Evidence Matters

LEARNING PAPER

An Implementation Science Framework for Embedding AI–Based Clinical Decision Support Tools in Primary Health Care Systems

Evidence from an AI Ultrasound Pilot in Sindh, Pakistan

About the learning paper

An Implementation Science Framework for Embedding AI-Based Clinical Decision Support Tools in Primary Health Care Systems: Evidence from an AI Ultrasound Pilot in Sindh, Pakistan is a learning paper developed from implementation evidence generated through a collaborative AI-enabled ultrasound pilot in Sindh's public primary healthcare system.

The paper draws on implementation learning from an AI-enabled obstetric ultrasound pilot conducted within public primary healthcare facilities in Sindh. It uses this experience to propose the EMBED-AI Framework, a practical implementation science framework for introducing, embedding, scaling, and sustaining AI-based clinical decision support tools (AI-CDSTs) in public health systems.

The paper is not intended as a product evaluation or a pilot report. Instead, it examines what is required for AI tools to move from technical promise to routine, trusted, and accountable use within public health delivery systems.

Intended audience for the Framework

The EMBED-AI Framework is intended primarily for public health systems, government health departments, and implementation partners responsible for introducing AI-based clinical decision support tools into routine care. They can use it to assess whether an AI tool addresses a real service delivery gap, fits provider workflows, supports safe clinical decision-making, links to referral and data systems, and can be governed, financed, and sustained beyond a pilot.

Secondary users include donors, product partners, and evaluators, who can use the framework to assess implementation readiness, design tools for real-world use, and measure adoption, safety, equity, and sustainability alongside clinical outcomes.

Acknowledgement

This learning paper reflects the collective efforts of the partners involved in the AI-enabled ultrasound pilot in Sindh. We are grateful to **the Gates Foundation** for funding the pilot and enabling the generation of implementation evidence.

We extend our sincere appreciation to **People's Primary Healthcare Initiative (PPHI) Sindh** for its leadership, facility network, frontline teams, district coordination, and operational stewardship, without which the pilot would not have been possible.

We acknowledge **the Sindh Social Protection Authority (SSPA)** for linking the intervention with the Mamta programme, supporting the broader maternal health and social protection pathway, and providing data for the evaluation. We recognise **Delft Imaging** for providing the BabyChecker AI-enabled ultrasound technology and supporting deployment, product refinement, local validation, and scan data. Most importantly, we thank the healthcare providers, facility staff, district teams, pregnant women, and families who participated in and supported the pilot. Their experience, feedback, and trust shaped the learning reflected in this paper.

We also gratefully acknowledge the contribution of the **CERP project team**, whose work across project design, implementation planning, field coordination, data systems, monitoring, and learning was essential to the successful rollout of the pilot. Their collective effort helped translate the intervention from concept into practice within routine public-sector primary healthcare settings. The CERP team included Imran Zia, EVP; Dr. Khizar Nasir, Associate Director; Sidra Bashir, Associate Director; Muhammad Zia Muneer, Program Manager; Haseeb Sajid, Project Manager; Muhammad Tahir, Senior Associate; and Ali Zulkifal Ahmed, Analyst.

About the Author

Ali Gardezi

Ali Gardezi is Director at the Centre for Economic Research in Pakistan (CERP). He has over 15 years of experience in public-sector reform, digital transformation, evidence-based delivery, and development consulting. His recent work focuses on the design, implementation, and evaluation of technology-enabled and AI-supported interventions in low-resource public systems. Before CERP, Ali worked with McKinsey & Company and Adam Smith International on public-sector reform and strategy advisory. He holds an MBA from Lahore University of Management Sciences (LUMS), with an exchange term at Tsinghua University, Beijing.

About CERP

The Centre for Economic Research in Pakistan (CERP) is an independent research, analytics, and policy advisory organisation that works to improve public decision-making through evidence, data, technology, and implementation learning. CERP partners with governments, development partners, researchers, and implementing organisations to design, test, evaluate, and scale solutions to complex development challenges.

CERP's work spans health, education, social protection, labour markets, digital systems, governance, and economic development.

Suggested Citation

Gardezi, A. (2026). An implementation science framework for embedding AI-based clinical decision support tools in Primary Health Care Systems: Evidence from a maternal health AI ultrasound pilot in Pakistan. Centre for Economic Research in Pakistan.

Table of Contents

Executive Summary	4
Why Implementation Science Matters for AI in Health	6
What Existing Implementation Science Frameworks Offer	7
Learning From the AI Ultrasound Pilot in Sindh	8
The EMBED-AI Framework	9
Implications for Governments, Implementers, Product Partners, & Donors	23
Annex 1: Implementation Science Frameworks and Their Relevance	24
Annex 2: Learnings From The AI Ultrasound Pilot in Sindh	27
Annex 3: Implementation Readiness Matrix	32
Contributors & Reviewers	34

Executive Summary

Artificial intelligence is increasingly viewed as a promising tool for strengthening health systems in low- and middle-income countries. AI-based clinical decision support tools can help extend scarce expertise, support frontline providers, improve diagnostic access, strengthen risk identification, and enable more timely referral and follow-up (Agweyu et al., 2026, pp. 607–608; Independent International Scientific Panel on AI, 2026, p. 21; World Health Organisation, 2024, pp. 1, 3–4).[1][2][3] These opportunities are particularly relevant in low-resource public health systems, where there is significant potential to enhance specialist access, expand diagnostic capacity, strengthen referral pathways, improve data integration, and optimise the use of available resources to support higher quality and more continuous care (Ciecierski-Holmes et al., 2022; World Health Organisation, 2024, pp. 1, 8; Independent International Scientific Panel on AI, 2026, pp. 20–21, 31).[4][2][3]

However, the movement from AI innovation to public health impact is not automatic. Many AI tools show strong technical potential but do not become routinely used, trusted, or sustainably integrated into healthcare delivery because real-world impact depends on workflow fit, clinical uptake, health-system readiness, user trust, governance, and sustainability (Khokhar & Mateen, 2026, pp. 5, 21; El Arab et al., 2025, p. 1).[5][6] The reason is rarely model performance alone. AI tools often struggle because they are introduced as technologies rather than as health-system interventions. They may not fit provider workflows, may not be trusted by frontline users, may lack clear clinical governance, may not connect to referral and follow-up systems, or may generate data that is not integrated into routine monitoring and decision-making (Hassan et al., 2024; Khokhar & Mateen, 2026, pp. 21–22).[5][7]

This learning paper tries to address that implementation gap. It draws on the field learning from an AI-enabled obstetric ultrasound pilot implemented in public primary healthcare facilities managed by PPHI in Sindh, Pakistan. The pilot was conducted through a collaboration among CERP, PPHI Sindh, SSPA, and Delft Imaging, with support from the Gates Foundation. The intervention introduced BabyChecker, a portable AI-enabled ultrasound device, within the Mamta maternal health platform to improve pregnancy screening, counselling, referral, and implementation learning. The Mamta program[8] is a maternal and child health initiative implemented by PPHI Sindh in collaboration with the Sindh Social Protection Authority (SSPA), with financial and technical support from the World Bank.

The paper does not present itself as a product evaluation. Instead, it uses the pilot as an empirical case to examine a broader implementation science question: **what does it take for an AI-based clinical decision support tool to become embedded within routine public health service delivery?**

The pilot generated several important lessons. AI-enabled ultrasound created the greatest value where it addressed a real diagnostic access gap, particularly in facilities without conventional ultrasound capacity. It supported clinical decision-making but did not replace provider judgement. Frontline providers were able to use the tool, but adoption depended on practical training, supervision, trust-building, troubleshooting, and workflow fit. Women valued ultrasound when it produced clear and actionable information, but continued care-seeking required counselling, family engagement, follow-up, and support to overcome demand-side barriers. Detection of risk created value only when connected to referral completion, feedback, and outcome tracking. Operational basics such as device charging, internet connectivity, gel supply, storage, troubleshooting, and data upload were central implementation conditions rather than minor logistical details.

These findings are consistent with the implementation science literature. Existing frameworks such as CFIR, RE-AIM, EPIS, NPT, NASSS, PRISM, the Behaviour Change Wheel, the Dynamic Sustainability Framework, Proctor's Implementation Outcomes Framework, and i-PARIHS provide important insights into context, adoption, routinisation, technology scale-up, behaviour change, sustainability, and implementation outcomes. However, they do not fully address the distinctive requirements of AI-based clinical decision support tools, particularly in low-resource health systems. AI-CDSTs are dynamic, data-dependent, trust-sensitive, clinically consequential, and dependent on downstream system action. They require not only adoption but also explainability, local validation, human-in-the-loop governance, workflow-data integration, product adaptation, and responsible institutionalisation.

To address this gap, the paper proposes the EMBED-AI Framework, an implementation science framework for embedding AI-CDSTs in low-resource health systems.

The central contribution of EMBED-AI is that it connects product, provider, patient, pathway, data, and institution into a single implementation logic. It reframes AI adoption as a system change process rather than a technology rollout. The framework argues that an AI tool should be scaled only when the system around it is ready, when the clinical use case is clear, providers are supported, workflows are coherent, referral pathways can respond, operational support is available, and institutional accountability is defined.

The broader message of the paper is that AI in health should not be judged only by whether a model can generate an output. The more important question is whether an AI-supported decision process can be safely, equitably, and sustainably embedded into routine public health service delivery. In the LMIC context, this requires implementation science, institutional partnership, and disciplined attention to the realities of care (OECD, 2026).[9]

The framework consists of seven stages:

	Stage	Description
E	**Establish the clinical and system need**	Define the priority clinical decision, service-delivery gap, target population, and why AI adds value.
M	**Map workflows and routines**	Understand how patients, providers, data, referrals, and operational routines currently move through the system.
B	**Build trust, capability, and governance**	Train and mentor providers, validate the tool locally, explain AI outputs clearly, and define clinical oversight for safe use.
E	**Embed into care pathways**	Convert AI outputs into action by linking them to counselling, provider interpretation, documentation, referral, follow-up, and care completion.
D	**Design operational support for routine use**	Put in place the practical systems needed for daily use: troubleshooting supplies, maintenance, supervision, , workload management, and incentives.
A	**Adapt through learning loops**	Use provider feedback, patient experience, dashboard data, validation findings, and field learning to refine the tool and workflow over time.
I	**Institutionalise for scale and sustainability**	Formalise ownership, workforce support, monitoring, financing, governance and scale-up criteria so the intervention becomes part of the health system.



[1] Agweyu, A., Mwaniki, P., Musau, W., Korom, R., Isaaka, L., Wanyama, C., Kiptinness, S., Adan, N., Emmanuel-Fabula, M., & Mateen, B. A. (2026). Safety of a large language model-based clinical decision support system in African primary healthcare. *Nature Health, 1*, 607–618. https://doi.org/10.1038/s44360-026-00082-5
[2] Independent International Scientific Panel on AI. (2026). *Preliminary report of the Independent International Scientific Panel on AI: Evidence-based assessment of opportunities, risks and impacts of artificial intelligence*. United Nations.
[3] World Health Organisation. (2024). *The role of artificial intelligence in sexual and reproductive health and rights: Technical brief*. World Health Organisation. https://www.who.int/publications/i/item/9789240090705
[4] Ciecierski-Holmes, T., Singh, R., Axt, M., Brenner, S., & Barteit, S. (2022). Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: A systematic scoping review. *npj Digital Medicine, 5*, 162. https://doi.org/10.1038/s41746-022-00700-y
[5] Khokhar, T., & Mateen, B. A. (2026). *Asking what matters: A practical guide to AI for health programme appraisal* (1st ed.). K4DD Learning Product. Institute of Development Studies. doi:10.19088/K4DD.2026.035
[6] El Arab, R. A., Abu-Mahfouz, M. S., Abuadas, F. H., Alzghoul, H., Almari, M., Ghannam, A., & Seweid, M. M. (2025). Bridging the gap: From AI success in clinical trials to real-world healthcare implementation—A narrative review. *Healthcare, 13*(7), 701. doi:10.3390/healthcare13070701
[7] Hassan, M., Kushniruk, A., & Borycki, E. (2024). Barriers to and facilitators of artificial intelligence adoption in health care: Scoping review. *JMIR Human Factors, 11*, e48633. doi:10.2196/48633
[8] Mamta Program (Mother and Child Support Program) managed by SSPA
[9] OECD. (2026). *Scaling artificial intelligence in health*. OECD Publishing. doi:10.1787/a436e12d-en

Why Implementation Science Matters for AI in Health?

AI in health is often discussed through the language of prediction, automation, accuracy, sensitivity, specificity, and efficiency. These concepts are important, but they do not by themselves explain whether an AI tool will improve health outcomes in real-world systems. An AI model may perform well in validation studies yet fail to generate impact if it is not adopted by providers, trusted by patients, integrated into workflows, supported by data systems, connected to referral pathways, or sustained by institutions (Khokhar & Mateen, 2026, pp. 20–22; El Arab et al., 2025; Greenhalgh et al., 2017) [5] [6] [10]

Implementation science is therefore essential for AI in health. It studies how evidence-based interventions are adopted, integrated, adapted, scaled, and sustained in real-world settings (Chu, 2024).[11] It shifts the focus from whether an intervention works under ideal conditions to how it can work within complex systems, for real users, under routine constraints. This distinction is particularly important in low-resource health systems, where implementation is shaped by contextual factors such as, availability of diagnostics, data systems, workforce capacity, referral pathways, levels of digital maturity, connectivity conditions, and financing environments.

