We used KTA to locate each AI application within the implementation process, identify where applications clustered, and highlight underrepresented phases [47].

Eligibility criteria

Eligible sources described or evaluated AI used to support implementation research or practice activities in health and health-related settings, including healthcare, public health, community health, and health policy, in any geographic or resource context. Implementation research referred to studies and methods used to understand, evaluate, or improve implementation; implementation practice referred to real-world activities supporting uptake, adaptation, monitoring, evaluation, sustainment, or scale-up [48]. Quality improvement studies were eligible when AI supported an implementation or change process rather than only clinical care. We included empirical studies, case reports, protocols, conference abstracts, reviews, conceptual or methodological papers, reports, preprints, grey literature, and policy documents. Reviews and conceptual sources were retained because this emerging field contains proposed applications, frameworks, and risks not yet represented in empirical evaluations; application-focused sources were analysed separately where appropriate.

We excluded AI used only as a clinical or public health intervention (e.g., diagnosis or risk prediction), a provider productivity tool (e.g., an AI scribe), generic systematic review-automation software, or a tool for generating knowledge products. We also excluded provider- or patient-facing AI whose primary purpose was care delivery rather than support for implementation work. We therefore distinguished AI that supports implementation work from AI that is itself being implemented, avoiding adjacent literatures with different outcomes and evaluation standards.

Information sources and search strategy

The search strategy was developed with a research librarian and refined with input from the AI experts (JC, JH, SAR) and the lead implementation scientist (GF) on the team. It combined terms for AI technologies and methods with terms for implementation science activities and constructs. Full search strategies, search dates, and filters were documented and reported in Supplementary File 3. We searched MEDLINE, Embase, CINAHL, PsycINFO, IEEE Xplore, and Web of Science for sources published between January 1, 2010, and April 6, 2026, in any language. We also ran targeted journal searches, handsearched relevant conference proceedings, screened reference lists, and conducted citation searches. Grey literature was eligible, but we did not conduct a dedicated grey-literature database; non-journal sources were identified through these supplementary methods. Full database and journal strategies, dates, and filters are reported in Supplementary File 3.

Study selection

All retrieved records were imported into Covidence [49] for deduplication and screening. After automated and manual deduplication, two independent reviewers (CZ, GF, JD, JE, LC, or ODL) screened titles and abstracts against the eligibility criteria. Two independent reviewers (ADMP, CZ, GF, JD, JE, LC, ODL, or RL) assessed potentially eligible full texts. Disagreements were resolved through discussion, and full-text exclusion reasons were documented. Because reviewer pairings varied, pair-specific disagreement counts and inter-rater coefficients were not prospectively retained.

Data extraction

We developed a structured data-charting form aligned with the review objectives and protocol. It was piloted on 10 included records to calibrate interpretation and refine operational definitions. GF extracted all sources, RL independently validated the full extraction, and two data science experts (JC, JD) conducted a third validation of AI-specific items.

We extracted the following domains:

- *Record characteristics:* publication year; country(ies) or region; setting (e.g., healthcare, public health, policy, community); health or social care domain; evidence type (e.g., empirical study, review, grey literature); and population/partners involved.
- *AI characteristics:* model(s), platform(s), or tool(s); model version where available; inputs and data sources; training and validation approaches; and maturity or deployment status. Each source was assigned one mutually exclusive primary AI family using the dominant architecture and output: generative AI (LLM-based); non-generative deep learning; multicomponent AI system combining two or more AI families; non-learning natural language processing and text analytics; conventional (non-deep) machine learning; symbolic or rules-based AI.
- *Tasks supported by AI:* The implementation practice or research task(s) supported by AI.
- *Evaluation methods:* Study design and evaluation approach; comparator type; the evaluation focus; and any explicit use of implementation science frameworks.
- *Outcome domains:* Outcomes included technical performance metrics, time- and cost-related measures, human factors, implementation outcomes, equity-related outcomes, and any clinical or health system outcomes reported in relation to AI-supported implementation activities.
- *Risks and adverse effects:* Reported or hypothesized risks (in the results or discussion sections) and unintended consequences related to AI use in implementation practice and science, including bias/fairness issues, privacy and data governance concerns, hallucinations or unreliable outputs, workflow disruptions and burden, automation bias/overreliance, model drift, transparency and accountability limitations, and environmental considerations.
- *Knowledge gaps:* Author-stated gaps and recommendations for future research.

For analyses of specific AI systems and their evaluation, we distinguished application-focused sources, defined as sources that described, developed, used, tested, implemented, or evaluated a specific AI system or prototype. We summarised all included sources descriptively and mapped each to one or more KTA steps. Analyses of AI families, evaluation approaches, and outcome domains focused on application-focused sources, whereas analyses of implementation tasks, frameworks, risks, and research gaps included all sources.

Use of LLMs

LLMs (OpenAI's ChatGPT 5.2, 5.4; Anthropic's Claude Sonnet 4.6 and Opus 4.7) were used to improve the readability and clarity of selected manuscript sections. They were not used for study screening decisions, data extraction, or synthesis judgments. All substantive content, classifications, and interpretations were developed and verified by the authors in accordance with the protocol, and responsibility for the final manuscript rests with the authors.

Results

Search results

As presented in Fig. 1, we identified 7,203 records through database searching and other sources. After deduplication, 6,604 records underwent title and abstract screening, and 184 full texts were reviewed in full. We excluded 144 full-text records, most often because they did not address AI for implementation science, resulting in 40 included sources.

Characteristics of included evidence sources

Table 1 summarises the included sources. Of 40 sources published between 2013 and 2026, 34 appeared since 2021 and 14 in 2025. Most originated from the United States ($n = 28$), followed by Canada ($n = 3$), Australia ($n = 2$), the United Kingdom ($n = 2$), and one each from Switzerland, Ireland, Germany, Singapore, and China. Twenty-two sources were primary research articles, comprising 12 quantitative, eight mixed-methods, and two qualitative studies. The remaining sources comprised eight conceptual, methodological, or review articles, three study protocols, three conference abstracts, two framework development articles, one comparative case study report, and one prototype development article. All were scholarly or research-oriented outputs, although several examined AI systems piloted or deployed in healthcare, public health, or quality improvement settings.

AI approaches and maturity

Figure 2 depicts the publication trend over time by AI approach. Nine sources discussed multiple potential AI applications without applying a specific AI system and were therefore not assigned an AI category. Among the 31 sources that described, developed, used, or evaluated a specific AI system or prototype, conventional non-deep machine learning was most common (10/31, 32%), followed by generative AI (7/31, 23%), multicomponent systems or studies comparing multiple AI approaches (5/31, 16%), and non-generative deep learning (3/31, 10%). Non-learning natural language processing and text analytics, symbolic or rules-based AI, and systems for which the underlying AI family was not reported were each represented in two sources (2/31, 6%).

Most specific AI systems remained at an early stage of maturity. Eighteen of the 31 sources (58%) described research prototypes or systems that were planned, under development, or undergoing methodological development. Four (13%) involved pilots, prototypes, or deployment evaluations in real-world settings, and four (13%) described systems deployed in trials or routine operational workflows. The remaining five sources (16%) used commercial AI products in research or as research aids, including generative AI tools for qualitative analysis, manuscript development, fidelity feedback, and framework development.

Use of implementation science theories, models and frameworks

Across the 40 included sources, 18 (45%) explicitly invoked one or more implementation science theory, model, or framework (TMF), drawing on 21 distinct TMFs (see Supplementary File 4). The most frequently used framework was the Consolidated Framework for Implementation Research (CFIR) [88], appearing in five studies [55, 68, 73, 77, 86]. For example, Chan et al. [55] cited CFIR alongside the Theoretical Domains Framework (TDF) [89], the Expert Recommendations for Implementing Change (ERIC) compilation [90], Proctor's implementation outcomes [91], and the Learning Health System framework [92] to inform contextual data extraction in the ImpleMATE platform's NLP and ontology pipeline. Knippen et al. [68] combined CFIR with the Model for Improvement [93] and Normalization Process Theory [94] to frame an AI-assisted reflexive thematic analysis of closing interviews. Moussa et al. [73] paired CFIR with TDF, the Tailored Implementation in Chronic Diseases (TICD) framework [95], and the Generic Implementation Framework (GIF) [96] to situate barrier categorisation prior to random forest modelling of facilitation-strategy effectiveness.

The Expert Recommendations for Implementing Change (ERIC) taxonomy [90] was also cited in four studies [55, 66, 75–77]. Rappon et al. [75] paired ERIC with the Dynamic Sustainability Framework [97] to contextualise a neural network model predicting QI sustainment, while Rinne et al. [76] used it alongside Proctor's implementation outcomes to evaluate a ChatGPT-3.5–generated synthesis of an expert-panel discussion on EHRs and implementation science.

The Reach, Effectiveness, Adoption, Implementation, Maintenance (RE-AIM) [98] framework appeared in four studies. El-Bassel et al. [60] embedded it within a PRISM-Capabilities model to structure equity-oriented AI evaluation, and Huguet et al. [64] referenced it alongside the Strategic Implementation Framework (SIF) and the

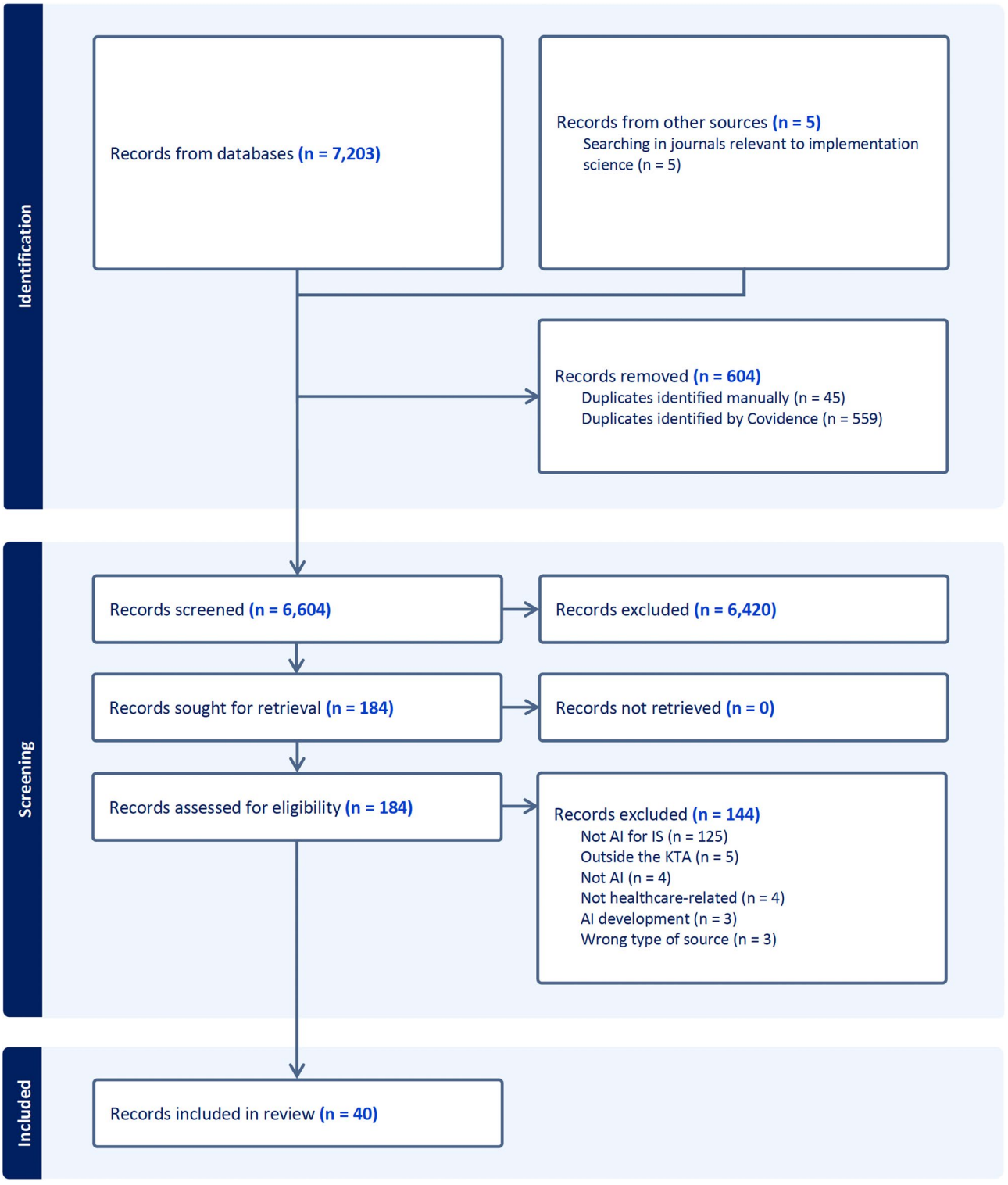


Fig. 1 PRISMA flow diagram

Table 1 Characteristics of included sources ($N=40$)