AI-based clinical decision support tools create a distinctive implementation challenge because they operate at the point where information becomes action. They do not simply collect data or deliver messages. They generate classifications, alerts, predictions, or recommendations that may influence clinical judgement, patient counselling, referral, treatment, or follow-up. This means that the implementation pathway is not complete when the AI output appears. The output must be understood by the provider, communicated to the patient, linked to an appropriate clinical action, documented in the system, and followed through until the outcome is known (Khokhar & Mateen, 2026, pp. 11, 20–22).[5]

The AI-enabled ultrasound pilot in Sindh illustrates this clearly. The BabyChecker device could generate automated obstetric assessments, but its health-system value depended on whether providers could use the tool within ANC workflows, whether risk classifications were clinically meaningful, whether women received understandable counselling, whether referral slips were issued, whether follow-up occurred, and whether data were linked to maternal health records. The intervention was therefore not only a diagnostic technology. It was also a workflow intervention, provider adoption intervention, referral intervention, data intervention, and institutional learning intervention.

Implementation science also identifies & manages implementation risks by examining barriers, facilitators, contextual determinants, and system-level conditions that influence adoption, scale-up, and sustainability (Nilsen & Bernhardsson, 2019).[12] AI-CDSTs can create unintended consequences if introduced without adequate system design and governance. False high-risk flags may generate anxiety or unnecessary referral. False low-risk outputs may create inappropriate reassurance. Unclear recommendations may weaken provider trust. Poor workflow fit may lead to inconsistent use. Weak data governance may expose patients and providers to privacy risks. These risks do not imply that AI should be avoided; they imply that AI must be implemented carefully (Wells et al., 2025).[13]

Existing implementation science frameworks offer complementary lenses for understanding and guiding implementation, including determinants of success, stages of implementation, integration into routine practice, challenges to scale and sustainability, and the need to measure implementation outcomes alongside clinical outcomes. However, AI-CDSTs require an additional implementation layer. They require explainability, local validation, product update processes, interoperability, human-in-the-loop governance, and closed-loop referral systems. For AI tools that identify risk, the intervention is incomplete until the system acts on that risk and verifies what happened next.

This is why the EMBED-AI Framework is needed. It translates implementation science into a practical pathway for AI-CDSTs in health systems. It recognises that AI impact depends on the alignment of six elements: the product, the provider, the patient, the care pathway, the data system, and the institution. Implementation science matters for AI because the central challenge is not whether AI can generate a decision-support output. The central challenge is whether that output can be safely, meaningfully, and sustainably embedded into the real decisions, routines, relationships, and responsibilities of a health system.

[10] Greenhalgh, T., Wherton, J., Papoutsi, C., Lynch, J., Hughes, G., A'Court, C., Hinder, S., Fahy, N., Procter, R., & Shaw, S. (2017). Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *JMIR Publications*; *19*(11), e367. https://doi.org/10.2196/jmir.8775

[11] Chu, F. (2024). Implementation science: Why should we care? *Journal of the Medical Library Association, 112*(3), 281–285. https://doi.org/10.5195/jmla.2024.1919

[12] Nilsen, P., & Bernhardsson, S. (2019). Context matters in implementation science: A scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. *BMC Health Services Research, 19*, 189. https://doi.org/10.1186/s12913-019-4015-3

[13] Wells, B. J., Nguyen, H. M., McWilliams, A., Pallini, M., Bovi, A., Kuzma, A., Kramer, J., Chou, S.-H., Hetherington, T., Corn, P., Taylor, Y. J., Cuison, A., Gagen, M., Isreal, M., & FAIR-AI Consortium. (2025). A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare. *npj Digital Medicine, 8*, 514. doi:10.1038/s41746-025-01900-y

What Existing Implementation Science Frameworks Offer?

Implementation science offers a rich set of frameworks for understanding why evidence-based interventions succeed, fail, scale, or fade in real-world settings. For AI-enabled health interventions, these frameworks are useful because they shift attention away from technology alone and toward context, users, organisations, workflows, governance, behaviour, data systems, implementation quality, and sustainability. (Nilsen, 2015; Proctor et al., 2011).[14] [15]

Several established frameworks are particularly relevant for understanding AI-CDSTs. The Consolidated Framework for Implementation Research (CFIR), the Practical, Robust Implementation and Sustainability Model (PRISM), and the Context Compass Framework help examine context, readiness, organisational fit, recipient perspectives, infrastructure, and sustainability conditions. The Exploration, Preparation, Implementation, Sustainment (EPIS) Framework provides a useful lifecycle perspective by showing that implementation should move deliberately from problem exploration and preparation to implementation and sustainment. Normalization Process Theory (NPT) and the Behaviour Change Wheel/COM-B Model help explain why provider adoption depends on sense-making, capability, opportunity, motivation, and the ability to integrate new practices into routine clinical work.

Other frameworks add important complementary insights. The NASSS Framework is especially relevant for AI and digital health because it highlights the risks of non-adoption, abandonment, and failed scale-up when technology complexity is underestimated. The Dynamic Sustainability Framework (DSF) shows that sustainment requires adaptation and learning rather than static replication. Proctor's Implementation Outcomes Framework clarifies how implementation success should be measured through outcomes such as acceptability, adoption, appropriateness, feasibility, fidelity, cost, penetration, and sustainability. The integrated Promoting Action on Research Implementation in Health Services (i-PARIHS) Framework reinforces the importance of facilitation, particularly through training, supervision, mentoring, troubleshooting, and stakeholder engagement. The What Works Hub Implementation Science Framework adds a useful public-sector scale lens by linking evidence generation, policy translation, delivery systems, practice at scale, and equity.

Taken together, general implementation science and emerging AI-health guidance address many of the conditions required for responsible AI-CDST implementation, but from different perspectives. Implementation science provides strong lenses on context, adoption, organisational change, implementation processes, adaptation, and sustainability, while AI-health guidance gives greater attention to clinical evaluation, safety, explainability, human oversight, governance, and lifecycle monitoring. The remaining challenge is not the absence of these individual concepts, but their integration into the service-delivery pathway through which an AI output must be understood, acted upon, connected to appropriate care, monitored, and sustained within routine health systems. (World Health Organization, 2021; World Health Organization, 2024).[16] [17]

EMBED-AI addresses this integration gap by organising the most relevant implementation and AI-health considerations around an end-to-end pathway for routine service delivery. It connects clinical need and system readiness with provider trust and capability, workflow integration, clinical governance, closed-loop referral and follow-up, operational support, continuous learning, and institutional scale-up. Its contribution is therefore not the introduction of entirely new implementation concepts, but their integration into a practical pathway through which AI-supported outputs can translate into clinical action, patient follow-up, outcome verification, system learning, and sustained public health service delivery.

In this sense, EMBED-AI reframes the implementation question. The issue is not only whether an AI model works, or whether an AI product can be deployed. The more important question is whether an AI-supported decision process can be safely, equitably, and sustainably embedded into routine public health service delivery. It is this gap between technical capability and integrated, routine service delivery that EMBED-AI is designed to address.

A detailed review of the implementation science frameworks, including their origin, main focus and relevance to EMBED-AI, is provided in Annexure 1: Implementation Science Frameworks and Their Relevance to EMBED-AI.

[14] Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science, 10*, 53. https://doi.org/10.1186/s13012-015-0242-0
[15] Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7
[16] World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. World Health Organization.
[17] World Health Organization. (2024). *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*. World Health Organization.

What the AI Ultrasound Pilot Revealed About AI–CDST Implementation?

The AI-enabled ultrasound pilot in Sindh was a Gates Foundation-funded implementation and learning initiative designed to assess whether AI-enabled handheld obstetric ultrasound could improve antenatal care uptake, early pregnancy risk identification, and referral pathways for high-risk pregnancies in public primary healthcare settings. The pilot was implemented through a collaboration among PPHI Sindh and SSPA as implementation partners, CERP as the research partner, and Delft Imaging as the product partner. It used a two-arm stepped-wedge randomised controlled design across public primary health facilities linked to the Mamta maternal health platform.

The pilot tested BabyChecker, a portable AI-enabled obstetric ultrasound device designed for frontline healthcare providers with limited ultrasound training. While the intervention aimed to strengthen maternal health service delivery, its broader value was in showing that AI tools do not create impact through technical performance alone. They become useful when they address a real service delivery gap, fit into provider workflows, support clinical judgement, strengthen patient counselling, connect risk identification to referral completion, and operate within strong institutional and data systems.

These pilot learnings informed the development of the EMBED-AI Framework. The detailed learning from the pilot is provided in Annexure 2. The twelve core learnings are listed below:

1



AI ultrasound added the most value in facilities where routine ultrasound services or expertise were unavailable

2



Clinical interpretation remained essential to every AI-supported decision

3



Local validation study, transparent communication, and hands-on support increased provider trust by reducing uncertainty

4



Successful frontline adoption required hands-on training, supervision, and workflow integration

5



Pregnant women were more likely to take action when providers explained the scan results clearly

6



AI created a meaningful impact when the intervention was embedded into clinical workflows

7



Simple referral tools strengthened the pathway from risk identification to care completion

8



Product adaptation strengthened implementation by turning feedback from the field into app changes

9



Operational readiness influenced uptake as much as technical performance

10



Provider continuity required onboarding, refresher training, and workload management

11



Data integration and advanced analytics created visibility across the full care pathway from scan to outcome

12



Strong public-sector ownership and institutional support from PPHI

The EMBED–AI Framework

The EMBED-AI Framework is a practical implementation science framework for embedding AI-based clinical decision support tools (AI-CDSTs) into low-resource primary health care systems. It was developed by triangulating three sources of learning: established implementation science frameworks, field learning from the AI-enabled ultrasound pilot in Sindh, and practical experience of introducing an AI-supported decision tool into routine antenatal care services delivered through public primary healthcare facilities.

The framework begins from a central implementation insight: AI creates health impact only when it becomes part of routine service delivery. A technically strong AI model may classify risk, generate predictions, or support diagnosis, but health outcomes improve only when the AI output is interpreted by providers, trusted by patients, connected to counselling and referral pathways, documented in health records, followed up through care systems, and monitored over time. The implementation challenge is therefore not merely to deploy an AI product, but to embed an AI-supported decision process into the everyday work of a health system.

The framework is designed for AI-CDSTs in low- and middle-income countries, where health systems are resource constrained. In these contexts, AI tools must be implemented as complex service delivery interventions rather than plug-and-play solutions.

The framework is both sequential and iterative

This Framework is built around the idea that AI-CDST impact is produced through a sequence of linked implementation conditions. These conditions can be summarised as an "AI-output-to-health-impact chain":

AI-supported output → provider interpretation → patient counselling → clinical action → referral or treatment → follow-up → outcome verification → system learning.

If any link in this chain is weak, the potential value of the AI tool is reduced. If the provider does not trust the output, the tool may not be used. If the woman does not understand the result, she may not act. If referral facilities do not provide feedback, the system cannot verify whether risk identification led to care completion. If data are not linked, implementation teams cannot distinguish between a product problem, a provider adoption problem, a patient-side barrier, or a referral-system failure.

The stages are presented in a logical order, but implementation rarely proceeds in a straight line. Learning from later stages should feed back into earlier decisions. For example, referral completion data may reveal that the key bottleneck is not risk identification but patient-side constraints; provider feedback may reveal that the AI output is technically valid but difficult to explain; or scale-up costing may show that procurement is affordable, but maintenance and supervision are not. EMBED-AI therefore treats AI implementation as a continuous learning process rather than a one-time rollout.

The framework therefore shifts the unit of analysis from the AI product to the AI-supported care pathway. EMBED-AI argues that AI implementation should not be evaluated only at the point where the algorithm produces an output. It should be evaluated across the full chain through which that output is translated into clinical action and health-system learning.

The framework is organized around patient impact. Its final objective is not adoption of AI for its own sake, but to improve quality of care and outcomes, strengthen system and workforce capacity, make health systems more efficient, and ensure equitable access to quality services.

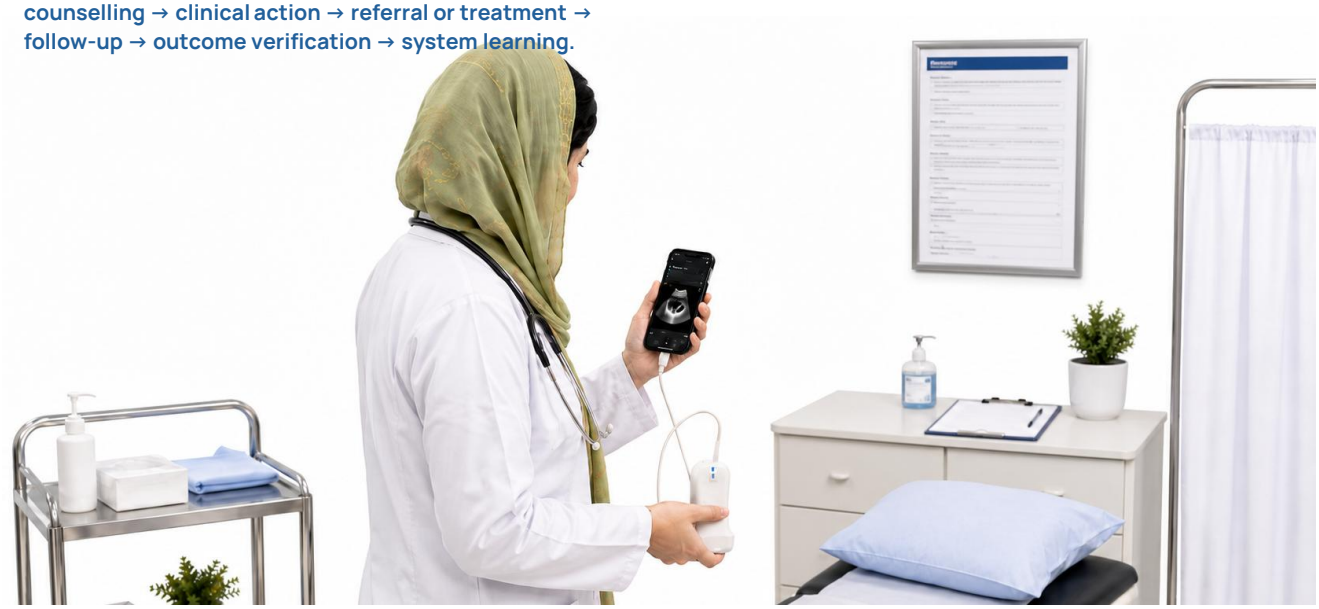
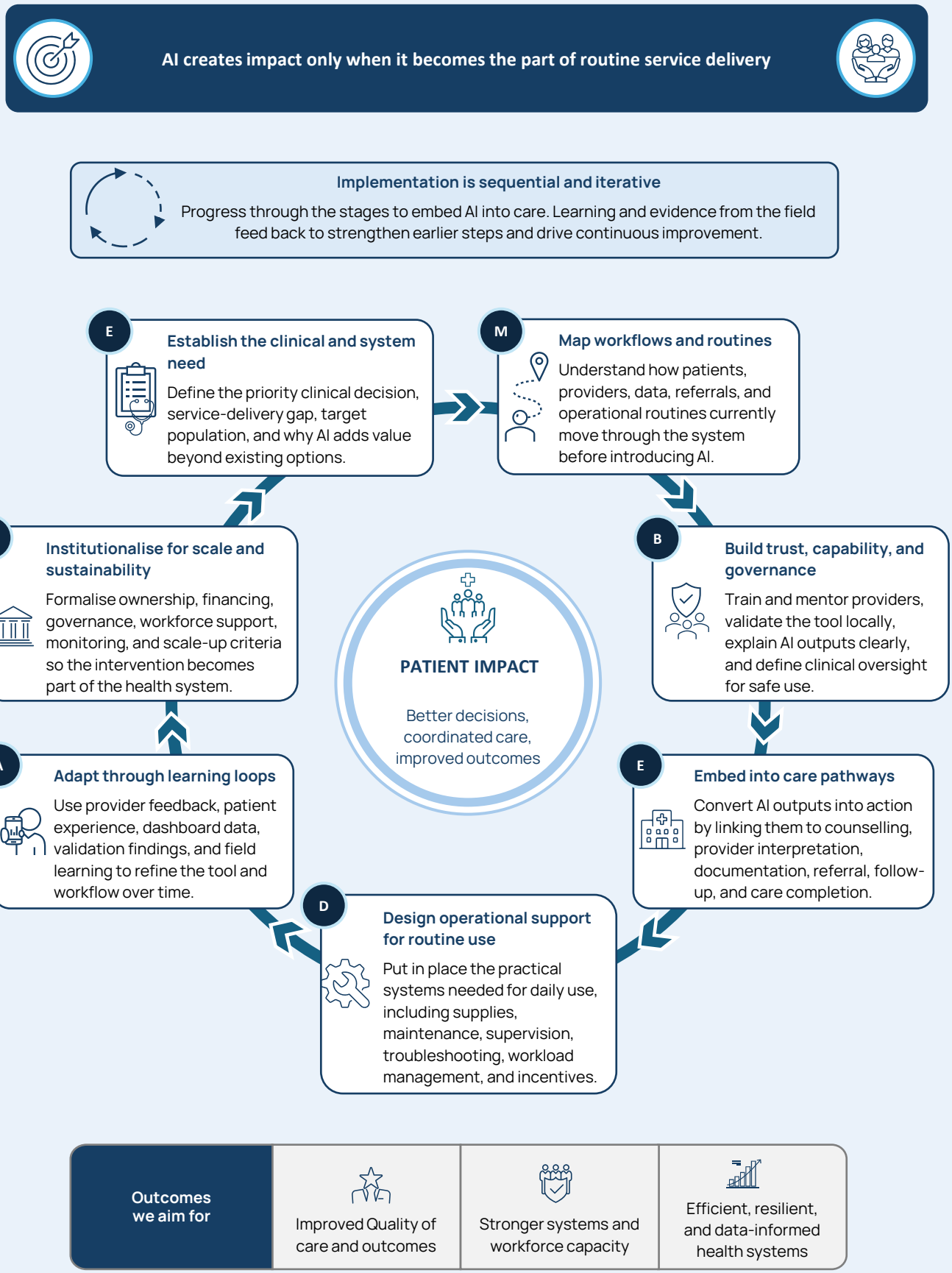


Figure 1: EMBED-AI Framework Illustration



E: Establish the clinical and system need

The first stage asks whether the AI-CDST addresses a real clinical and health-system problem. This stage is deliberately placed before product selection, procurement, or deployment. Many AI health interventions begin with the availability of a tool and then search for an implementation context. EMBED-AI reverses this logic. It begins with the service delivery gap.

In the Sindh pilot, the problem was not simply that an AI ultrasound device existed and needed to be tested. The problem was that many pregnant women had limited access to routine ultrasound services, frontline providers often lacked diagnostic support for early risk identification, and referral pathways for high-risk pregnancies required strengthening. The AI tool was therefore relevant because it addressed a real access gap in antenatal care and supported frontline decision-making where conventional ultrasound expertise was unavailable.

This stage requires clarity on six fronts:

The clinical decision must be defined:

The intervention team should specify whether the AI-CDST supports screening, diagnosis, referral, triage, treatment, counselling, follow-up, or monitoring. A tool that produces a risk score but does not change any clinical action is unlikely to generate meaningful health impact.

The health-system gap must be explicit:

Implementers should define what is currently unavailable, delayed, costly, inconsistent, or inaccessible. In low-resource settings, this may include limited specialist availability, lack of diagnostic equipment, delayed referral, weak records, or poor patient follow-up.

The AI value proposition must be justified against non-AI alternatives:

AI should not be used because it is novel.[5] It should be used because it offers a meaningful advantage over available alternatives, such as extending scarce expertise, reducing travel burden, improving triage, enabling task-sharing, or generating actionable data for follow-up.

The system's ability to act on the AI output must be assessed:

A risk prediction tool has limited value if there is no referral pathway. A diagnostic triage tool may create harm if it identifies high-risk cases without connecting patients to care. A provider-facing tool may reduce efficiency if it duplicates documentation / creates additional workload.[10 30]

The equity and access implications must be clear:

AI interventions in public health should be assessed by whether they expand access for underserved groups or risk deepening existing inequities. In low-resource health systems, the strongest case for AI often lies in extending scarce expertise, bringing diagnostic support closer to communities, and reducing the need for travel.

The theory of change and minimum success criteria must be defined:

Implementers should specify how the AI output is expected to lead to provider action, patient behavior change, referral completion, improved care, or better outcomes. They should also define the minimum conditions under which the intervention should proceed, pause, or be redesigned. This prevents AI deployment from becoming an open-ended pilot and creates a clear basis for learning and accountability.



M: Map workflows and routines

Once the clinical and system need is established, the next step is to understand how care actually works before the AI tool is introduced. This stage maps the everyday routines through which patients enter care, providers make decisions, data are recorded, referrals are made, and operational problems are resolved. In EMBED-AI, workflow mapping is the method through which implementation readiness is assessed. Readiness is not treated as a fixed facility attribute; it is understood as the degree to which an AI tool can be absorbed into existing clinical, operational, data, and governance routines without disrupting care delivery or overburdening providers.

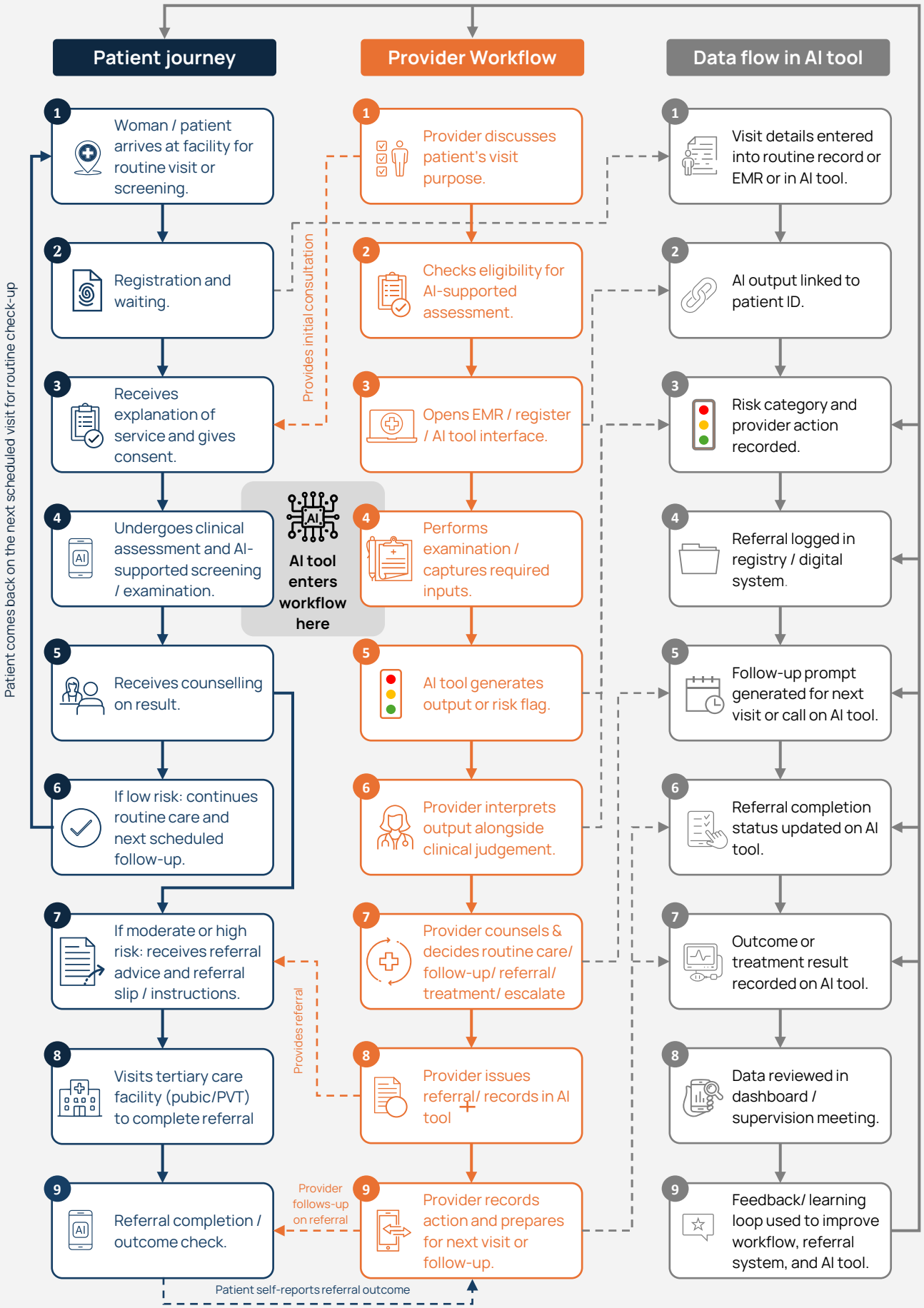
This stage is especially important for AI-CDSTs because these tools do not operate in isolation. They enter a chain of tasks, decisions, relationships, records, and follow-up actions. Mapping workflows and routines helps identify where the AI output will enter the care pathway, who will use it, what information will be needed, how decisions will be documented, and where friction, duplication, delay, or informal workarounds may affect implementation. EMBED-AI therefore proposes a 5P Workflow and Readiness Map.

Mapping workflows and routines ensures that AI implementation is grounded in the real conditions of care delivery rather than in assumed or idealised processes. It helps implementers see whether the AI tool will fill a genuine operational gap, where it may create additional burden, what adaptations are required, and which facilities are ready for early deployment. It also prevents AI from becoming a parallel activity by identifying how the tool can be integrated into existing patient pathways, provider routines, data systems, referral mechanisms, and institutional responsibilities. The output of this stage is a patient-provider-system map that shows patient journey, provide workflow and data flow.