Source, country	Source type; study design (if applicable)	Health or service domain(s)	AI approach; maturity	Brief purpose or focus	Primary implementation function
Bartholomew (2022), USA [50]	Primary research article; quantitative cross-sectional	Public health; policy implementation	Conventional (non-deep) machine learning; Prototype (research)	Applied Poisson LASSO and Random Forest to identify predictors of acute HCV and prioritise locations for syringe service programme implementation.	Identify/prioritise problem or evidence-based intervention
Barwise (2026), USA [51]	Primary research article; quantitative RCT	Inpatient care; interpreter services; health equity	Conventional (non-deep) machine learning; Deployed in trial; continued after trial completion	Used an EHR-integrated predictive ML algorithm to rank inpatients with non-English language preference and trigger nurse outreach for interpreter use.	Implement and monitor implementation
Berkel (2019), USA [52]	Conceptual, methodological or review article	Prevention; community mental health	Not applicable (multiple approaches discussed)	Proposed a technology-enabled system using text and speech classification to automate implementation monitoring and feedback.	Implement and monitor implementation
Berkel (2023), USA [53]	Primary research article; quantitative	Primary care; implementation monitoring	Multiple AI approaches compared: conventional non-deep ML and non-generative deep learning; Prototype (research)	Compared TF-IDF/support-vector regression and multilingual BERT-LSTM pipelines to predict fidelity and competent adherence from English- and Spanish-language session transcripts.	Implement and monitor implementation
Brown (2013), USA [54]	Conceptual, methodological or review article	Public health; HIV prevention	Not applicable (multiple approaches discussed)	Illustrated potential uses of mobile technologies, machine learning, computational linguistics, speech recognition, social-network analysis, and simulation across HIV prevention implementation.	Cross-cutting / multiple
Chan (2025), Australia [55]	Protocol; mixed methods	Health systems; implementation infrastructure	Multicomponent AI system / multiple approaches compared; Prototype / under development	Described ImpleMATE, integrating ontologies, BERT, topic modelling, retrieval, and locally hosted LLMs for implementation evidence extraction and decision support.	Cross-cutting / multiple
Chen (2025), USA [56]	Primary research article; quantitative before-after	Health services	Conventional (non-deep) machine learning; Prototype (research)	Applied structural topic modelling to EHR audit logs to quantify workflow changes before and after tobacco-cessation decision-support implementation.	Implement and monitor implementation
Chin-Purcell (2025), USA [57]	Conference abstract; mixed methods	Implementation science methods; public health	Generative AI (LLM-based); Prototype (research)	Compared GPT-4-assisted qualitative coding with manual coding for a public-health implementation needs assessment.	Assess barriers and facilitators
Creed (2022), USA [58]	Protocol	Community mental health	Non-generative deep learning; Real-world deployment and evaluation	Described LyssnCBT, using speech recognition and transformer models to estimate CBT fidelity and generate feedback.	Implement and monitor implementation
Dabravolskaj (2025), Canada [59]	Primary research article; quantitative	School-based health promotion	Conventional (non-deep) machine learning; Prototype (research)	Used supervised k-nearest-neighbours text classifiers to distinguish intervention activities from implementation strategies in school action plans.	Sustain and scale
El-Bassel (2025), USA [60]	Framework development article	Community-engaged implementation science; public health	Not applicable (framework development)	Proposed PRISM-Capabilities to guide a portfolio of AI analyses and safeguards in community-engaged implementation science.	Cross-cutting / multiple
Goedken (2025), USA [61]	Conference abstract; mixed-methods cross-sectional	Implementation evaluation; health-system AI	Generative AI (LLM-based); Pilot in real setting	Used VA GPT(Beta) to organise and summarise interview and survey data, with human validation, to evaluate VA AI initiatives.	Evaluate impact
Golz (2022), Switzerland [62]	Primary research article; qualitative	Mental health; psychiatry	Non-learning NLP and text analytics; Prototype (research)	Used lemmatisation, n-grams, and lexicon-based sentiment analysis to examine professionals' views of implemented health information technology.	Assess barriers and facilitators
Griffen (2024), USA [63]	Primary research article; quantitative	Developmental disabilities; skill acquisition	Symbolic or rules-based AI; Prototype in real setting	Evaluated GAINS, a rules-based tablet system providing real-time prompts to support implementor fidelity and learner performance.	Implement and monitor implementation
Huguet (2024), USA [64]	Conceptual, methodological or review article	Implementation science methods	Not applicable (multiple approaches discussed)	Discussed how machine learning and NLP could support classification, prediction, tailoring, monitoring, and evaluation across implementation stages.	Cross-cutting / multiple

Table 1 (continued)

Source, country	Source type; study design (if applicable)	Health or service domain(s)	AI approach; maturity	Brief purpose or focus	Primary implementation function
Ironside (2025), USA [65]	Conference abstract; mixed-methods secondary analysis of two separate implementation trials	Behavioural health; motivational interviewing	AI family not reported; Pilot in real setting	Examined predictors and patterns of engagement in an AI-feedback trial and contrasted them descriptively with a separate expert-human-feedback trial.	Assess barriers and facilitators
Jones (2022), USA [66]	Report; comparative case-study report	HIV prevention and care; implementation research methods	AI family not reported; Under development / not stated	Described development of a social-media text classification and targeting algorithm to identify a virtual cohort and inform PrEP implementation.	Cross-cutting / multiple
King (2023), USA [67]	Conceptual, methodological or review article	Critical care	Not applicable (multiple approaches discussed)	Reviewed data-science approaches, including reinforcement learning, recommender systems, NLP, and dynamic decision support, for critical-care implementation.	Cross-cutting / multiple
Knippen (2025), USA [68]	Primary research article; mixed-methods before-after	Nutrition; gestational diabetes care	Generative AI (LLM-based); Commercial AI products used in research	Used AILYZE and ChatGPT-4 to support qualitative analysis within a PDSA-guided guideline implementation study.	Assess barriers and facilitators
Lee (2026), USA [69]	Primary research article; quantitative	Palliative care; serious-illness communication	Non-generative deep learning; Production / deployed in two pragmatic trials	Used a fine-tuned BioClinicalBERT classifier with human adjudication to identify documented goals-of-care discussions in pragmatic trials.	Evaluate impact
Li (2021), USA [70]	Primary research article; mixed-methods methodological study	Transitional care	Conventional (non-deep) machine learning; Prototype (research)	Used Random Forest and boosted trees to examine transitional-care strategy groupings and their relationships with 30-day readmission.	Evaluate impact
Lindamer (2022), USA [71]	Primary research article; mixed-methods cross-sectional	Cognitive rehabilitation	Non-learning NLP and text analytics; Prototype (research)	Used n-gram analysis of free-text surveys to complement implementation assessment of CogSMART/CCT.	Implement and monitor implementation
Mac Aonghusa (2020), Ireland [72]	Prototype development article	Public health; behavioural science	Multicomponent AI system / multiple approaches compared; Prototype (research)	Described the Human Behaviour-Change Project Knowledge System combining ontology-based representation, NLP extraction, ML prediction, and automated reasoning.	Synthesise implementation-relevant knowledge
Michie (2017), UK [5]	Protocol	Behaviour-change interventions; evidence synthesis	Multicomponent AI system / multiple approaches compared; Protocol / planned	Specified the planned Human Behaviour-Change Project Knowledge System integrating automated searching, ontology-based NLP extraction, ML prediction, and reasoning.	Synthesise implementation-relevant knowledge
Moussa (2021), Australia [73]	Primary research article; mixed methods	Community pharmacy services	Conventional (non-deep) machine learning; Prototype (research)	Used Random Forest classification to predict which facilitation strategies would resolve specific implementation barriers.	Select and tailor implementation strategies
Pearson (2020), USA [74]	Conceptual, methodological or review article	Precision health; public health	Not applicable (multiple approaches discussed)	Reviewed predictive-analytics and implementation-research approaches for precision health and learning health systems.	Cross-cutting / multiple
Rappon (2021), Canada [75]	Primary research article; quantitative	Sustainability of evidence-based interventions	Conventional (non-deep) machine learning; Prototype (research)	Used a supervised self-organising map to link sustainability determinants with outcomes and generate tailored recommendations.	Sustain and scale
Rinne (2024), USA [76]	Primary research article; mixed-methods methodological study	Implementation science methods; health services	Generative AI (LLM-based); Commercial product used in research	Used ChatGPT-3.5 to draft a manuscript from expert-panel notes and evaluated the output against human revision.	Synthesise implementation-relevant knowledge

Table 1 (continued)