The 5P Workflow and Readiness Map

Domain	What to assess	How to assess
Patient pathway	How patients enter care, receive counselling, understand advice, complete referrals, and return for follow-up. This should include patient-side barriers such as travel, cost, family permission, fear, timing, and trust.	Conduct patient journey mapping. (observe patient flow, interview patients and caregivers, review referral and follow-up records, and examine where patients drop out between screening, referral, and outcome).
Provider workflow	Who will use the AI tool, when it will be used, how long it takes, what tasks are added or removed, how outputs are interpreted, and how decisions are recorded.	Shadow providers during routine service delivery: (conduct workflow observations, interview frontline staff, map task allocation, assess workload implications, and identify where the AI tool may create duplication or efficiency).
Place and facility readiness	Whether the facility has the physical, operational, and logistical conditions needed for routine use, including space, privacy, electricity, internet, supplies, storage, safety, patient flow, and device security.	Conduct facility readiness assessments: (verify infrastructure and supplies, observe room layout and patient flow, check charging and connectivity conditions, review storage arrangements, and identify minimum operational requirements).
Platform and data systems	What data are collected, where they are stored, how records are linked, whether AI outputs can be integrated into existing registers, EMRs, dashboards, referral systems, and reporting routines.	Review existing registers and digital systems: (Map data flows, assess interoperability and identifiers, identify duplicate reporting risks, consult data managers, and define minimum data fields for AI-supported care and follow-up).
Policy, governance, and partnerships	Which protocols, actors, approvals, roles, referral rules, product partner responsibilities, data-sharing arrangements, and accountability structures are needed for implementation.	Conduct stakeholder mapping: (review clinical and operational protocols, clarify partner roles, identify the champions, assess referral governance, examine procurement and approval requirements, and define accountability for training, supervision, data, safety, and product support).

Figure 2: Illustrative patient-provider-system map



B: Build trust, capability, and governance

AI-CDSTs depend on trust. Providers must trust the tool enough to use it, but not so much that they stop exercising clinical judgement. Patients must trust the counselling they receive. Institutions must trust that the tool is safe, valid, accountable, and aligned with clinical protocols. Trust is therefore not a soft implementation issue; it is a core condition for adoption, routine use, and scale.

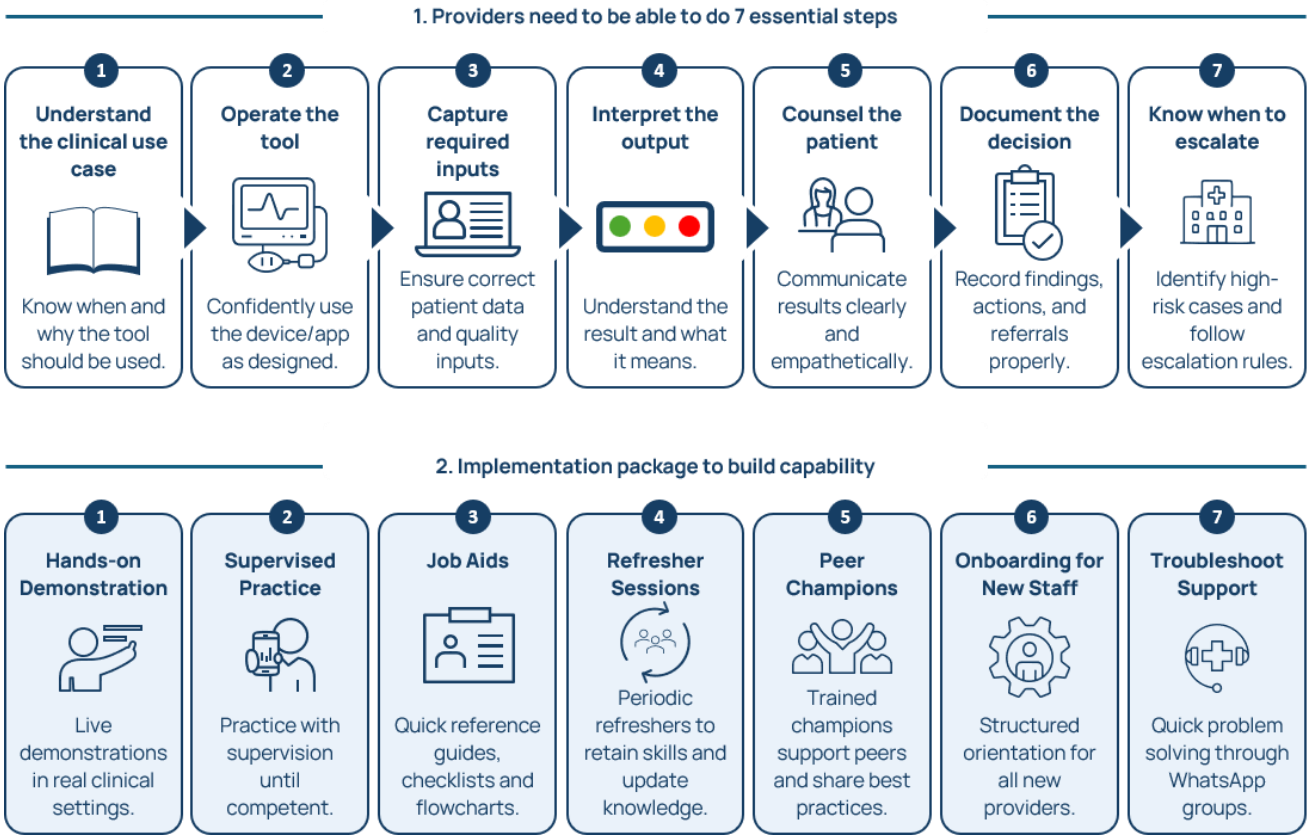
EMBED-AI proposes a Trust Architecture with four components: capability, validation, explainability, and governance. Each one answers a simple question: Can I use it correctly (capability)? Does it actually work here (validation)? Do I understand what it is telling me (explainability)? And is it being used safely and responsibly (governance)? When providers can use the tool well, see that it works in their setting, understand its outputs, and know there are clear rules for safe use, they are more likely to trust it and use it appropriately.

The Trust Architecture

Component	Explanation
a). Capability	Providers have the knowledge, skills, confidence, and practical support needed to use the AI tool correctly within routine care.
b). Validation	The AI tool is tested against an accepted clinical reference standard in the local context, and cases of disagreement are reviewed to assess performance, risk, and limitations.
c). Explainability	Providers and patients can understand what the AI output means, why it matters, and what action should follow.
d). Governance	The AI tool is used within clear clinical, ethical, data, safety, and accountability rules, with defined institutional ownership.

a). Capability: Capability should be built through repeated, practical, role-specific training rather than one-time orientation. Providers need to understand the clinical use case, operate the tool, capture required inputs, interpret the output, counsel the patient, document the decision, and know when to escalate. The implementation package should include hands-on demonstrations, supervised practice, job aids, refresher sessions, peer champions, onboarding for new staff, and rapid troubleshooting support. Capability is strongest when training is embedded into routine supervision and when providers receive feedback on actual use, not only classroom instruction. *(Figure 3: Building Provider Capability)*

Figure 3: Building Provider Capability



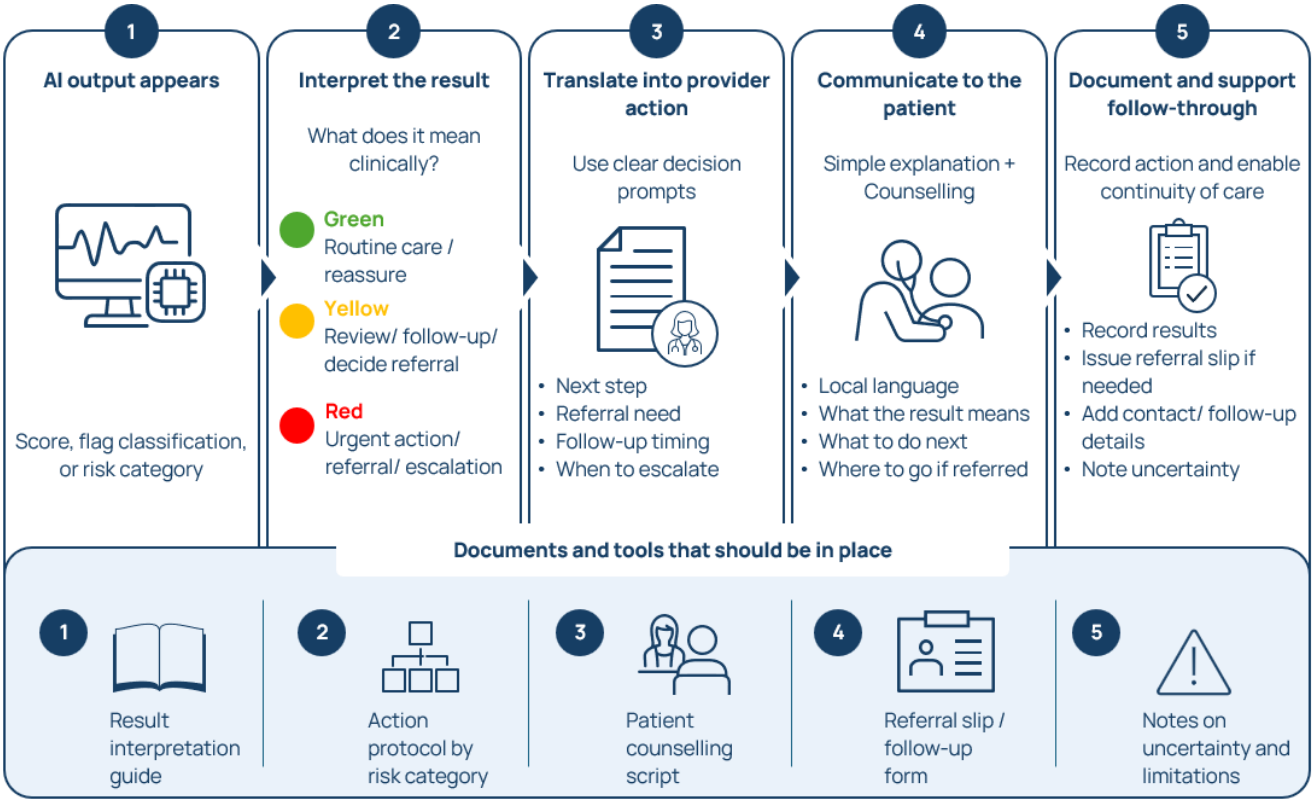
b). Validation: Validation should be treated as a local trust-building process, not only as a technical performance exercise. Before scale-up, the AI tool should be compared with an accepted clinical reference standard in the setting where it will be used. The validation process should define the parameters being assessed, identify the reference standard, ensure correct data linkage, review scan or input quality, and analyse cases of disagreement. The AI ultrasound validation study followed this logic by comparing AI-generated outputs with standard ultrasound across selected clinical parameters in primary healthcare facilities. It also showed why data linkage, date matching, scan quality, and traceability matter for credible validation. Local validation should produce evidence of how the AI tool performs in the real setting, practical guidance on when providers can rely on its output or seek confirmation and clear communication of limitations and uncertainty in clinical care. (Figure 4: Building Local Validation)

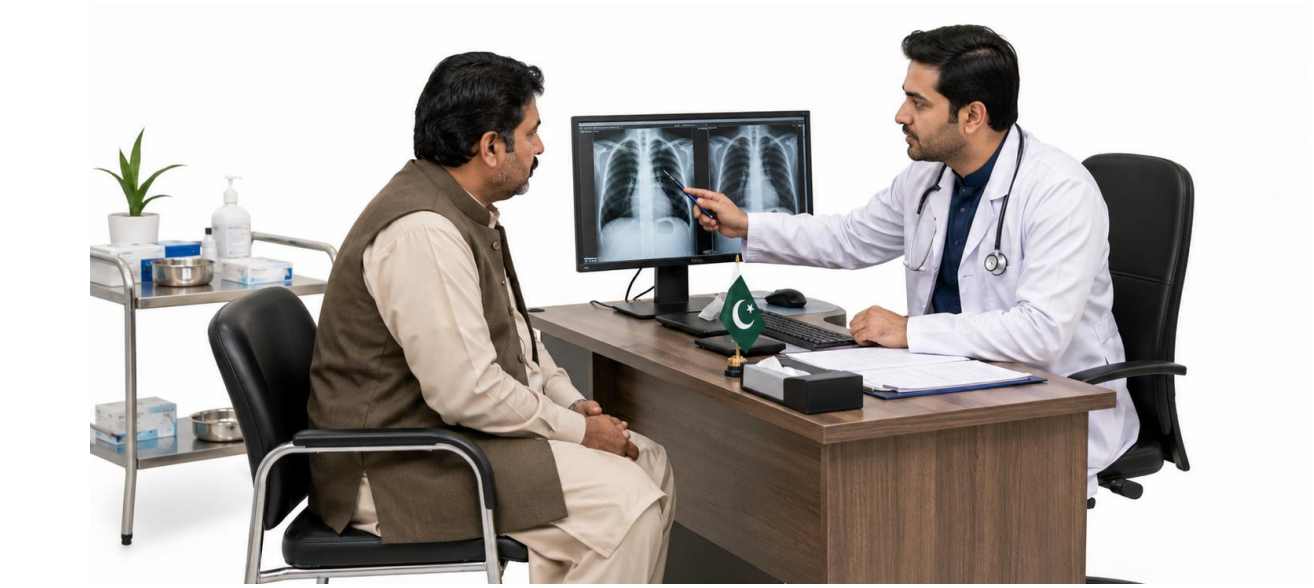
Figure 4: Building Local Validation



c). Explainability: Explainability should convert AI output into clinical meaning and patient action. Providers should not receive an unexplained score, flag, or classification without guidance on what it means and what should happen next. AI outputs should be translated into clear risk categories, simple explanations, clinical decision prompts, and patient counselling messages. Where possible, outputs should be supported by local language explanations, visual risk categories, referral guidance, and notes on uncertainty or limitations. For example, a traffic-light risk result should not only say "green," "yellow," or "red"; it should explain what the category means, what action is recommended, and what the patient should do next. Explainability protects provider confidence, reduces patient anxiety, and prevents both over-referral and false reassurance. (Figure 5: Maing AI Results Explainable and Actionable)