Source, country	Source type; study design (if applicable)	Health or service domain(s)	AI approach; maturity	Brief purpose or focus	Primary implementation function
Rosen (2025), USA [77]	Primary research article; RCT	Mental health; PTSD psychotherapy	Symbolic or rules-based AI; Deployed in trial	Used cTAKES and validated rule-based logic to classify psychotherapy notes by evidence-based treatment type in an implementation trial.	Evaluate impact
Rosen (2026), Germany [78]	Primary research article; qualitative focus groups	Physical therapy; guideline adherence	Generative AI (LLM-based); Prototype (research)	Explored physical therapists' experiences with GAEP, a GPT-4 Turbo guideline assistant grounded in German clinical practice guidelines.	Select and tailor implementation strategies
Shafran (2026), UK [79]	Conceptual, methodological or review article	Mental health; CBT training and supervision	Not applicable (multiple approaches discussed)	Reviewed AI tools for CBT training, simulation, fidelity feedback, supervision, and governance.	Cross-cutting / multiple
Shankar (2025), Singapore [80]	Primary research article; mixed-methods before-after	Health-system quality improvement	Multicomponent AI system: conventional non-deep ML and non-learning NLP/text analytics; Production / deployed	Used a multicomponent NLP and ML pipeline to transform 126,134 patient-feedback records into topics, sentiment, and targeted quality-improvement actions.	Evaluate impact
Sibley (2025), USA [81]	Primary research article; mixed-methods randomized controlled trial	Child and adolescent mental health	Non-generative deep learning; Commercial product used in research	Evaluated Lyssn-generated motivational-interviewing fidelity feedback delivered through a clinical dashboard in a cluster RCT.	Implement and monitor implementation
Stanfield (2024), USA [82]	Primary research article; quantitative simulation	Cancer screening; preventive care	Conventional (non-deep) machine learning; Prototype (research)	Used Random Forest surrogate models to predict colorectal-cancer outcomes across implementation scenarios.	Evaluate impact
Swanson (2025), USA [83]	Primary research article; mixed-methods cross-sectional	Implementation science capacity building	Generative AI (LLM-based); Commercial product used in research	Used ChatGPT Plus to support thematic synthesis of implementation consultation notes with iterative human validation.	Evaluate impact
Trinkley (2024), USA [84]	Conceptual, methodological or review article	Implementation science; learning health systems	Not applicable (multiple approaches discussed)	Discussed opportunities and cautions for generative AI, predictive ML, NLP, computer vision, and causal methods in implementation science.	Cross-cutting / multiple
Wang (2016), USA [85]	Primary research article; quantitative	Mental health; child-welfare implementation	Conventional (non-deep) machine learning; Prototype (research)	Used semi-supervised non-negative matrix factorisation, SVM, and Naive Bayes to classify implementation-broker communications into implementation stages.	Implement and monitor implementation
Ye (2025), USA [86]	Primary research article; quantitative observational cross-sectional analysis with stakeholder review	Primary care; practice facilitation	Conventional (non-deep) machine learning; Prototype (research)	Used principal component analysis and structural equation modelling to examine relationships among practice characteristics, facilitation strategies, and outcomes.	Select and tailor implementation strategies
Younas (2024), Canada [4]	Conceptual, methodological or review article	Implementation science; nursing	Not applicable (multiple approaches discussed)	Discussed potential uses of predictive analytics, machine learning, NLP, expert systems, robotics, and adaptive tools across the implementation continuum.	Cross-cutting / multiple
Zhong (2025), China [87]	Framework development article	Clinical practice guideline implementability	Generative AI (LLM-based); Commercial product used as a research aid	Used ChatGPT-4.0 to classify and label candidate guideline-implementability constructs, followed by stakeholder refinement and expert validation.	Synthesise implementation-relevant knowledge

Abbreviations: AI, artificial intelligence; BERT, Bidirectional Encoder Representations from Transformers; CBT, cognitive behavioural therapy; CCT, compensatory cognitive training; CogSMART, Cognitive Symptom Management and Rehabilitation Training; cTAKES, clinical Text Analysis and Knowledge Extraction System; EHR, electronic health record; GAINS, Guidance, Assessment, and Information Systems; GPT, generative pre-trained transformer; HCV, hepatitis C virus; HIV, human immunodeficiency virus; LASSO, least absolute shrinkage and selection operator; LLM, large language model; LSTM, long short-term memory; ML, machine learning; NLP, natural language processing; PDSA, Plan–Do–Study–Act; PrEP, pre-exposure prophylaxis; PRISM, Practical, Robust Implementation and Sustainability Model; PTSD, post-traumatic stress disorder; RCT, randomized controlled trial; SVM, support vector machine; TF-IDF, term frequency–inverse document frequency; VA, Veterans Affairs

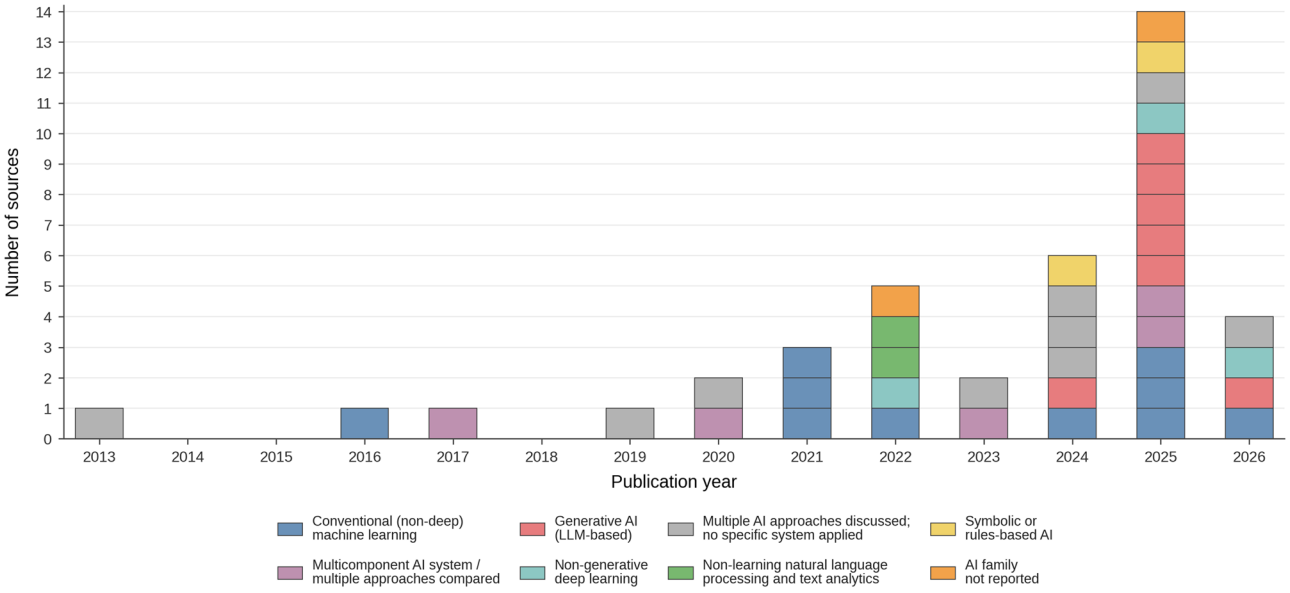


Fig. 2 Trend in AI approaches and systems over time (2013–2026)

Exploration, Preparation, Implementation, Sustainment (EPIS) framework [99] when describing where ML methods might apply across the implementation continuum. Lindamer et al. [71] used RE-AIM domains to structure survey items and organise NLP-derived findings about adoption and maintenance.

Use of data science or AI-specific frameworks

Five sources explicitly named an AI- or data science-specific framework, process, or ontology in addition to, or instead of, an implementation science TMF. Shankar et al. [80] organised an NLP pipeline using the Knowledge Discovery in Databases (KDD) process (selection, preprocessing, transformation, mining, interpretation), combined with action-research cycles and the Linguistic Inquiry and Word Count (LIWC) lexicon. Shafran et al. [79] cited the Fairness Metrics Ontology (FMO) to guide evaluation of synthetic data in AI-assisted CBT training and supervision. Michie et al. [5] and Mac Aonghusa et al. [72] both built on the Behaviour Change Intervention Ontology (BCIO) as an AI-aligned knowledge-representation framework within the Human Behaviour-Change Project pipeline; this is more accurately characterised as an intervention/knowledge-representation ontology than an implementation science TMF, and we therefore catalogue it under AI/data science frameworks. Chan et al. [55] similarly developed and applied the Learning Implementation System (LIS) ontology as the knowledge-representation structure underpinning the ImpleMATE platform, alongside the implementation science TMFs noted above.

Mapping the uses of AI across the implementation process (KTA)

Sources mapped to a mean of 3.3 KTA steps (range 0–8), reflecting that many applications spanned multiple implementation activities. Figure 3 presents the within-family distribution of sources across KTA steps and primary AI approaches, with conceptual and review papers shown separately; Table 2 summarises the associated tasks, inputs, outputs, and contributing studies. Sources most frequently addressed evaluating impact (25/40), selecting and tailoring implementation strategies (23/40), assessing barriers and facilitators (21/40), and implementing and monitoring implementation (20/40). Sustainment and scale-up were addressed in 13/40 sources, synthesis of implementation-relevant knowledge in 11/40, and both problem or intervention prioritisation and adaptation to local context in 9/40. Several additional mappings were coded as unclear.

1. Implementation activities are based on the Knowledge-to-Action (KTA) model. 2. The limited number of sources included for the KTA step is explained by the fact that generic studies focusing on the use of AI to automate systematic reviews were excluded from the review. Sources were included only if they focused on synthesized implementation evidence (implementation barriers, facilitators, strategies, mechanisms, outcomes) in the context of broader projects or infrastructures (i.e., outside of a review).

Thirty-one sources were application-focused; nine were conceptual, methodological, review, or framework-focused sources that did not apply a specific AI system. Figure 4 provides a graphical summary of four concrete applications from Bartholomew et al. [50], Berkel et al.

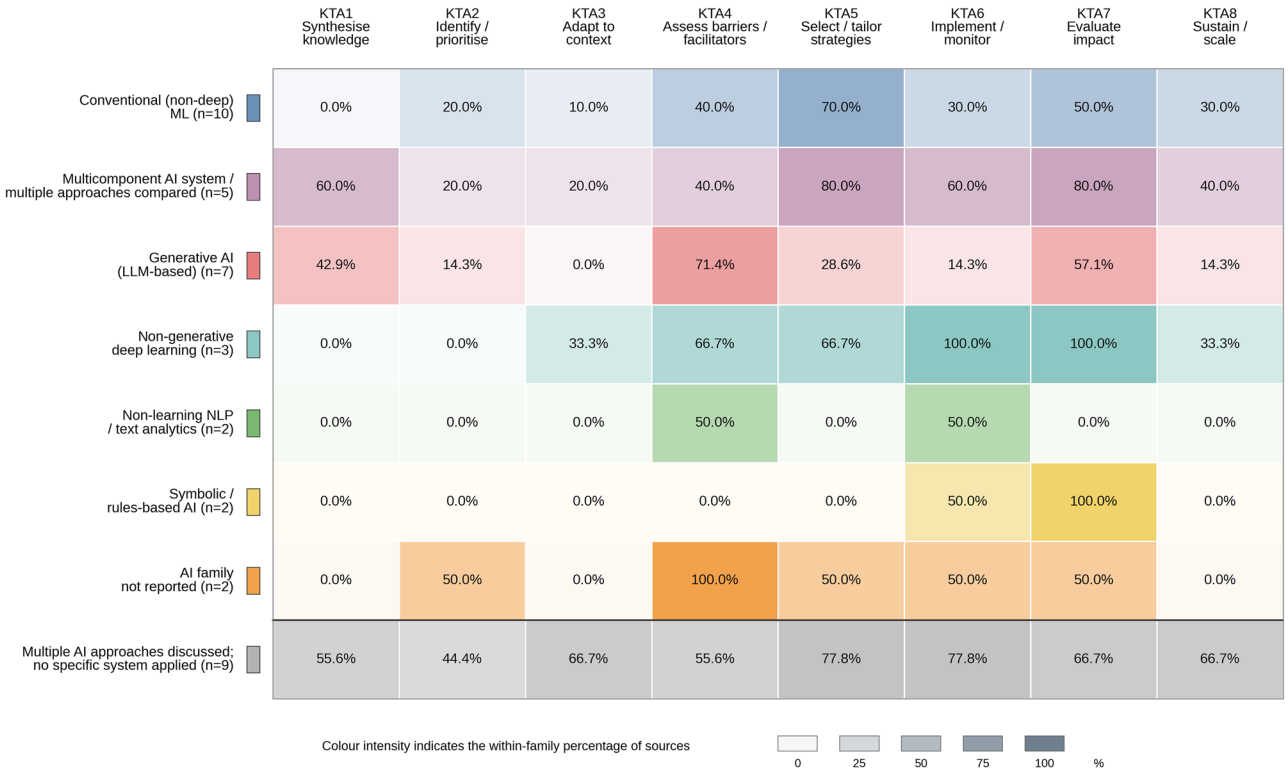


Fig. 3 AI applications and proposed uses across implementation activities, corresponding to the steps of the knowledge-to-action (KTA) model

[53], Moussa et al. [73] and Swanson et al. [83]. Several systems spanned multiple implementation activities. ImpleMATE proposed a living platform to extract evidence and contextual information, support qualitative coding, organise implementation knowledge, and inform planning, tailoring, monitoring, and evaluation [55]. The Human Behaviour-Change Project similarly combined NLP, supervised ML, and ontologies to structure implementation-relevant evidence and support downstream recommendations [5, 72].