Figure 5: Making AI Results Explainable and Actionable

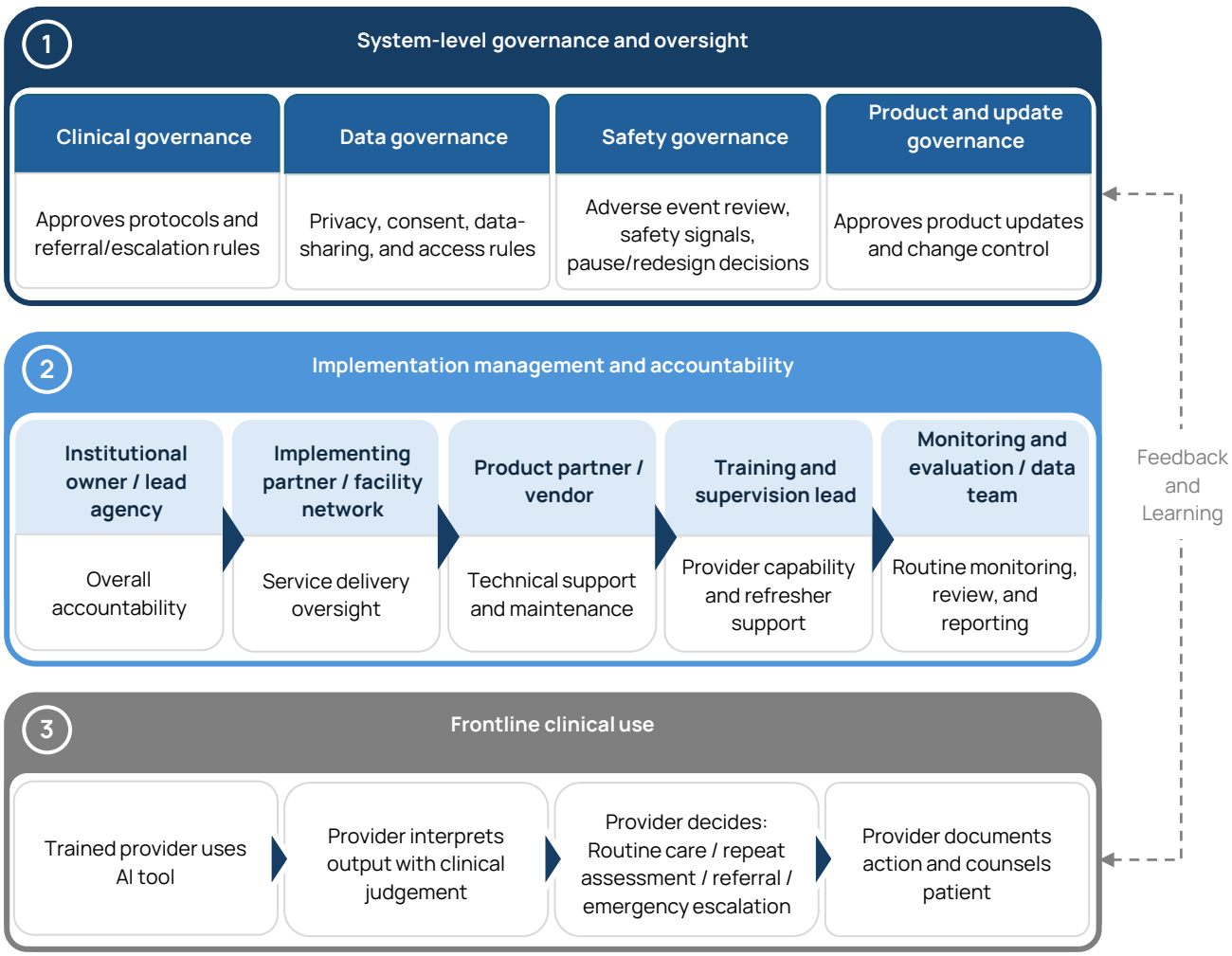




d). Governance: Governance should define who is responsible for safe and appropriate use of the AI tool. AI-CDSTs should support clinical judgement, not replace it. Protocols should clarify that final decisions remain with trained providers and should define when referral, repeat assessment, doctor review, or emergency escalation is required. Governance should also cover informed consent, privacy, data-sharing, safety review, adverse event reporting, product updates, and conditions under which use should be paused or redesigned. Institutional ownership must be explicit: one actor should not be responsible for training, another for data, another for device maintenance, and no one for overall safety. A clear governance structure creates confidence for providers, patients, public institutions, and product partners. (Figure 6: Illustrative Governance Structure for AI-CDSTs)

The four components reinforce one another. Capability without validation can produce confident but unsafe use. Validation without explainability can produce technically accurate but poorly understood outputs. Explainability without governance can produce action but not accountability. Governance without capability can create protocols that are not followed in practice. Trust is therefore built when providers are trained, the tool is locally validated, outputs are understandable, and accountability is clear.

Figure 6: Illustrative Governance Structure for AI-CDSTs



E: Embed into care pathways

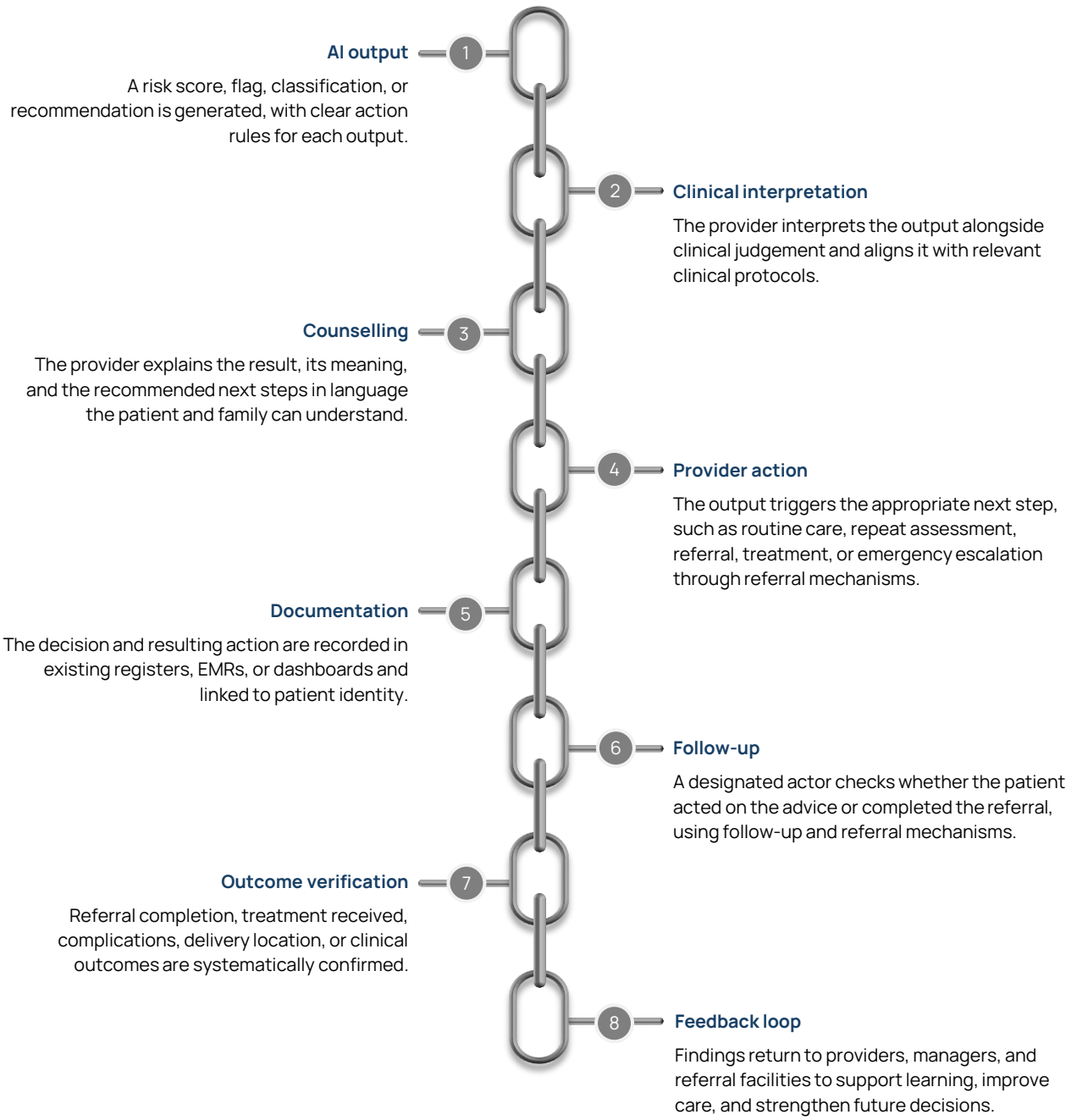
This is the core stage of EMBED-AI. It converts AI from a product into a clinical process. An AI output has limited value unless it triggers the right action and becomes part of the health record and care pathway.

This stage is particularly important in low-resource settings because referral pathways are often fragmented. A provider may identify risk, but the patient may lack transport, family permission, money, information, or confidence. The receiving facility may not share feedback. Records may not link across facilities. Without these connections, AI can improve detection without improving care completion.

The pilot demonstrated that workflow embedding matters more than the device itself. AI Ultrasound became useful when the scan was linked to registration or EMR entry, provider interpretation, counselling, referral slip issuance, maternal health record entry, follow-up prompts, and referral tracking. App modifications that prompted HCPs to issue referral slips, record EMR numbers, explain follow-up steps, and document reasons for non-referral are examples of AI-CDST workflow embedding. For clinical decision support, the critical implementation question is: what happens after the AI result appears?

EMBED-AI proposes an AI-output-to-care-action chain. This chain should be explicitly designed, tested, and monitored.

Figure 7: an AI-output-to-care-action chain



D: Design operational support for routine use

AI implementation often fails not because the technology lacks technical potential, but because the operational system around it, including infrastructure, workflow integration, training, technical support, maintenance, and governance, is too weak to support routine use (Khokhar and Mateen, 2026)[18]. Charging, internet connectivity, device storage, consumables, technical troubleshooting, staff turnover, reporting burden, and supervision can determine whether an AI-CDST is used consistently or quietly abandoned.

The Sindh pilot showed that operational readiness influenced uptake as much as technical performance. Connectivity delays affected verified results and increased administrative burden. Gel supply, charging, safe storage, device handling, spare devices, and theft-risk management were not peripheral concerns; they were implementation conditions. Provider workload also mattered because the tool introduced additional documentation, upload monitoring, referral recording, and follow-up tasks.

EMBED-AI therefore recommends a Minimum Viable Implementation Package before scale-up.

The Minimum Viable Implementation Package

For AI-CDSTs to be used routinely, implementation teams need more than a device, software, or training session. They need a minimum operational package that makes the tool usable within everyday service delivery. The purpose of this package is to prevent AI implementation from depending on ad hoc problem-solving. A pilot may work with intensive project support, but routine use requires defined systems for training, logistics, supervision, data, maintenance, risk response, and financing.

The purpose of this package is to prevent AI implementation from depending on ad hoc problem-solving. A pilot may work with intensive project support, but routine use requires defined systems for training, logistics, supervision, data, maintenance, risk response, and financing. The output of this stage is a complete operational support package that allows the AI-CDST to function as part of routine care rather than as a parallel technology activity.

Component	Minimum implementation requirement
Clinical workflow SOPs	Define eligibility criteria, intended use cases, decision rules, referral criteria, documentation requirements, and escalation protocols. Align SOPs with national/institutional clinical guidelines and clearly specify provider actions for each AI output or risk category to avoid informal interpretation.
Training	Provide hands-on training, supervised practice, refresher sessions, job aids, peer champions, and structured onboarding for new staff. Embed training into routine supervision and ensure multiple providers are trained per facility to manage turnover and continuity.
Supplies	Ensure availability of essential consumables, functioning devices, charging infrastructure, secure storage, protective materials, maintenance schedules, and backup arrangements. Assign clear responsibility for equipment upkeep and replacement planning.
Technical support	Establish defined support channels, escalation pathways, response timelines, and maintenance logs for device, software, connectivity, or usage issues. Clearly assign responsibility across product partners and implementation teams.
Supervision	Implement routine facility supervision, district-level oversight, performance reviews, and structured escalation mechanisms. Clearly define accountability across facility, district, and system levels to ensure safe and consistent use.
Data systems	Integrate AI outputs into existing registers, EMRs, or reporting systems. Define minimum data requirements, ensure data quality checks, maintain patient linkage, and establish privacy and governance safeguards.
Workload management	Assess whether the AI tool reduces, shifts, or increases provider workload. Remove duplication in reporting, simplify workflows, clarify task allocation, and ensure integration into routine service delivery rather than parallel systems.
Incentives and motivation	Strengthen adoption through supportive supervision, continuous feedback on real usage, peer learning mechanisms, recognition of good performance, and alignment with provider roles and responsibilities to sustain engagement.
Risk management	Define protocols for device failure, incorrect outputs, data breaches, clinical disagreement, adverse events, and escalation procedures. Include clear criteria for pausing, reviewing, or redesigning the intervention when safety concerns arise.
Budgeting	Plan for recurrent costs including maintenance, consumables, connectivity, supervision, refresher training, technical support, staffing, and equipment replacement. Ensure financial planning covers full lifecycle implementation, not just initial deployment.

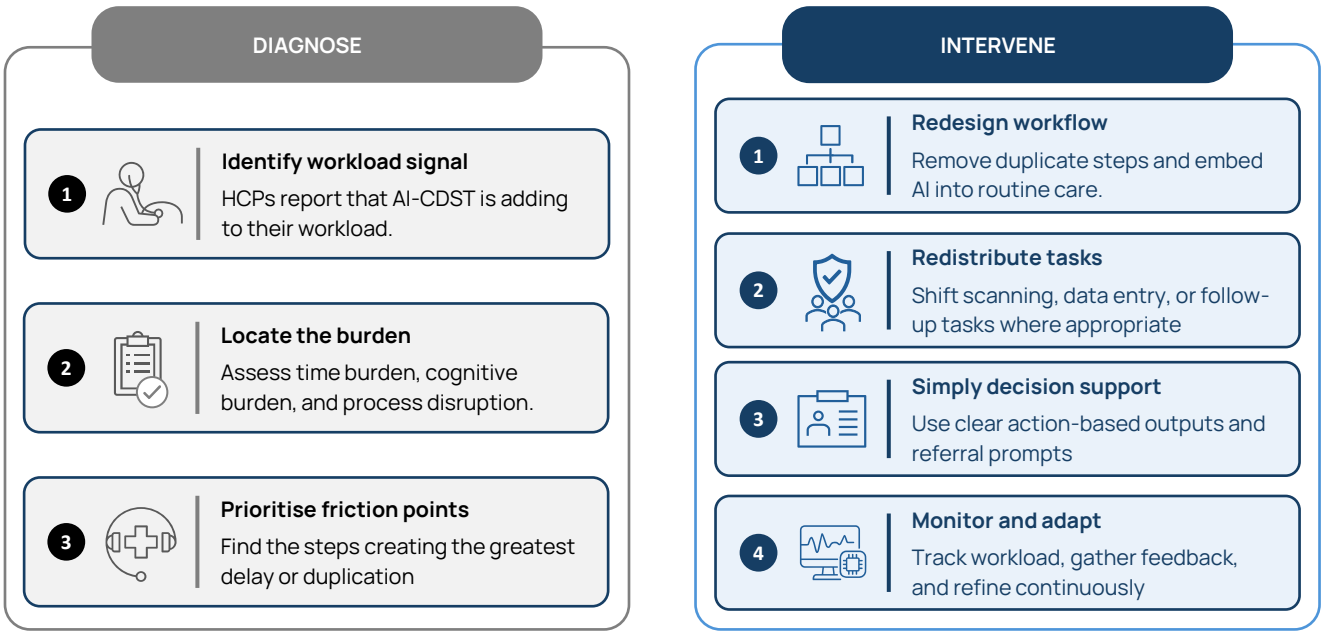
[18] Khokhar, T., & Mateen, B. A. (2026). Asking what matters: A practical guide to AI for health programme appraisal (1st ed.). Institute of Development Studies. https://doi.org/10.19088/K4DD.2026.035

Provider Workload Management

Provider workload management is the most critical operational support element in AI-CDST implementation because it directly determines whether the tool integrates into routine clinical workflows or becomes an additional burden that reduces adoption and sustainability.

Figure 8 presents a simple two-stage approach for managing provider workload during AI-CDST implementation. The first stage, Diagnose, focuses on understanding the workload problem before proposing solutions: teams first identify the signal from HCPs that the AI-CDST is adding work, then locate whether the burden is coming from extra time, cognitive effort, or disruption to existing care processes, and finally prioritize the specific friction points causing the greatest delay or duplication. The second stage, Intervene, translates this diagnosis into practical implementation fixes by redesigning the workflow, redistributing tasks across appropriate cadres, simplifying decision-support outputs into clear action prompts, and continuously monitoring workload feedback to refine the model. The core message is that AI-CDST should not be treated as an additional task for providers; it should be embedded into routine care in a way that makes the workflow more compatible, efficient, and clinically useful.

Figure 8: Provider Workload Management Strategy





A: Adapt through learning loops

The sixth stage recognizes that AI implementation is never finished at deployment. AI tools must evolve through field learning. The pilot highlighted the need for product improvements such as fetal heart rate measurement, clearer risk explanations, real-time scan guidance, scan quality scoring, local language usability, and better referral prompts. These are not simply product preferences; they affect trust, clinical usefulness, and standardization.

Learning loops should be formalized. Providers and patients generate important insights about usability, workflow fit, clinical confidence, and patient understanding. These insights should be captured systematically, reviewed regularly, and translated into product and implementation changes.

Product partners play an important role in this stage. AI health tools deployed in low-resource systems should have a documented product roadmap that responds to field needs. Implementers and governments should not treat the product as fixed. At the same time, changes must be governed carefully to preserve safety, fidelity, and accountability.

Learning loops should also include data review. Dashboards can help identify where the tool is being used, where usage is low, where referrals are not completed, where false flags are reported, and where providers need additional support. This turns implementation into a continuous improvement process.

EMBED-AI proposes four learning loops.

Loop	Key question	Data source
Clinical performance loop	Is the AI output accurate, useful, and safe in this context?	Validation data, clinical review, false positive and false negative cases
Workflow loop	Is the tool fitting into routine service delivery without excessive burden?	Observations, provider interviews, usage data, supervision reports
Patient pathway loop	Are patients able to act on AI-supported advice and complete care?	Referral tracking, follow-up calls, patient interviews, outcome records
Product modification loop	What product changes are needed to improve usability, trust, & standardisation?	Provider feedback, helpdesk logs, app analytics, field observations

Product modification categories for AI-CDSTs

AI-tool modifications should be classified before they are built. This prevents every field complaint from becoming a product feature and helps teams decide whether the right response is product redesign, training, workflow change, or governance action

Modification	Shortcoming	Action
Data capture modification	Missing, incomplete, or inaccurate data needed for clinical decision-making, follow-up, or monitoring.	Add a unique patient ID, make key fields mandatory, restrict invalid entries, capture provider name, or add visit date/time.
Data linkage modification	Weak connection between AI outputs, patient records, repeat visits, referrals, or outcomes.	Link current visit with prior results, show previous risk outputs, match follow-up records to the same patient, or connect app data with routine registers.
Clinical decision-support modification	AI output does not sufficiently guide the provider on what action should follow.	Add risk-based prompts, recommended next steps, escalation triggers, or reminders to consult a clinician.
Referral and follow-up modification	AI identifies risk but does not support referral completion or follow-up verification.	Add referral slip confirmation, referral reason fields, follow-up questions, non-completion reasons, or call-back prompts.
Explainability modification	Providers or patients do not understand what the AI output means.	Add simple explanations, local language text, traffic-light categories, counselling scripts, or uncertainty notes.
Workflow modification	The app does not fit the provider's routine or creates unnecessary burden.	Reduce duplicate fields, reorder screens, add auto-fill, simplify navigation, or place prompts at the correct point in the visit.
Quality-control modification	Poor input quality affects output credibility or clinical usefulness.	Add input-quality checks, scanning guidance, error alerts, required image quality thresholds, or training reminders.

AI-tool modification should follow a disciplined pathway:

Field insight → problem definition → classification → prioritization → product requirement → build → field testing → deployment → monitoring → next learning cycle.

I: Institutionalize for scale and sustainability

The final stage asks whether the AI-CDST can become a routine part of the health system. Institutionalisation is not the same as expansion. Expansion increases the number of sites. Institutionalisation embeds ownership, financing, governance, workforce capacity, monitoring, and accountability into the system.

The Sindh pilot demonstrates why this matters. The intervention was feasible because it was implemented with strong public-sector ownership and operational support from PPHI, linked to the Mamta programme through SSPA, supported by CERP as the research partner, and enabled by Delft Imaging as the product partner. This division of roles allowed the AI tool to be tested inside real public primary healthcare settings rather than as a parallel project.

For any scale-up, such roles must be formalised. Public institutions need clarity on who owns the intervention, who maintains devices, who supervises providers, who manages data, who governs clinical protocols, who funds recurrent costs, and who is accountable for safety and performance. Product partners need clarity on technical support, software updates, data access, product roadmap, and response timelines. Implementers need clarity on training, supervision, referral follow-up, and troubleshooting.

Institutionalisation readiness test [19]

We need to have answers to the following questions:

- ❑ Which institution owns implementation, data, protocols, and accountability for performance?
- ❑ Who pays for procurement, maintenance, consumables, training, supervision, connectivity, and replacement?
- ❑ How will training, onboarding, certification, supervision, and role clarity be maintained?
- ❑ What clinical, ethical, data, and safety governance structures oversee the tool?
- ❑ Are purchasing decisions based on clinical value, interoperability, usability, support, and total cost of ownership?
- ❑ Does the tool meet relevant regulatory, privacy, quality, and clinical safety requirements?
- ❑ Can data flow into routine systems without creating unsustainable parallel reporting?
- ❑ Will scale-up prioritise facilities and populations where the tool adds the greatest value?
- ❑ Does the tool produce sufficient system value relative to cost and alternatives?
- ❑ What happens if the tool underperforms, becomes obsolete, or the vendor exits?

Institutionalisation also requires the right scale-up metrics. AI implementation should not be scaled based only on the number of scans, predictions, log-ins, or devices deployed. The relevant indicators include reach, adoption, fidelity, referral completion, outcome verification, provider confidence, patient comprehension, equity, cost, safety, implementation burden, and sustainability.

Key Takeaways

1. First, AI must be clinically useful, not merely technically impressive. The framework starts with a clinical and system need because performance metrics alone do not guarantee health impact.
2. Second, AI must support human judgement. In low-resource health systems, AI can strengthen frontline decision-making, but it should not transfer clinical accountability to an opaque model.
3. Third, implementation must be designed around workflows. The more an AI tool requires parallel systems, duplicate reporting, or unnatural provider behaviour, the less likely it is to be sustained.
4. Fourth, referral and follow-up are part of the AI intervention. For risk-identification tools, the intervention is incomplete until the patient receives the recommended care and the outcome is known.
5. Fifth, scale must be earned through evidence. Scale-up should depend on demonstrated feasibility, acceptability, safety, adoption, equity, cost-effectiveness, and institutional readiness.

.



[19] Khokhar, T., & Mateen, B. A. (2026). Asking what matters: A practical guide to AI for health programme appraisal (1st ed.). Institute of Development Studies. https://doi.org/10.19088/K4DD.2026.035

Implications for Governments, Implementers, Product Partners, & Donors

The EMBED-AI Framework reframes AI adoption in health systems as a question of stewardship, not technology deployment. Sections 4 and 5 show that AI-based clinical decision support tools only create value when they are embedded into service delivery, provider practice, referral pathways, data systems, and institutional accountability. The implication is that different actors must move beyond narrow success metrics such as model accuracy, device distribution, or pilot completion, and instead focus on whether AI can be safely, equitably, and sustainably absorbed into routine public health systems.

6.1 Implications for governments and public health authorities

Governments should treat AI-CDSTs as public health service delivery interventions, not as standalone digital products. This means that decisions on adoption, procurement, and scale-up should be based on clinical need, equity, readiness, safety, interoperability, affordability, and long-term value.

A public-sector AI procurement process should therefore ask five questions: Does the tool address a priority service gap? Can frontline providers use it safely and confidently? Can the health system act on its outputs? Can the data be integrated into routine systems? Can the intervention be financed, governed, and maintained beyond the pilot?

Governments should also define clear institutional accountability for AI-enabled care. This includes clinical oversight, data governance, regulatory compliance, vendor accountability, safety review, and criteria for pausing or discontinuing a tool if it underperforms. Responsible AI scale-up requires public stewardship, not passive acceptance of technology claims.

6.2 Implications for implementers

Implementers should position themselves as system integrators. Their role is not only to deploy devices, train providers or monitor usage. It is to connect the AI tool with the real operating environment of care: facilities, providers, patients, supervisors, referral centres, and data systems.