Applications focused on implementation problem identification and prioritisation used administrative, geographic, interview, or simulation data to generate ranked targets, coded needs, or scenario estimates that could support implementation planning [50, 57, 82].

AI was used to support context-sensitive adaptation, barrier assessment, and strategy tailoring. Few sources treated adaptation as a distinct AI task; instead, AI contributed indirectly by generating feedback or extracting locally relevant contextual information [55, 58, 70]. For example, Dabravolskaj et al. [59] used supervised text classification to separate intervention activities from implementation strategies in school action plans, supporting adaptation of evidence-based interventions to local contexts. Barrier and facilitator assessment drew on the widest range of data sources and methods, including text mining, sentiment analysis, supervised ML, neural networks, and LLM-assisted coding, to identify determinant themes, stage labels, and pattern clusters. For example, Golz et al. [62] used text mining and sentiment analysis to examine how health professionals in psychiatric hospitals discussed and implemented information technologies and found that many expressions were negative, illustrating how sentiment analysis could provide information about implementation climate. Shankar et al. extended this approach into deployed practice, using a multi-component NLP pipeline (topic modelling, aspect-based sentiment, and named-entity recognition) to convert open-text patient feedback into actionable quality-improvement insights [80]. LLM-assisted qualitative coding also appeared in this step, especially in recent abstracts and proof-of-concept work [57].

Several studies treated implementation strategy selection as a matching problem, using information about barriers, context, or target populations to recommend strategies. Moussa et al. [73] provided one empirical example, using random forest models to predict which facilitation strategies were most likely to resolve specific barriers in community pharmacy implementation. Ye et al. [86] similarly used data-driven modeling to relate facilitation strategies to practice characteristics, with the aim of informing more tailored use of practice facilitation.

Table 2 AI-supported tasks across the implementation process ($N = 40$)

Implementation activity[1]	AI-supported tasks	Typical inputs	Typical outputs	Sources, n/40	Contributing sources
Synthesise implementation evidence[2]	Extract, structure and synthesise evidence on barriers, strategies, mechanisms and outcomes	Research literature, annotated reports and implementation datasets	Structured variables, coded concepts, annotations and evidence summaries	11	[4, 5, 54, 55, 67, 72, 76, 78, 79, 84, 87]
Identify and prioritise problems or interventions	Detect needs and prioritise problems, populations, settings or interventions	Administrative and geographic data, interview transcripts and simulation data	Priority rankings, needs profiles, coded themes and scenario estimates	9	[4, 50, 54, 57, 64, 66, 80, 82, 84]
Adapt evidence to local context	Support adaptation of evidence, interventions or implementation approaches to local conditions	Local contextual data, stakeholder input, guidelines and implementation evidence	Context profiles, adapted content and locally tailored recommendations	9	[4, 54, 55, 58, 59, 64, 67, 79, 84]
Assess barriers and facilitators	Identify and characterise implementation determinants	Logs, interviews, free text, programme and administrative data, and literature	Determinant themes, sentiment signals, clusters and modelled associations	21	[4, 54, 55, 57, 58, 61, 62, 64–66, 68, 73, 75, 76, 78–81, 84–86]
Select and tailor implementation strategies	Match determinants to strategies and tailor implementation strategies	Barrier–strategy datasets, electronic health record data and other implementation datasets	Ranked strategies, tailored bundles, recommended targets and implementation options	23	[4, 5, 50–52, 54, 55, 58, 59, 64, 66, 67, 70, 72, 73, 75, 76, 78–81, 84, 86]
Implement and monitor	Monitor fidelity, implementation stage, uptake and workflow change	Audit logs, clinical notes, transcripts, audio recordings, surveys and free text	Fidelity scores, implementation-stage labels, workflow indicators and monitoring summaries	20	[4, 51–56, 58, 63–65, 67, 69, 71, 76, 79–81, 84, 85]
Evaluate outcomes and impact	Derive outcomes and support rapid or comparative evaluation	Audio, text, audit logs, consultation notes and simulation outputs	Implementation, workflow, service, clinical and health system outcomes	25	[4, 51–56, 58, 61, 63–65, 68–70, 72, 74, 76, 77, 80–84, 86]
Sustain and scale	Support sustainment planning, scale-up and ongoing learning	Longitudinal implementation, quality-improvement and sustainment data	Sustainability predictions, tailored recommendations, and readiness or scalability indicators	13	[4, 52–55, 58, 59, 64, 74–76, 84, 85]

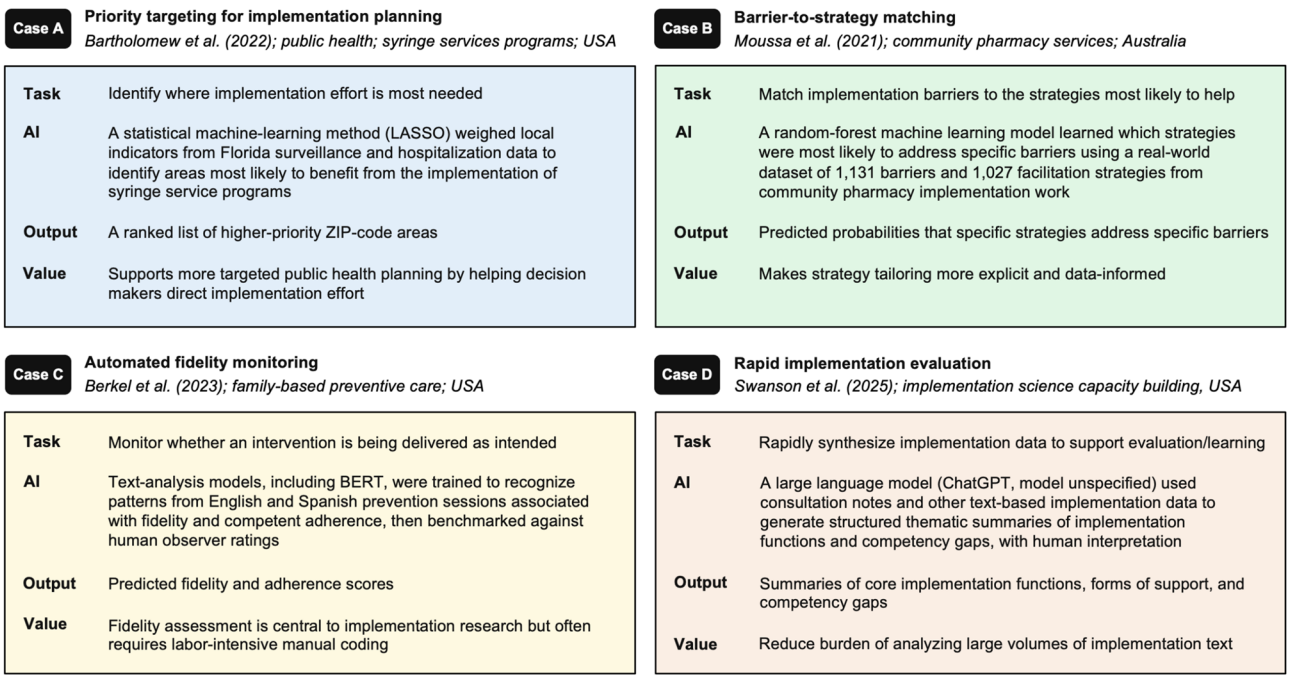


Fig. 4 Four illustrative applications of AI in implementation science. BERT = Bidirectional encoder representations from Transformers; LASSO = least absolute shrinkage and selection operator

Monitoring and evaluation were the most developed application areas. These studies used transcripts, audio, audit logs, consultation notes, and other routine traces to estimate fidelity, classify implementation stage, detect workflow change, or generate rapid qualitative summaries. Chen et al. [56] used topic modeling of EHR audit logs to estimate how clinicians' time spent on different EHR activities changed after the implementation of a tobacco cessation clinical decision support alert. Wang et al. [85] analyzed communication logs from implementation brokers to classify sites by implementation stage, identifying where programs were progressing or stalling. Lee et al. [69] deployed a fine-tuned BioClinicalBERT pipeline to screen electronic health record passages for goals-of-care discussions, routing positive hits to human adjudicators as the outcome measure of a pragmatic trial. In some cases, the evaluation focused on predicting fidelity or adherence from transcripts and audio. [53, 58] Goedken et al. [61] and Swanson et al. [83] used LLM-assisted rapid qualitative methods to evaluate AI initiatives and implementation consultations respectively. Li et al. [70] linked data-driven groupings of transitional care strategies to readmissions and patient experience.

A smaller group of sources addressed sustainment and scale-up by predicting sustainment risk or longitudinal learning. Rappon et al. [75] used a self-organizing map to identify sustainability patterns and support tailored recommendations. Other sources suggested that AI could help detect drift or changes in implementation patterns over time by analyzing longitudinal logs, fidelity data, or continuously updated evidence platforms [55, 64, 84].

Evaluation approaches and outcomes

Among the 31 application-focused sources, 12 used quantitative approaches, 12 used mixed methods, two were mixed-methods protocols, two were qualitative, one was a comparative case-study report, one was a quantitative methodological-development study, and one was a mixed-methods framework-development study (Table 3). Twenty included an explicit comparator or reference condition.

AI-specific evaluation or validation was reported or prospectively specified in 23/31 application-focused sources, although only 19/31 reported or specified a definite technical accuracy or performance outcome (Table 4). Human-centred evaluation was reported or planned

Table 3 Evaluation domains and practical evaluation questions in studies evaluating an AI system ($N=31$)

Evaluation domain	Practical evaluation question	Typical comparator/reference, when applicable	Sources, n/31	Contributing studies
A. Technical validity and reliability of AI system	Does the AI work on unseen data? (cross-validation or external validation)	Gold-standard labels; human ratings in some studies	11	[50, 53, 55, 56, 58, 59, 69, 73, 75, 77, 82]
	Where does the AI fail and why? (error analysis or agreement and error checking)	Usually human coding/judgment as reference	5	[55, 57, 72, 83, 85]
	Can users understand the AI output enough to act on it? (explainability or interpretability assessment)	Sometimes paired with manual workflow review	1	[56]
B. Transportability, robustness, and fairness of AI system	Will the AI work in a different setting or population? (external validation)	Cross-site or cross-dataset assessment	1	[69]
	Is the performance of the AI equitable across groups? (fairness or bias analysis)	Subgroup comparisons	1	[55]
	Will the AI keep working over time? (robustness or drift testing)	—	0	—
C. Human-centered and sociotechnical fit of AI system	Is the AI usable or acceptable to the people doing the implementation work? (usability, satisfaction)	Often versus manual workflow or human-only process	14	[5, 51, 55, 58, 63, 65, 69, 71, 76, 78, 80, 81, 83, 86]
D. Comparative value of AI system	Is the AI better than usual practice or a non-AI alternative? (any explicit comparator)	Standard manual process; human coding/ratings; non-AI technology; alternative AI model	20	[5, 50, 51, 53, 55–59, 63, 68–70, 72, 76, 77, 80–82, 85]
E. Implementation impact and resource consequences of AI system	Does the AI influence implementation outcomes?	Sometimes compared with usual/manual/human-only processes	25	[50, 51, 53, 55, 56, 58, 59, 61, 63, 65, 66, 68–71, 73, 75–78, 80, 81, 83, 85, 86]
	Does the AI reduce workload or speed up work? (time outcomes)	Often manual workflow or human-only process	10	[5, 51, 55, 56, 58, 59, 69, 76, 78, 81]
	Does the AI reduce cost or resources required?	Typically planned as part of broader evaluation	2	[55, 69]
	Does the AI affect equity-relevant outcomes?	Usually subgroup or access-related comparisons	3	[51, 55, 80]

Table 4 Outcome domains and measures in studies evaluating an AI system (*N* = 31)

Outcome domain	Evaluation focus	Sources, n/31	Example measures / indicators
Technical performance	Accuracy or performance of the AI output	19	Precision, recall, and F1 for extraction/classification [55]; mean squared error/concurrent validity for monitoring predictions [53]; Krippendorff's alpha and coding agreement [57]; diagnostic odds ratio, sensitivity, and specificity [75]; variance explained and absolute percent error [82]; prediction accuracy / predicted resolution probability [73]
Efficiency and resource use	Whether AI reduced time, workload, or resource requirements	10 for time; 2 for cost	Time spent in EHR activities / workflow time-in-system [56]; planned scalability and efficiency indicators [58]; planned cost-effectiveness and resource requirements [55]
Human-centered outcomes	User experience or response to the AI-enabled approach	12	System Usability Scale [63]; satisfaction and acceptability [71]; sentiment toward implemented technologies [62]; planned usability evaluation [55, 58]
Implementation-specific outcomes	Outcomes directly relevant to implementation processes or success	25	Fidelity [58, 63]; engagement / implementor behavior indicators [53]; implementation-stage classification [85]; workflow change signals [56]; readiness domains [58]; competency gap analysis [83]; strategy–barrier fit / predicted resolution probability [73]; sustainment risk classification [75]
Equity-related outcomes	Differential access, benefit, or performance across groups	3	Equity-related utilization/disparity reduction [51]; planned fairness/equity checks [55]
Clinical or health system outcomes	Downstream patient, population, or service effects linked to AI-supported implementation work	9	Readmissions / patient outcomes linked to strategy groupings [70]; child independent correct handwashing responses [63]; client outcomes and dropout [58]; service/public health targeting outputs [50]

in 12/31. Only one source externally validated the AI system; none completed a formal fairness analysis, although one protocol planned bias monitoring, and none assessed robustness or model drift.

Implementation outcomes were the most frequently reported outcome domain (25/31), typically represented by proximal indicators such as fidelity, implementation-stage classification, workflow change, or engagement rather than broader implementation success. Time or efficiency outcomes were reported in 10/31 sources, and clinical or health system outcomes in 9/31. Cost or resource outcomes (2/31) and equity-related outcomes (3/31) were rarely assessed. No source explicitly evaluated the overall appropriateness of using AI for the task as a distinct construct.

Risks, adverse effects, and unintended consequences

Twenty-two of 40 sources discussed risks, adverse effects, or unintended consequences of AI use; 16 did not, and reporting was unclear in two. Risks were discussed across empirical and non-empirical sources, but were generally described as limitations or anticipated concerns rather than measured as prospective safety or harm outcomes. AI-specific concerns, including bias, hallucinations, prediction error, data limitations, and opacity, were explicitly discussed in 17/40 sources. Workflow-related risks were discussed in 11/40 and user-level risks such as overreliance, trust, acceptability, or de-skilling in 9/40. Only one source explicitly discussed clinical or patient-level harms. Environmental impacts were substantively addressed by only one source. Broader societal impacts were raised in three sources through equity implications, cultural/linguistic misalignment in low-resource settings, or widening of health disparities [66, 67, 79], but were otherwise absent.

Two empirical sources nevertheless identified concrete problems. Rinne et al. [76] found that ChatGPT-3.5 outputs lacked nuance, required extensive human editing, and produced erroneous citations, requiring close supervision throughout the process. Sibley et al. [81] observed an unintended provider-level effect: motivational interviewing quality declined over time in the AI-assisted group, suggesting that AI fidelity feedback may foster a task-oriented therapist mindset that prioritises efficiency over relational therapeutic process. No source reported abandonment or withdrawal of an AI system because of harm or poor fit.

Discussion

This living scoping review mapped how AI is being used and evaluated in implementation science and practice. The evidence base remains small, although publication has accelerated since 2021. Current applications cluster in codifiable, data-rich tasks, including assessing barriers and facilitators, selecting and tailoring strategies, monitoring implementation and evaluating impact, while earlier-cycle and relational work is less represented. Evaluation is methodologically narrow, as proximal technical or implementation signals predominate, while external validation, equity, cost, sociotechnical effects, and environmental consequences are rarely examined. The field is therefore moving faster in developing new AI systems than in generating evidence on evaluation, safety, and governance for wider adoption.

Most AI systems identified relied on traditional ML methods, with emerging use of generative AI

and LLMs. We found no evaluated examples of several AI approaches that are now prominent elsewhere in applied AI, including agentic systems [25], custom GPT-style implementations [100], multimodal models [20], reinforcement learning approaches [101], or more autonomous decision-support architectures. By 2026, adjacent clinical AI literatures were already reporting EHR-embedded, agentic decision-support systems that combine retrieval, reasoning, recommendation, and workflow support [102–104], whereas these were absent from the implementation science literature captured in our review. The absence of these approaches likely reflects their recent diffusion and the longer timelines required for deployment, evaluation, and publication in health research.

AI use concentrated where implementation work can be represented as classifications, rankings, scores, or summaries. Earlier and more relational activities [105], including framing problems, negotiating priorities, co-adapting interventions, and resolving value-laden trade-offs, were less visible, as were AI uses for selecting or refining theories, models, and frameworks. [106] Importantly, frequent use does not mean that AI is appropriate for a given task. For any proposed use, teams should ask whether AI offers incremental value over a simpler method and whether it is locally valid, acceptable, feasible, equitable, governable, affordable, and compatible with relational and workflow requirements.

The absence of published accounts of routine implementation practice should not be interpreted as absence of use. Our search was anchored in implementation science and related health terminology; operational applications may instead be described as quality improvement, process improvement, change management, service redesign, analytics, or not published at all. Future updates will broaden practice-oriented terminology and targeted sources, including technical reports, organisational case studies, conference materials, and other grey literature.

Patient, workforce, equity, societal, and environmental consequences of AI use were rarely examined. An AI system may improve speed or classification accuracy while increasing cognitive burden, weakening trust, shifting decision-making away from frontline actors, or disadvantaging groups poorly represented in the underlying data. [35, 107–109] Sibley et al. [81] provide a concrete warning, as motivational interviewing quality declined over time among therapists receiving AI-generated fidelity feedback. Evaluation that stops at technical performance of the AI system can therefore miss effects on relationships, professional judgement, and implementation quality.

No single existing framework fully captures the requirements for evaluating AI used in implementation science and practice. Proctor's implementation outcomes can guide the assessment of whether an AI-supported approach is acceptable, appropriate, feasible, adopted, delivered with fidelity, affordable, and sustained [91]. PRISM-Capabilities extends this perspective by emphasizing multilevel context, implementation infrastructure, equity, community engagement, and the organizational capabilities needed to introduce AI responsibly [60]. FUTURE-AI can contribute AI-specific criteria, including fairness, universality, traceability, usability, robustness, and explainability [108]. Environmental assessment is a further dimension that is largely absent from these frameworks. It should document, at minimum, the model and deployment mode, frequency and scale of use, and available information on computing or energy requirements [110, 111].

Box 2 Research implications of this review for AI in implementation science

Establish appropriateness and broaden applications

1. Before selecting AI, define the implementation problem, affected users, and why AI is expected to add value beyond a non-AI method; assess local appropriateness using acceptability, feasibility, workflow fit, equity, governance, cost, and environmental burden.
2. Evaluate AI systems for additional implementation tasks in the exploration and preparation phases.
3. Test whether AI-supported barrier–strategy matching improves downstream implementation outcomes, rather than only predicting strategy fit or generating recommendations.

Strengthen real-world evaluation

1. Compare AI-assisted implementation workflows with usual practice, human-only workflows, simpler non-AI tools and/or other AI models in prospective, randomized, controlled studies.
2. Validate existing AI systems in new settings, languages, lower-resource contexts, and populations underrepresented in the training data.
3. Study AI model drift, recalibration needs, and sustainment over time, and separate sustainment of the AI tool from sustainment of the implementation it supports.

Assess equity, human factors, and safety

1. Measure subgroup performance and equity effects to determine whether AI-supported implementation narrows or widens disparities.
2. Examine AI system usability, trust, interpretability, workflow fit, and harms such as overreliance, hallucinations, privacy risks, and workflow disruption before scale-up.
3. Examine the resource, cost and environmental implications of AI use in the field.

Study emerging AI systems

1. Test newer AI systems, including custom GPT–type tools and agentic AI systems, for multi-step implementation tasks and compare them with simpler single-purpose tools.
2. Develop and evaluate modular, interoperable AI architectures that support multiple implementation functions within a single workflow, and test whether integrated systems outperform single-purpose tools.

Build standards and governance

1. Develop consensus reporting and outcome standards for AI-assisted implementation studies so future research uses comparable measures.
2. Build and evaluate governance, consent, and data infrastructure for responsible AI use, including data access, privacy, accountability, community oversight, and minimum AI-literacy competencies for implementation teams.

The review maps where AI has been used; it does not by itself determine whether those uses were warranted. The next phases of the AIM-IS program will examine when AI is appropriate for implementation research and practice tasks. The program aims to develop a framework, toolkit, and reporting standard for the responsible use of AI in implementation science and practice [41]. These products will be refined through stakeholder consultation, consensus methods, and usability testing.

These findings suggest five overarching priorities for future research and practice (Box 2).

Determining what was "in" or "out" of scope for this review was conceptually challenging. Adjacent literatures are already large and rapidly expanding, particularly in three areas. One is the use of AI for evidence synthesis, including citation screening, study classification, data extraction, and evidence mapping [112, 113]. A second is the use of AI for generating knowledge products, such as drafting evidence summaries, guideline text, implementation briefs, decision aids, audit-and-feedback reports, or tailored educational materials [114–116]. A third is the use of AI within provider- or patient-facing implementation strategies, [117] for example adaptive e-learning platforms, [8] clinical decision support systems, [109] conversational agents used to support clinician training, [118] and patient-facing chatbots or digital coaches used to increase uptake, adherence, or engagement [119]. These applications are relevant to implementation, but they are also large enough to constitute distinct review topics. As this living review evolves, it may become appropriate to broaden eligibility criteria or develop parallel review streams.