This requires implementation teams to diagnose bottlenecks continuously and distinguish between different kinds of failure. Low adoption may reflect weak training, poor workflow fit, provider mistrust, patient-side barriers, limited referral capacity, or product limitations. Without this diagnostic discipline, a potentially useful tool may be wrongly judged ineffective, or a weak implementation model may be scaled prematurely.

Implementers should therefore maintain adaptive learning systems that combine routine monitoring, provider feedback, patient follow-up, referral verification, and structured issue resolution. In AI-CDST implementation, learning is not an evaluation activity at the end of a project; it is part of implementation itself.

6.3 Implications for product partners

Product partners should design AI-CDSTs for real-world public health delivery rather than ideal clinical environments. In low-resource settings, the most important product features may not be the most technically sophisticated ones. Usability, explainability, local language support, low-connectivity functionality, interoperability, durability, and responsive technical support are central to adoption and scale.

AI outputs must be clinically meaningful and actionable. A risk flag, score, or recommendation is insufficient if providers cannot understand why it appears, what it means, and what action should follow. Product design should therefore support clinical interpretation, not simply prediction.

Product partners should also accept that deployment creates new evidence. Field use will reveal missing features, workflow constraints, trust issues, and data gaps. A credible product partner should maintain a responsive product roadmap, document changes, communicate updates clearly, and remain accountable for implementation support after initial deployment.

6.4 Implications for donors and funders

Donors should fund AI health interventions as implementation learning platforms, not only as innovation pilots. Short pilots that focus on product use or technical feasibility are unlikely to generate the evidence needed for responsible scale-up. Funding should support the full pathway from need definition to workflow integration, provider adoption, referral strengthening, data governance, value-for-money analysis, and sustainability planning.

Donors should also require stronger evidence before supporting scale. This evidence should include implementation outcomes such as acceptability, adoption, appropriateness, feasibility, fidelity, cost, penetration, and sustainability, in addition to clinical or service delivery outcomes (Proctor et al., 2011). For AI-CDSTs, scale should be earned through demonstrated system readiness and not assumed because the technology is promising.

Finally, donors can shape the AI-for-health market by incentivising tools that are interoperable, explainable, equity-oriented, implementation-ready, and aligned with public-sector priorities. This is especially important in LMIC settings where governments may face pressure to adopt AI products before sufficient implementation evidence exists.

Annex 1: Implementation Science Frameworks and Their Relevance

Framework	Focus	Relevance to EMBED-AI
Consolidated Framework for Implementation Research (CFIR) developed by **Damschroder et al. in 2009**[20] and updated in **2022** [21]	CFIR is a determinant framework that identifies the contextual factors influencing whether an intervention is successfully implemented. It examines implementation across five domains: innovation, the outer setting, the inner setting, individuals, and the implementation process. It is particularly useful for understanding how organisational readiness, leadership, workflow compatibility, provider capability, implementation climate, policy context, infrastructure, and system-level barriers shape implementation outcomes.	CFIR is highly relevant to EMBED-AI because AI-CDST implementation depends on whether the tool fits the clinical, organisational, workforce, infrastructure, and policy context. It supports EMBED-AI's emphasis on assessing clinical fit, facility readiness, provider capability, leadership support, workflow alignment, referral capacity, and system barriers before deployment. CFIR is strongest for identifying what may influence implementation; it is less prescriptive about the sequence of strategies and operational actions required to address those determinants. (Edwards et al., 2026).[22]
RE-AIM Framework developed by **Glasgow, Vogt, and Boles in 1999**[23]	RE-AIM is an evaluation and planning framework that assesses the public health impact of interventions through five dimensions: Reach, Effectiveness, Adoption, Implementation, and Maintenance.[24] It moves evaluation beyond whether an intervention works under controlled conditions and asks whether it reaches the intended population, produces meaningful effects, is adopted by providers, is implemented with quality, and is sustained over time.	RE-AIM is relevant to EMBED-AI because AI-CDSTs should not be assessed only through technical performance or model accuracy. EMBED-AI uses this logic to emphasise reach among underserved populations, provider and facility adoption, equity, implementation quality, referral completion, and sustainability. Re-AIM is stronger for evaluation than for implementation design. It tells implementers what to measure, but not how to build trust, strengthen referral system & redesign workflow.
Exploration, Preparation, Implementation, Sustainment (EPIS) Framework[25] developed by **Aarons, Hurlburt, and Horwitz in 2011**[26]	EPIS is a process framework that conceptualises implementation as a staged pathway. The Exploration phase focuses on identifying the problem and deciding whether an intervention is needed. The Preparation phase involves planning, readiness assessment, adaptation, and stakeholder engagement. Implementation phase involves active delivery and support. The Sustainment phase focuses on embedding the intervention into routine practice over time.	EPIS is relevant to EMBED-AI because AI-CDSTs should not move directly from technical validation to deployment. EMBED-AI reflects this phased logic by moving from clinical need identification and readiness assessment to workflow embedding, provider support, learning loops, and institutionalisation. Its limitation is that it does not specify AI-specific activities such as explainability, false alert management, human-in-the-loop governance, data interoperability, product updates, or closed-loop referral systems.
Normalization Process Theory (NPT) developed by **May, Finch, and colleagues, with major publications in 2009** [27]	NPT explains how new practices become embedded and normalised within everyday work. It focuses on four constructs: coherence, cognitive participation, collective action, and reflexive monitoring.[28] These constructs examine whether users understand the intervention, commit to it, carry out the work required to use it, and assess whether it remains valuable in practice.	NPT is relevant to EMBED-AI because AI-CDSTs only create value when providers integrate them into routine clinical workflows rather than using them as parallel project tools. EMBED-AI draws on this logic through its emphasis on provider trust, sense-making, workflow embedding, counselling routines, supervision, and continued use. It focuses mainly on social embedding and routine practice, while offering less guidance on referral systems, product governance, data integration, financing, and scale-up.

[20] Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. Implementation Science, 4, 50. https://doi.org/10.1186/1748-5908-4-50
[21] Damschroder, L. J., Reardon, C. M., Widerquist, M. A. O., & Lowery, J. (2022). The updated Consolidated Framework for Implementation Research based on user feedback. Implementation Science, 17, 75. https://doi.org/10.1186/s13012-022-01245-0
[22] Edwards, J., Armstrong, M., Begum, S., Dunphy, E., Lau, R., Naughton, M., Taylor, S. J., & Ross, J. (2026). Implementation of digital health interventions in primary care in high-income countries: A systematic review. The Lancet Primary Care, Article 100175. https://doi.org/10.1016/j.lanprc.2026.100175
[23] Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. American Journal of Public Health, 89(9), 1322–1327. https://doi.org/10.2105/AJPH.89.9.1322
[24] RE-AIM. (n.d.). Welcome to RE-AIM. https://re-aim.org/
[25] EPIS Framework. (n.d.). EPIS framework publications. https://episframework.com/publications/
[26] Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. Administration and Policy in Mental Health and Mental Health Services Research, 38(1), 4–23. https://doi.org/10.1007/s10488-010-0327-7

Framework	Focus	Relevance to EMBED-AI
Behaviour Change Wheel and COM-B Model developed by **Michie, van Stralen, and West in 2011** [29]	The Behaviour Change Wheel is a behaviour change design framework built around the COM-B model, which argues that behaviour is shaped by Capability, Opportunity, and Motivation. Capability includes knowledge and skills. Opportunity includes the physical and social conditions that enable behaviour. Motivation includes reflective and automatic processes that drive action. The framework links these behavioural determinants to intervention functions such as training, education, enablement, environmental restructuring, persuasion, modelling, and incentivisation.	This framework is relevant to EMBED-AI because provider adoption is a behavioural challenge, not only a technical one. Providers must learn how to use the tool, interpret outputs, counsel patients, document decisions, refer appropriately, and continue using the tool under routine conditions. EMBED-AI reflects COM-B through its focus on practical training, supervision, workflow opportunity, motivation, trust, incentives, and support systems. COM-B and the Behaviour Change Wheel are deliberately behaviour-focused; they do not by themselves provide the wider organisational, technological, and service-delivery architecture required for implementation..
NASSS Framework developed by **Greenhalgh, Wherton, Papoutsi, and colleagues in 2017** [30]	NASSS stands for Non-adoption, Abandonment, Scale-up, Spread, and Sustainability. It was developed to explain why health and care technologies are not adopted, are abandoned, or fail to scale and sustain. It examines complexity across the condition, the technology, the value proposition, adopters, organisations, the wider system, and adaptation over time. Its central insight is that technology implementation often fails when complexity is underestimated.	NASSS is one of the most relevant frameworks for EMBED-AI because AI-CDSTs are complex technologies introduced into complex clinical and organisational systems. EMBED-AI reflects NASSS by treating AI implementation as more than product deployment and by addressing technology usability, value proposition, provider adoption, organisational readiness, policy alignment, interoperability, and scale-up risk. Its limitation is that it is primarily diagnostic; it explains why technologies may fail but does not provide a detailed operational playbook for implementing AI-CDSTs in public health system.
Practical, Robust Implementation and Sustainability Model (PRISM) developed by **Feldstein and Glasgow in 2008** [31]	PRISM is a context-oriented implementation and sustainability model that examines how interventions interact with recipients, organisations, implementation infrastructure, sustainability infrastructure, and the external environment. It is closely linked to RE-AIM because it helps explain the contextual factors that influence reach, adoption, implementation, and maintenance.[32] PRISM pays particular attention to intervention-context fit and to how different recipients perceive the intervention.	PRISM is relevant to EMBED-AI because AI-CDSTs must fit the realities of patients, providers, facilities, referral systems, data systems, and public health institutions. EMBED-AI reflects PRISM through its emphasis on recipient perspectives, workflow fit, organisational readiness, operational infrastructure, sustainability planning, and external policy alignment. PRISM is particularly strong on context, recipient perspectives, implementation infrastructure, and sustainability, but is less explicit about the sequence of operational activities required to move an intervention from readiness to routine delivery.
Dynamic Sustainability Framework (DSF) developed by **Chambers, Glasgow, and Stange in 2013** [33]	DSF argues that sustainability should not be understood as preserving an intervention exactly as originally designed. Instead, sustainability requires ongoing learning, adaptation, and fit between the intervention, the practice setting, and the wider ecological system. It is particularly useful for interventions operating in changing systems where rigid replication may weaken long-term effectiveness.	DSF is relevant to EMBED-AI because AI-CDSTs are dynamic interventions. Their interfaces, clinical prompts, product features, data fields, workflows, and governance requirements may need to evolve as they are used in real settings. EMBED-AI reflects DSF through its emphasis on learning loops, product refinement, workflow adaptation, dashboard review, field feedback, and governed evolution from pilot to scale. DSF is strongest for understanding adaptation and sustainment over time rather than for designing the complete implementation pathway from initial need assessment through early deployment..

[27] May, C., & Finch, T. (2009). Implementing, embedding, and integrating practices: An outline of Normalization Process Theory. *Sociology, 43*(3), 535–554. https://doi.org/10.1177/0038038509103208
[28] May, C. R., Mair, F., Finch, T., MacFarlane, A., Dowrick, C., Treweek, S., Rapley, T., Ballini, L., Ong, B. N., Rogers, A., Murray, E., Elwyn, G., Légaré, F., Gunn, J., & Montori, V. M. (2009). Development of a theory of implementation and integration: Normalization Process Theory. *Implementation Science, 4*, 29. https://doi.org/10.1186/1748-5908-4-29
[29] Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science, 6*, 42. https://doi.org/10.1186/1748-5908-6-42
[30] Greenhalgh, T., Wherton, J., Papoutsi, C., Lynch, J., Hughes, G., A'Court, C., Hinder, S., Fahy, N., Procter, R., & Shaw, S. (2017). Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *Journal of Medical Internet Research, 19*(11), e367. https://doi.org/10.2196/jmir.8775
[31] Feldstein, A. C., & Glasgow, R. E. (2008). A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. *Joint Commission Journal on Quality and Patient Safety, 34*(4), 228–243. https://doi.org/10.1016/S1553-7250(08)34030-6
[32] RE-AIM. (n.d.). *What is PRISM?*https://re-aim.org/learn/prism/
[33] Chambers, D. A., Glasgow, R. E., & Stange, K. C. (2013). The dynamic sustainability framework: Addressing the paradox of sustainment amid ongoing change. *Implementation Science, 8*, 117. https://doi.org/10.1186/1748-5908-8-117