For practice and policy, based on the evidence identified in this review, AI applications are better supported as augmentation tools rather than as autonomous implementation solutions. At present, no AI tools identified in this review can be recommended for general use. Decision-makers should require local validation, human oversight, subgroup monitoring, clear model-update procedures, and appropriate privacy and governance safeguards. Implementation teams also need to develop practical AI literacy: recognising uncertainty and hallucinations, protecting confidential data, documenting models and prompts, checking subgroup effects and drift, maintaining human oversight, and knowing when not to use AI. Co-design with affected partners should precede scale-up.

Several limitations should be considered. Conference abstracts, protocols, and conceptual papers helped capture an emerging field but often provided limited methodological detail. Grey literature was eligible, but we did not undertake a dedicated grey-literature database or systematic web search. Search terminology anchored in implementation science and health may have missed operational applications described as process improvement, change management, service redesign, or analytics. Classification of AI family, implementation function, maturity, and KTA step involved judgement, particularly for cross-cutting systems. Finally, restricting the review to the previous 16 years may have excluded earlier work.

Conclusion

AI is beginning to be integrated into implementation science and practice, but the evidence remains limited, narrow in scope, and uneven in maturity. Identified applications cluster in codifiable, later-stage implementation tasks, including determinant assessment, strategy tailoring, monitoring, and evaluation, and usually involve human-in-the-loop tools that generate signals, rankings, or summaries rather than autonomous systems embedded in routine practice. The most important next step is therefore not more AI, but better evidence on when, where, for whom, and under what conditions AI improves, or undermines, implementation work. Future cycles of this review will track changes in applications, evaluation methods, risks, and remaining evidence gaps as AI technologies evolve.

Abbreviations

AI	Artificial intelligence;
BERT	Bidirectional encoder representations from transformers;
CBT	Cognitive behavioural therapy;
CDSS	Clinical decision support system;
DL	Deep learning;
EHR	Electronic health record;
KTA	Knowledge-to-action;
LLM	Large language model;
ML	Machine learning;
NLP	Natural language processing;
PCC	Population–concept–context;
PRISMA-LSR	Preferred reporting items for systematic reviews and meta-analyses extension for living systematic reviews;
PRISMA-ScR	Preferred reporting items for systematic reviews and meta-analyses extension for scoping reviews;
RE-AIM	Reach, effectiveness, adoption, implementation, and maintenance;
TMF	Theories, models and frameworks.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13012-026-01535-x.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Acknowledgements

Not applicable.

Author contributions

GF conceptualized the review. GF led the study with senior input from NT, IDG, JP, SES, SDL, SM, BJP, VW, JT, FL, and SAR. GF, ODL, RL, JD, CZ, JC, LC, ADMP, JE and TR contributed to study selection and/or data extraction. GF led the

synthesis and drafted the manuscript. ODL, ADMP, JD, and LC contributed to data checking and manuscript development. JC, RA, RSB, CF, EG, JH and TR contributed content expertise on AI, implementation science, and health systems applications. All authors contributed to the interpretation of findings, critically revised the manuscript for important intellectual content, and approved the final version.

Funding
GF holds a Tier 2 Canada Research Chair in Implementation Science for Health Systems Transformation (Canadian Institutes of Health Research). GF's salary was funded partly by a Fonds de recherche du Québec - Santé (FRQ-S) Junior 1 Research Scholar Award. The AI Methods for Implementation Science (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. FL is supported by a Chaire de recherche du Canada sur la décision partagée et la mobilisation des connaissances niveau 1.

Data availability
The extracted dataset and full search strategies will be made available in the Open Science Framework alongside the review protocol: https://doi.org/10.17605/OSF.IO/2Q5DV.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
GF, JC, JH and NT are authors on the ImpleMATE protocol, included in this review as an evidence source. GF received speaker fees from AbbVie for unrelated work. GF is an Associate Editor for *Implementation Science Communications*. SM is the Principal Investigator on the Human Behaviour Change Project, linked to two sources included in this review. RSB receives funding from the National Institutes of Health, Northwestern Medicine, National Psoriasis Foundation, American Foundation for Suicide Prevention, West Health Institute, Sanofi, and Pfizer for her research; is principal at Implementation Science & Practice, LLC; is currently an appointed member of the National Advisory Mental Health Council; serves on the scientific advisory board for AIM Youth Mental Health Foundation, Klingenstein Third Generation Foundation, and the American Foundation for Suicide Prevention; and serves as Co-Editor-in-Chief for *Implementation Science*. She will not be involved in any decisions related to this manuscript. The other authors declare that they have no competing interests.

Author details
[1]Ingram School of Nursing, Faculty of Medicine and Health Sciences, McGill University, Montréal, Canada
[2]Department of Global and Public Health, Faculty of Medicine and Health Sciences, McGill University, Montréal, Canada
[3]Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Montréal, Canada
[4]Centre for Implementation Research, Ottawa Hospital Research Institute, Ottawa, Canada
[5]Kirby Institute, University of New South Wales, Sydney, Australia
[6]Sir Mortimer B. Davis Jewish General Hospital, CIUSSS West-Central Montreal, Montréal, Canada
[7]Implementation to Impact (I2i), School of Population Health, UNSW Sydney, Sydney, Australia
[8]Department of Family Medicine, McGill University, Montreal, Canada
[9]Mila – Quebec Artificial Intelligence Institute, Montréal, Canada
[10]Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, Canada
[11]Silver School of Social Work, New York University, New York, USA
[12]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[13]Li Ka Shing Knowledge Institute, Unity Health Toronto, Toronto, Canada
[14]Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, Canada
[15]Division of Infectious Diseases, John T. Milliken Department of Medicine, Washington University in St. Louis, St. Louis, Missouri, USA
[16]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada
[17]Idiap Research Institute, Martigny, Switzerland
[18]Swiss Institute of Bioinformatics, Lausanne, Vaud, Switzerland
[19]St. Mary's Research Centre, CIUSSS West-Montreal, Montréal, Canada
[20]Department of Family and Emergency Medicine, Université Laval, Québec City, Canada
[21]VITAM Research Centre for Sustainable Health, Québec City, Canada
[22]Centre for Behaviour Change, University College London, London, UK
[23]Brown School, Washington University in St. Louis, St. Louis, Missouri, USA
[24]School of Psychology, University of Ottawa, Ottawa, Canada
[25]Department of Medicine, University of Toronto, Toronto, ON, Canada
[26]EPPI Centre, Social Research Institute, University College London, London, UK
[27]Bruyère Health Research Institute, Ottawa, ON, Canada

Received: 28 May 2026 / Accepted: 4 August 2026
Published online: 14 August 2026

References

1. Angus DC, Khera R, Lieu T, Liu V, Ahmad FS, Anderson B, et al. AI, health, and health care today and tomorrow: the JAMA summit report on artificial intelligence. JAMA. 2025, Nov, 11;334(18):1650–64. https://doi.org/10.1001/jama.2025.18490.
2. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med. 2023, Mar, 30;388(13):1201–08. https://doi.org/10.1056/NEJMra2302038.
3. Rashid AB, Kausik MDAK. AI revolutionizing industries worldwide: a comprehensive overview of its diverse applications. Hybrid Adv. 2024;7:100277. https://doi.org/10.1016/j.hybadv.2024.100277.
4. Younas A, Reynolds SS. Leveraging artificial intelligence for expediting implementation efforts. Creative Nurs. 2024, May;30(2):111–17. https://doi.org/10.1177/10784535241239059.
5. Michie S, Thomas J, Johnston M, Aonghusa PM, Shawe-Taylor J, Kelly MP, et al. The human behaviour-change project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. Implementation Sci. 2017, Oct, 18;12(1):121. https://doi.org/10.1186/s13012-017-0641-5.
6. West R, Bonin F, Thomas J, Wright AJ, Mac Aonghusa P, Gleize M, et al. Using machine learning to extract information and predict outcomes from reports of randomised trials of smoking cessation interventions in the human behaviour-change project. Wellcome Open Res. 2024;8:452. https://doi.org/10.12688/wellcomeopenres.20000.2.
7. Hastings J, Glauer M, West R, Thomas J, Wright AJ, Michie S. Predicting outcomes of smoking cessation interventions in novel scenarios using ontology-informed, interpretable machine learning. Wellcome Open Res. 2023;8:503. https://doi.org/10.12688/wellcomeopenres.20012.1.
8. Fontaine G, Cossette S, Maheu-Cadotte MA, Mailhot T, Deschênes M-F, Mathieu-Dupuis G, et al. Efficacy of adaptive e-learning for health professionals and students: a systematic review and meta-analysis. BMJ Open. 2019, Aug, 28;9(8):e025252. https://doi.org/10.1136/bmjopen-2018-025252.
9. Team G. Gemma: open models based on Gemini research and technology. 2024.
10. OpenAI. GPT-4 technical report. 2024.
11. Hugo Touvron TL, Izacard G, Martinet X, Lachaux M-A, Lacroix T, Rozière B, et al. Guillaume Lample. LLaMA: Open efficient foundation language model. 2023.
12. US FDA. FDA digital health and artificial intelligence glossary - educational resource. https://www.fda.gov/science-research/artificial-intelligence-and-medical-products/fda-digital-health-and-artificial-intelligence-glossary-educational-resource?utm_source=chatgpt.com.
13. Schwartz R, Vassilev A, Greene K, Perine L, Burt A, Hall P. Towards a standard for identifying and managing bias in artificial intelligence. 2022. https://doi.org/10.6028/nist.Sp.1270.

14. WHO. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. 2025. https://www.who.int/publications/i/item/9789240084759?utm_source=chatgpt.com.
15. Bommasani RDA, Adeli E, Altman R, Arora S, von Arx S., … Liang P. On the opportunities and risks of foundation models [v3]. arXiv. 2022, https://doi.org/10.48550/arXiv.2108.07258.
16. Horvitz E, Mulligan D. Policy forum. Data, privacy, and the greater good. Science. 2015, Jul, 17;349(6245):253–55. https://doi.org/10.1126/science.aac4520.
17. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023, Aug;29(8):1930–40. https://doi.org/10.1038/s41591-023-02448-8.
18. Liu F, Zhou H, Gu B, Zou X, Huang J, Wu J, et al. Application of large language models in medicine. Nat Rev Bioeng. 2025;3(6):445–64. https://doi.org/10.1038/s44222-025-00279-5.
19. Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. Inf Fusion. 2022;81:84–90. https://doi.org/10.1016/j.inffus.2021.11.011.
20. Yin S, Fu C, Zhao S, Li K, Sun X, Xu T, et al. A survey on multimodal large language models. Natl Sci Rev. 2024, Dec;11(12):nwae403. https://doi.org/10.1093/nsr/nwae403.
21. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015, May, 28;521(7553):436–44. https://doi.org/10.1038/nature14539.
22. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [v7]. arXiv. 2023. https://doi.org/10.48550/arXiv.1706.03762.
23. He R, Cao J, Tan T. Generative artificial intelligence: a historical perspective. Natl Sci Rev. 2025, May;12(5):50. https://doi.org/10.1093/nsr/nwaf050.
24. Hirschberg J, Manning CD. Advances in natural language processing. Science. 2015, Jul, 17;349(6245):261–66. https://doi.org/10.1126/science.aaa8685.
25. Abou Ali M, Dornaika F, Charafeddine J. Agentic AI: a comprehensive survey of architectures, applications, and future directions. Artif Intell Rev. 2025;59(1). https://doi.org/10.1007/s10462-025-11422-4.
26. Glasziou P, Straus S, Brownlee S, Trevena L, Dans L, Guyatt G, et al. Evidence for underuse of effective medical services around the world. Lancet. 2017, Jul, 8;390(10090):169–77. https://doi.org/10.1016/S0140-6736(16)30946-1.
27. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, et al. Evidence for overuse of medical services around the world. Lancet. 2017, Mar, 5;390(10090):156–68. https://doi.org/10.1016/S0140-6736(16)32585-5.
28. Rapport F, Smith J, Hutchinson K, Clay-Williams R, Churruca K, Bierbaum M, et al. Too much theory and not enough practice? The challenge of implementation science application in healthcare practice. Evaluation Clin Pract. 2022, Dec;28(6):991–1002. https://doi.org/10.1111/jep.13600.
29. Morrow A, Tyedmers E, Debono D, Chan P, Steinberg J, Tiernan G, et al. Trials and tribulations: a qualitative exploration of researcher perspectives on navigating the challenges of health system implementation research. BMJ Open. 2025, Jan, 15;15(1):e087926. https://doi.org/10.1136/bmjopen-2024-087926.
30. Baumann AA, Cabassa LJ. Reframing implementation science to address inequities in healthcare delivery. BMC Health Serv Res. 2020, Mar, 12;20(1):190. https://doi.org/10.1186/s12913-020-4975-3.
31. Beidas RS, Dorsey S, Lewis CC, Lyon AR, Powell BJ, Purtle J, et al. Promises and pitfalls in implementation science from the perspective of US-based researchers: learning from a pre-mortem. Implementation Sci. 2022, Aug, 13;17(1):55. https://doi.org/10.1186/s13012-022-01226-3.
32. Xie CX, Franklin PD, Walunas TL, Beidas RS. Data-driven implementation trials: realizing their full potential in achieving the promise of learning health systems. Learn Health Syst. 2026, Jan;10(1):e70043. https://doi.org/10.1002/lrh2.70043.
33. Marko JGO, Neagu CD, Anand PB. Examining inclusivity: the use of AI and diverse populations in health and social care: a systematic review. BMC Med Inf Decis Mak. 2025, Feb, 5;25(1):57. https://doi.org/10.1186/s12911-025-02884-1.
34. Dankwa-Mullan I. Health Equity and ethical considerations in using artificial intelligence in Public health and medicine. Prev Chronic Dis. 2024, Aug, 22;21:E64. https://doi.org/10.5888/pcd21.240245.
35. Goktas P, Grzybowski A. Shaping the future of healthcare: ethical clinical challenges and pathways to trustworthy AI. JCM. 2025, Feb, 27;14(5):1605. https://doi.org/10.3390/jcm14051605.
36. Chustecki M. Benefits and risks of AI in health care: narrative review. Interact J Med Res. 2024, Nov, 18;13:e53616. https://doi.org/10.2196/53616.
37. Nundy S, Cooper LA, Mate KS. The quintuple aim for health care improvement: a new imperative to advance health equity. JAMA. 2022, Feb, 8;327(6):521–22. https://doi.org/10.1001/jama.2021.25181.
38. Wathen CN, Browne AJ, Wilson E, Bungay V, Varcoe C. Operationalizing the quintuple aim of health system improvement through Equity-oriented health Care. Health?serv?Insights. 2025;18:11786329251381442. https://doi.org/10.1177/11786329251381442.
39. Lee P, Bubeck S, Petro PJ. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med. 2023, Mar, 30;388(13):1233–39. https://doi.org/10.1056/NEJMsr2214184.
40. Katirai A. The environmental costs of artificial intelligence for healthcare. ABR. 2024, Jul;16(3):527–38. https://doi.org/10.1007/s41649-024-00295-4.
41. Fontaine G, Michie S, Beidas RS, Geng E, Fahim C, Powell BJ, et al. 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. Implementation Sci. 2026, Apr, 14;21(1). https://doi.org/10.1186/s13012-026-01503-5.
42. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: scoping reviews (2020 version). in. In: Aromataris E, Munn Z, editors. Joanna Briggs institute (JBI) manual for evidence synthesis. JBI; 2020.
43. 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.
44. Fontaine G G, Di Lalla O, Michie S, J. Powell B, Welch V, Thomas J, 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;14:1135. https://doi.org/10.12688/f1000research.171774.2.
45. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018, Oct, 2;169(7):467–73. https://doi.org/10.7326/M18-0850.
46. Akl EA, Khabsa J, Iannizzi C, Piechotta V, Kahale LA, Barker JM, et al. Extension of the PRISMA 2020 statement for living systematic reviews (PRISMA-LSR): checklist and explanation. BMJ. 2024, Nov, 19;387:e079183. https://doi.org/10.1136/bmj-2024-079183.
47. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: Time for a map? J Contin Educ Health Prof. 2006, Winter;26(1):13–24. https://doi.org/10.1002/chp.47.
48. Harvey G, Rycroft-Malone J, Seers K, Wilson P, Cassidy C, Embrett M, et al. Connecting the science and practice of implementation – applying the lens of context to inform study design in implementation research. Front Health Serv. 2023;3:1162762. https://doi.org/10.3389/frhs.2023.1162762.
49. Veritas Health Innovation. Covidence systematic review software. https://doi.org/www.covidence.org
50. 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, Dec;54(1):2137–50. https://doi.org/10.1080/07853890.2022.2105391.
51. Barwise A, Strechen I, Eltalhi T, Piche K, Wilson P, Tschida-Reuter D, et al. Improving in-person interpreter utilization in complex care: findings from a stepped-wedge cluster randomized trial of integrated artificial intelligence. J Gen Intern Med. 2026, Jan, 5;41(5):1312–21. https://doi.org/10.1007/s11606-025-10104-9.
52. Berkel C, Gallo CG, Sandler IN, Mauricio AM, Smith JD, Brown CH. Redesigning implementation measurement for monitoring and quality improvement in community delivery settings. J Primary Prevent. 2019, Feb;40(1):111–27. https://doi.org/10.1007/s10935-018-00534-z.
53. Berkel C, Knox DC, Flemotomos N, Martinez VR, Atkins DC, Narayanan SS, et al. A machine learning approach to improve implementation monitoring of family-based preventive interventions in primary care. Implement Res Pract. 2023, Jan-Dec;4:26334895231187906. https://doi.org/10.1177/26334895231187906.
54. Brown CH, Mohr DC, Gallo CG, Mader C, Palinkas L, Wingood G, et al. A computational future for preventing HIV in minority communities: how advanced technology can improve implementation of effective programs. JAIDS J Acquired Immune Deficiency Syndromes. 2013, Jun, 1;63(Suppl 1):S72–84. https://doi.org/10.1097/QAI.0b013e31829372bd.
55. Chan J, Hastings J, Tiernan G, Paterson J, Lovell N, Betbeder-Matibet L, et al. Combining implementation and data sciences to accelerate evidence integration into healthcare – ImpleMATE. F1000Research. 2025;14(1184):1184. https://doi.org/10.12688/f1000research.169999.1.
56. Chen J, Cutrona SL, Dharod A, Moses A, Bridges A, Ostasiewski B, et al. Electronic health record activity changes around new decision support

implementation: monitoring using audit logs and topic modeling. JAMIA Open. 2025, Aug;8(4):ooaf050. https://doi.org/10.1093/jamiaopen/ooaf050.
57. 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). Maryland, United States: 2025.
58. Creed TA, Salama L, Slevin R, Tanana M, Imel Z, Narayanan S, et al. Enhancing the quality of cognitive behavioral therapy in community mental health through artificial intelligence generated fidelity feedback (project AFFECT): a study protocol. BMC Health Serv Res. 2022, Sep, 20;22(1):1177. https://doi.org/10.1186/s12913-022-08519-9.
59. Dabravolskaj J, Kalubi J, Moore J, Mandour BA, Honorato C, Veugelers PJ, et al. Intervention activities and implementation strategies for school-based health promotion: identifying core functions and forms to facilitate scale-up of an effective intervention. Implement Res Pract. 2025, Jan-Dec;6:26334895251385936. https://doi.org/10.1177/26334895251385936.
60. El-Bassel N, David J, Mukherjee TI, Aggarwal M, Wu E, Gilbert L, et al. The practical, robust implementation and sustainability (PRISM)-capabilities model for use of artificial intelligence in community-engaged implementation science research. Implementation Sci. 2025, Aug, 7;20(1):37. https://doi.org/10.1186/s13012-025-01447-2.
61. Goedken CC, Hook D, Hill D, Holman J, Morgan D, Brinkley K, et al. 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). Maryland, United States: 2025.
62. Golz C, Aarts S, Hacking C, Hahn S, Zwakhalen SMG. Health professionals' sentiments towards implemented information technologies in psychiatric hospitals: a text-mining analysis. BMC Health Serv Res. 2022, Nov, 28;22(1):1426. https://doi.org/10.1186/s12913-022-08823-4.
63. Griffen B, Lorah ER, Caldwell N, Hantula DA, Nosek J, Tincani M, et al. The effects of artificial intelligence on implementors' fidelity of instructional strategies during handwashing acquisition in children with autism. J Dev Phys Disabil. 2024;36(5):793–819. https://doi.org/10.1007/s10882-023-09937-1.
64. Huguet N, Chen J, Parikh RB, Marino M, Flocke SA, Likumahuwa-Ackman S, et al. Applying Machine Learning Techniques to Implementation Science. Online J Public Health Inf. 2024, Apr, 22;16:e50201. https://doi.org/10.2196/50201.
65. Ironside M SM, Willging C, Ehrhart M, Aarons G. Enhancing delivery of evidence-based behavioral healthcare with AI vs expert human feedback tools: predictors and patterns of provider engagement. Implement Sci. 2025;20(29)(1): Abstract S4. https://doi.org/10.1186/s13012-025-01433-8.
66. Jones J, Knox J, Meanley S, Yang C, Lounsbury DW, Huang TT, et al. Explorations of the role of digital technology in HIV-Related implementation research: case comparisons of five ending the HIV epidemic supplement awards. JAIDS J Acquired Immune Deficiency Syndromes. 2022, Jul, 1;90(S1):S226–34. https://doi.org/10.1097/QAI.0000000000002983.
67. King AJ, Kahn JM. The role of data science in closing the implementation gap. Crit Care Clin. 2023, Oct;39(4):701–16. https://doi.org/10.1016/j.ccc.2023.03.005.
68. Knippen KL, Woodcock L, Papoutsakis C, Gonzales-Pacheco DM. Implementation of clinical practice guidelines using the plan–do–study–act framework: the methodology and experiences of the Academy of nutrition and dietetics health informatics infrastructure registry study on gestational diabetes mellitus. Nut Clin Prac. 2025, Dec;40(6):1465–82. https://doi.org/10.1002/ncp.70043.
69. Lee RY, Li KS, Sibley J, Cohen T, Lober WB, O'Brien J, et al. A modular pipeline for natural language processing-screened human abstraction of a pragmatic trial outcome from electronic health records. Clin Trials. 2026, Apr;23(2):145–54. https://doi.org/10.1177/17407745251405386.
70. Li J, Du G, Clouser JM, Stromberg A, Mays G, Sorra J, et al. Improving evidence-based grouping of transitional care strategies in hospital implementation using statistical tools and expert review. BMC Health Serv Res. 2021, Jan, 7;21(1):35. https://doi.org/10.1186/s12913-020-06020-9.
71. Lindamer L, Almklov E, Pittman JOE, Shi S, Maye J, Jak A, et al. Multi-method study of the implementation of cognitive symptom management and rehabilitation training (CogSMART) in real-world settings. BMC Health Serv Res. 2022, Dec, 17;22(1):1542. https://doi.org/10.1186/s12913-022-08941-z.
72. Mac Aonghusa P, Michie S. Artificial intelligence and behavioral science through the looking glass: challenges for real-world application. Ann Behav Med. 2020, Dec, 1;54(12):942–47. https://doi.org/10.1093/abm/kaaa095.
73. 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. Implementation Sci. 2021, Jul, 19;16(1):73. https://doi.org/10.1186/s13012-021-01138-8.
74. Pearson TA, Califf RM, Roper R, Engelgau MM, Khoury MJ, Alcantara C, et al. Precision health analytics with predictive analytics and implementation research: JACC state-of-the-art review. J Am Coll Cardiol. 2020, Jul, 21;76(3):306–20. https://doi.org/10.1016/j.jacc.2020.05.043.
75. Rappon T. The sustainment and Sustainability of quality improvement initiatives for the health care of older adults. University of Toronto; 2021.
76. Rinne ST, Brunner J, Hogan TP, Ferguson JM, Helmer DA, Hysong SJ, et al. A use case of ChatGPT: summary of an expert panel discussion on electronic health records and implementation science. Front Digit Health. 2024;6:1426057. https://doi.org/10.3389/fdgth.2024.1426057.
77. Rosen CS, Peterson AL, Riggs DS, Haddock CK, Fong C, Young-McCaughan S, et al. Tailored strategies to increase the use of an evidence-based psychotherapy for posttraumatic stress disorder: a stepped-wedge randomized trial in military clinics. J Anxiety Disord. 2025, Dec;116:103080. https://doi.org/10.1016/j.janxdis.2025.103080.
78. Rosen D, Zwanzig D, Vogel B, Erhart M, Reiter NL. Physical therapists' perspectives on a large language model-powered knowledge translation tool for guideline adherence: a qualitative focus group study. Physiother Theory Pract. 2026, Jan, 6;1–14. https://doi.org/10.1080/09593985.2025.2606058.
79. Shafran R, Bond L, Carlbring P, Cohen ZD, Creed TA, Davey E, et al. From innovation to implementation: artificial intelligence in cognitive behaviour therapy training and supervision. Behav Res Ther. 2026, Feb;197:104945. https://doi.org/10.1016/j.brat.2025.104945.
80. Shankar R, Yip A. Transforming patient feedback into actionable insights through natural language processing: knowledge discovery and action research study. JMIR Form Res. 2025, Aug, 26;9:e69699. https://doi.org/10.2196/69699.
81. Sibley MH, Bickman L, Atkins D, Coxe S, King J, Tanana M, et al. Improving community-based care for adolescents with ADHD: a randomized controlled trial of artificial intelligence-assisted fidelity supports. Prev Sci. 2025, Dec, 24. https://doi.org/10.1007/s11121-025-01868-x.
82. Stanfield A, Mayorga ME, O'Leary MC, Lich KH. Metamodel of a simulation Model of colorectal cancer with diverse clinic populations and intervention scenarios. presented at: 2024 Winter Simulation Conference; 2024, https://ieeexplore.ieee.org/xpl/conhome/10838618/proceeding.
83. Swanson KN, Stadnick NA, Aarons GA, Brookman-Frazee L, Bouchard I, Du Z, 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, Aug, 18;6(1):87. https://doi.org/10.1186/s43058-025-00775-0.
84. Trinkley KE, An R, Maw AM, Glasgow RE, Brownson RC. Leveraging artificial intelligence to advance implementation science: potential opportunities and cautions. Implementation Sci. 2024, Feb, 21;19(1):17. https://doi.org/10.1186/s13012-024-01346-y.
85. Wang D, Ogihara M, Gallo C, Villamar JA, Smith JD, Vermeer W, et al. Automatic classification of communication logs into implementation stages via text analysis. Implementation Sci. 2015, Sep, 6;11(1):119. https://doi.org/10.1186/s13012-016-0483-6.
86. Ye J, Bannon J, Kho A, Smith JD, Walunas T. Leveraging machine learning approach to identify relationships between practice facilitation strategies and practice characteristics based on the implementation research logic model. Implement Sci Commun. 2025, Dec, 26;7(1). https://doi.org/10.1186/s43058-025-00850-6.
87. Zhong D, Wu Y, Aarons GA, Hutchinson AM, Wong WC, Lv S, et al. Implementability of clinical practice guidelines: the review and development of a comprehensive framework for guideline implementability (CFGI). Implement Sci Commun. 2025, Oct, 16;6(1):106. https://doi.org/10.1186/s43058-025-00780-3.
88. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. Implementation Sci. 2022, Oct, 29;17(1):75. https://doi.org/10.1186/s13012-022-01245-0.
89. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implementation Sci. 2012, Apr, 24;7(1):37. https://doi.org/10.1186/1748-5908-7-37.
90. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the expert recommendations for implementing change (ERIC) project.

Implementation Sci. 2015, Feb, 12;10(1):21. https://doi.org/10.1186/s13012-015-0209-1.
91. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011, Mar;38(2):65–76. https://doi.org/10.1007/s10488-010-0319-7.
92. Olsen L AD, Jm M. Institute of Medicine The learning healthcare system: workshop summary. National Academies Press; 2007.
93. Langley Gl MR, Nolan KM, Nolan TW, Norman CL, Provost LP. The improvement guide: a Practical approach to Enhancing organizational performance. 2. Jossey-Bass Publishers; 2009.
94. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: normalization process theory. Implementation Sci. 2009, May, 21;4(1):29. https://doi.org/10.1186/1748-5908-4-29.
95. Wensing M, Oxman A, Baker R, Godycki-Cwirko M, Flottorp S, Szecsenyi J, et al. Tailored implementation for chronic diseases (TICD): a project protocol. Implementation Sci. 2011, Sep, 7;6(1):103. https://doi.org/10.1186/1748-5908-6-103.
96. Moullin JC, Sabater-Hernandez D, Benrimoj SI. Model for the evaluation of implementation programs and professional pharmacy services. Res Soc Adm Pharm. 2016, May-Jun;12(3):515–22. https://doi.org/10.1016/j.sapharm.2015.08.003.
97. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implementation Sci. 2013, Oct, 2;8(1):117. https://doi.org/10.1186/1748-5908-8-117.
98. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, 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. https://doi.org/10.3389/fpubh.2019.00064.
99. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. Adm Policy Ment Health. 2011, Jan;38(1):4–23. https://doi.org/10.1007/s10488-010-0327-7.
100. Shalong W, Yi Z, Bin Z, Ganglei L, Jinyu Z, Yanwen Z, et al. Enhancing self-directed learning with custom GPT AI facilitation among medical students: a randomized controlled trial. Med Teach. 2025, Jul;47(7):1126–33. https://doi.org/10.1080/0142159X.2024.2413023.
101. Zhao B, Zhang X, Chen M, Wang R, Tian T, Yan Z, et al. A novel approach for recommending doctors in online health communities based on the coordination of deep learning and reinforcement learning. Decis Support Syst. 2026;205:205. https://doi.org/10.1016/j.dss.2026.114656.
102. Hu Y, Xuan W, Zhou Q, Li Z, Li Y, Hu J, et al. A self-correcting agentic graph RAG for clinical decision support in hepatology. Front Med. 2025;12:1716327. https://doi.org/10.3389/fmed.2025.1716327.
103. Gaber F, Shaik M, Allega F, Bilecz AJ, Busch F, Goon K, et al. Evaluating large language model workflows in clinical decision support for triage and referral and diagnosis. NPJ Digit Med. 2025, May, 9;8(1):263. https://doi.org/10.1038/s41746-025-01684-1.
104. Komenda A, Makowski M, Can E, Prucker P, Busch F, Wachter A, et al. Development and evaluation of a retrieval-augmented generation system for radiology guidelines. J Digit Imag. Inf. Med. 2026, Feb, 12. https://doi.org/10.1007/s10278-025-01835-6.
105. Do TT, Martyn M, McEwen A, Gaff C, McClaren BD. Relational work in implementation: a qualitative analysis of intra- and inter- professional strategies leveraged in genomic multidisciplinary teams. Implement Sci Commun. 2025, Dec, 16;6(1):137. https://doi.org/10.1186/s43058-025-00828-4.
106. 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. Implementation Sci. 2025;20(1). https://doi.org/10.1186/s13012-025-01436-5.
107. Hassan M, Kushniruk A, Borycki E. Barriers to and facilitators of artificial intelligence adoption in health Care: scoping review. JMIR Hum Factors. 2024, Aug, 29;11:e48633. https://doi.org/10.2196/48633.
108. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. 2025, Feb, 5;388:e081554. https://doi.org/10.1136/bmj-2024-081554.
109. Tun HM, Rahman HA, Naing L, Malik OA. Trust in artificial intelligence–based clinical decision support systems among health care workers: systematic review. J Med Internet Res. 2025, Jul, 29;27:e69678. https://doi.org/10.2196/69678.
110. International Telecommunication Union. ITU-T L. 1801: guidelines for assessing the environmental impact of artificial intelligence systems. https://www.itu.int/rec/T-REC-L.1801-202602-I?utm_source=chatgpt.com.
111. Ramachandran A, Sarabu C, Gupta U, Ghose S, Lee VS. Sustainably advancing health AI: a decision framework to mitigate the energy, emissions, and cost of AI implementation. NEJM Catalyst. 2025;6(10). https://doi.org/10.1056/cat.25.0125.
112. Flemyng E, Noel-Storr A, Macura B, Gartlehner G, Thomas J, Meerpohl JJ, et al. Position statement on artificial intelligence (AI) use in evidence synthesis across Cochrane, the Campbell collaboration, JBI, and the collaboration for environmental evidence 2025. Campbell Systematic Rev. 2025, Dec;21(4):e70074. https://doi.org/10.1002/cl2.70074.
113. 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.
114. Tang L, Sun Z, Idnay B, Nestor JG, Soroush A, Elias PA, et al. Evaluating large language models on medical evidence summarization. NPJ Digit Med. 2023, Aug, 24;6(1):158. https://doi.org/10.1038/s41746-023-00896-7.
115. Aydin S, Karabacak M, Vlachos V, Margetis K. Large language models in patient education: a scoping review of applications in medicine. Front Med. 2024;11:1477898. https://doi.org/10.3389/fmed.2024.1477898.
116. Giorgetti C, Giorgetti A, Pelotti S, Contissa G. AI and shared decision-making: a systematic review. AI Soc. 2026. https://doi.org/10.1007/s00146-026-02955-5.
117. Fontaine G, Vinette B, Weight C, Maheu-Cadotte M-A, Lavallée A, Deschênes M-F, et al. Effects of implementation strategies on nursing practice and patient outcomes: a comprehensive systematic review and meta-analysis. Implementation Sci. 2024, Sep, 30;19(1):68. https://doi.org/10.1186/s13012-024-01398-0.
118. Huynh AL, Roy TJ, Jackson KN, Lee AG, Liaw W, Hossain MM. Applications of artificial intelligence-based conversational agents in healthcare: a systematic umbrella review. Int J Med Inf. 2026, Mar, 1;207:106204. https://doi.org/10.1016/j.ijmedinf.2025.106204.
119. Singh B, Olds T, Brinsley J, Dumuid D, Virgara R, Matricciani L, et al. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. NPJ Digit Med. 2023, Jun, 23;6(1):118. https://doi.org/10.1038/s41746-023-00856-1.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.