Framework	Focus	Relevance to EMBED-AI
Proctor's Implementation Outcomes Framework developed by **Proctor et al. in 2011** [34]	Proctor's framework distinguishes implementation outcomes from clinical and service outcomes. It identifies eight implementation outcomes: acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability.[35] This distinction is important because an intervention can be clinically promising but poorly implemented, or well implemented but unable to produce desired clinical effects.	Proctor's framework is relevant to EMBED-AI because AI-CDST pilots should not be evaluated only through accuracy or clinical outcomes. EMBED-AI can use Proctor's outcomes to assess whether the tool is acceptable to providers and patients, adopted by facilities, appropriate for the setting, feasible under routine conditions, used with fidelity, affordable to implement, integrated into routine care, and sustained over time. Proctor's framework defines what successful implementation should be measured against, but does not prescribe the implementation strategies required to achieve those outcomes.
Integrated Promoting Action on Research Implementation in Health Services (i-PARIHS) developed by **Harvey and Kitson in 2016,**[36] building on the earlier **PARIHs frameworks (1990s / 2000s)** [37 38]	i-PARIHS explains successful implementation as the result of facilitation of an innovation with recipients in context. Its core constructs are innovation, recipients, context, and facilitation. The framework's distinctive contribution is its emphasis on facilitation as the active ingredient that helps people adopt, adapt, and sustain an intervention in real settings.	i-PARIHS is relevant to EMBED-AI because AI-CDST implementation requires more than device deployment or one-time training. Providers may need mentoring, reassurance, troubleshooting, supervision, local adaptation, and repeated engagement to build confidence. EMBED-AI reflects this through its emphasis on facilitation, provider handholding, supportive supervision, stakeholder engagement, and implementation support systems. i-PARIHS provides depth on facilitation, recipients, innovation, and context, but does not address complete architecture for technology-enabled service delivery.
What Works Hub Implementation Science Framework developed by the **What Works Hub for Global Education, University of Oxford** and partners, with recent public-facing work in **2025–2026** [39]	The What Works Hub framework focuses on how evidence-based interventions move from controlled studies to sustained effectiveness at scale with equity. It links efficacy trials, effectiveness trials, policy plans, and practice at scale,[40] and emphasises that evidence, policy, and implementation practice interact continuously rather than following a simple linear pathway. Although developed for education, it offers a useful public-sector implementation lens.	The framework is relevant to EMBED-AI because AI-CDSTs must move beyond validation and pilot evidence into policy alignment, public-sector delivery, routine practice, and equity-focused scale-up. EMBED-AI reflects this by emphasising government ownership, delivery system integration, value for money, institutionalisation, and equitable reach. Its principal strength is understanding how evidence moves through policy and public-sector delivery to scale; because it was developed in education, its application to clinical service delivery requires adaptation to health-system and patient-safety considerations.
Context Compass Framework developed by **The Center for Implementation,** with public-facing tools released **in 2026** [41]	Context Compass is a practical context assessment framework that helps implementation teams examine contextual factors across setting, organisation, inter-organisational relationships, and wider system conditions.[42] It links context to fit, readiness, implementation, and sustainability. It is designed as a practical tool to help teams understand what in the environment may support or constrain implementation.	Context Compass is relevant to EMBED-AI because AI-CDST implementation in low-resource settings depends heavily on facility, district, institutional, and policy context. It supports readiness assessment by helping teams examine staffing, infrastructure, leadership, referral linkages, relationships, resources, policy alignment, and sustainability conditions. Its limitation is that it is a general context assessment tool, not a clinical AI or health technology framework, and it does not directly address algorithmic safety, explainability, data interoperability, or product updates.

[34] Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7

[35] Implementation Outcome Repository. (n.d.). *Implementation outcomes*. https://implementationoutcomerepository.org/implementation-outcomes/

[36] Harvey, G., & Kitson, A. (2016). PARIHS revisited: From heuristic to integrated framework for the successful implementation of knowledge into practice. *Implementation Science, 11*, 33. https://doi.org/10.1186/s13012-016-0398-2

[37] Kitson, A., Harvey, G., & McCormack, B. (1998). Enabling the implementation of evidence based practice: A conceptual framework. *Quality in Health Care, 7*(3), 149–158. https://doi.org/10.1136/qshc.7.3.149

[38] Rycroft-Malone, J. (2004). The PARIHS framework: A framework for guiding the implementation of evidence-based practice. *Journal of Nursing Care Quality, 19*(4), 297–304. https://doi.org/10.1097/00001786-200410000-00002

[39] Kilburn, L. (2026). *What is implementation science in education?* What Works Hub for Global Education. https://doi.org/10.35489/BSG-WhatWorksHubforGlobalEducation-BL_2026/010

[40] What Works Hub for Global Education. (n.d.). *Our approach and framework*. https://www.wwhge.org/what-we-do/implementation-science/our-approach-framework/

[41] The Center for Implementation. (2026). *The Context Compass Framework explained*. https://thecenterforimplementation.com/toolbox/context-compass-framework-explained

[42] The Center for Implementation. (2026). *Context Compass rapid reflection*. https://thecenterforimplementation.com/toolbox/context-compass-rapid-reflection

Annex 2: Learnings From The AI Ultrasound Pilot in Sindh

The AI-enabled ultrasound pilot in Sindh, Pakistan, was a Gates Foundation-funded implementation and learning initiative that assessed the feasibility, usability, and integration of AI-enabled handheld obstetric ultrasound within public primary healthcare settings. It was implemented through a collaboration among People's Primary Health Initiative Sindh (PPHI) and Sindh Social Protection Authority (SSPA) as implementation partners, the Centre for Economic Research (CERP) as the research partner, and Delft Imaging as the product partner. The desired outcomes of the pilot were linked to the Mamta programme being implemented by SSPA. Mamta is a maternal and child health social protection initiative that provides conditional cash transfers to pregnant women and promotes antenatal care, institutional delivery, postnatal care, and child immunisation. The intervention introduced BabyChecker, a portable AI-enabled ultrasound device designed for frontline healthcare providers with limited to no ultrasound training. The device generates automated obstetric assessments and presents findings through a traffic-light classification system to identify risk and support referral decisions.

The pilot aimed to improve uptake of antenatal care (ANC) services and referral pathways for high-risk pregnancies among Mamta beneficiaries, while generating implementation evidence on the integration of AI-enabled ultrasound into routine maternal healthcare services. To achieve these aims, the intervention combined AI-assisted ultrasound screening with provider training, workflow integration, referral and follow-up mechanisms, digital data capture, and linkage with maternal health records.

The AI Ultrasound pilot was designed as a two-arm stepped-wedge randomised controlled trial to evaluate both the impact and implementation feasibility of the BabyChecker AI-enabled ultrasound device within routine public-sector antenatal care services in Sindh. The study was implemented across 20 public primary health facilities in six districts, with all facilities beginning in the control condition at baseline. Arm 1 facilities received the device from 15 October 2025 and remained under intervention through midline and endline, while Arm 2 facilities remained in control at midline and received the device from 15 January 2026. This staggered rollout enabled comparisons across the control and treatment periods while accounting for different exposure durations: approximately 6 months for Arm 1 and 4 months for Arm 2 by endline. The phased approach also supported operational learning and adaptation during rollout.

The pilot was not intended solely to test the technical performance of BabyChecker. It sought to understand how AI-supported ultrasound could be incorporated into routine antenatal care in resource-constrained settings and what organisational, behavioural, and system-level conditions were necessary for successful adoption. The pilot generated evidence on provider adoption, patient acceptability, referral practices, workflow integration, and operational feasibility within the public health system.

The pilot is particularly valuable for implementation science because it demonstrates that the success of AI in healthcare depends not only on technical performance but also on how well the tool fits in clinical workflows, supports frontline providers, enables effective counselling and referral, and connects with broader health-system processes. In this sense, it reinforces a central insight from implementation science: the effectiveness of an innovation is inseparable from the context and system into which it is introduced (Damschroder et al., 2022; Greenhalgh et al., 2017).[10] [20]

The following section synthesises the key lessons and implementation insights derived from this experience.

1. AI ultrasound added the most value in facilities where routine ultrasound services or expertise were unavailable

The pilot showed that AI-enabled ultrasound delivered the greatest value in facilities where conventional ultrasound was unavailable. In these settings, the device brought basic obstetric screening closer to pregnant women who would otherwise have had to travel to higher-level facilities or private providers. Implementation data supported this pattern: Mamta enrolment increased by 7.5%, from 2,677 to 2,877, after BabyChecker implementation, and interrupted time series analysis showed a statistically significant increase in enrolment trends only in facilities without conventional ultrasound. This is an important implementation lesson: AI tools should not be deployed simply because they are innovative; they should be deployed where they address a clearly defined service delivery gap.

The value proposition was strongest where trained sonographers were unavailable, diagnostic infrastructure was limited, and pregnancy risk was often identified late. In such contexts, AI-enabled ultrasound helped frontline providers estimate gestational age, identify fetal presentation, assess placental location, identify multiple pregnancies, classify pregnancy risk, and support counselling and referral decisions. Future scale-up should therefore prioritise remote and underserved facilities where AI expands access rather than replicates what the system already provides.

2. Clinical interpretation remained essential to every AI-supported decision

The pilot showed that AI outputs were most useful when healthcare providers interpreted them alongside their own clinical assessment. Providers used BabyChecker findings to counsel women, decide whether referral was needed, and plan follow-up care. This helped move practice from uncertainty-based referral to more evidence-informed triage.

This distinction matters for AI implementation. In low-resource settings, AI can strengthen frontline decision-making, but it should not replace clinical responsibility. During implementation, 340 women, representing 20.3% of those scanned, received yellow or red AI scan results. However, HCPs referred 153 of these women, or 45.0%, after applying clinical assessment. This shows that providers did not treat AI flags as automatic referral instructions; they used them as decision-support inputs. Yellow and red risk flags require interpretation, counselling, and, where necessary, escalation. Green results also require caution because a normal AI output should not override clinical symptoms or provider concern. Scale-up protocols should therefore define when AI outputs require referral, when repeat scanning or follow-up is sufficient, and when a doctor or higher-level clinician should review the case.

The pilot's referral protocol reflected this principle. PPHI referral slips were issued for red or yellow scans where the HCP judged referral to be clinically appropriate. The final referral decision remained with the HCP, and app prompts were introduced to record whether a referral slip was issued and, where it was not issued, why the HCP made that decision. Validation data further reinforced the need for HCP judgement: the referral validation dataset identified 6 correct referrals, 2 false positives, and 1 false negative. This preserved clinical judgement while making decision-making more explicit and auditable.

3. Local validation study, transparent communication, and hands-on support increased provider trust by reducing uncertainty

Trust emerged as a core implementation condition. HCPs were more likely to use BabyChecker confidently when they understood the tool, saw its clinical value, and had evidence that its outputs broadly aligned with standard ultrasound. The pilot, therefore, treated trust-building as an implementation activity rather than a soft or secondary consideration.

Several mechanisms helped build trust. First, providers received hands-on training, handholding, refresher support, and access to a WhatsApp support channel. Second, the team used transparent communication to explain what the device could and could not do. Third, the project conducted a local validation study comparing BabyChecker outputs with standard ultrasound results for gestational age, fetal presentation, and placenta location across six primary health facilities. The validation study included 200 paired scans; 164 scans agreed with conventional ultrasound, while 31 were false positives and 5 were false negatives. This local validation was important because providers are more likely to trust AI outputs when they see evidence from the same type of setting in which they work.

The pilot also showed that trust can weaken quickly when HCPs experience false flags or unexplained risk classifications. Doctors and midwives reported that inaccurate high-risk flags could create clinical anxiety, increase unnecessary referrals, and affect the credibility of the facility. This learning has wider relevance for AI-CDSTs: trust depends not only on model accuracy, but also on explainability, escalation rules, provider feedback, and visible mechanisms for reviewing errors. Future implementation should therefore combine local validation, transparent risk explanations, clear escalation pathways, and human-in-the-loop governance before scale-up.

4. Successful frontline adoption required hands-on training, supervision, and workflow integration

The pilot demonstrated that non-specialist providers, including midwives, nurses, community midwives, and lady health workers, can operate AI-enabled ultrasound devices in routine primary care settings. This finding is important for low-resource health systems because it suggests that AI can support task-sharing where specialist staff are unavailable.

However, the pilot also showed that adoption does not happen automatically. Frontline providers needed practical training, supervised use, refresher sessions, troubleshooting support, and ongoing encouragement. Across implementation, providers completed 2,698 scans; 1,674 scans, or 62%, were quality sufficient, while 1,024 scans, or 38%, required quality improvement. Most insufficient scans resulted from sweep technique issues, which accounted for 59.6% of poor-quality scans. Simply installing the device was not enough; it became useful because providers received support to integrate it into their daily work.

Qualitative insights suggest that the device also strengthened provider confidence and professional authority. Some midwives reported that the scan helped validate their physical assessment and enabled more confident counselling. Others felt that women trusted them more because they could now provide ultrasound-backed information at the facility. This matters because provider adoption improves when technology strengthens the provider's role rather than making the provider feel bypassed or replaced.

This learning aligns with behaviour change and normalization theories. Providers need capability, opportunity, and motivation to use an AI tool consistently (Michie, van Stralen and West, 2011).[28] They also need to understand the tool, value it, and incorporate it into routine work (May and Finch, 2009).[26] Future AI-CDST implementation should therefore treat frontline providers as central implementation actors, not passive end-users.

5. Pregnant women were more likely to take action when providers explained the scan results clearly

The pilot showed that pregnant women valued ultrasound when providers converted scan findings into understandable and actionable information. Information on gestational age, fetal presentation, fetal condition, and referral need helped women and families make more informed decisions. This was especially important in settings where women had limited access to ultrasound and where pregnancy risk was often identified late.

The device also gave providers a concrete counselling tool. Instead of giving general advice, HCPs could explain the scan result, the colour category, the reason for referral or follow-up, and the need for facility delivery where relevant. This supported birth preparedness and care-seeking because women received information that felt visible, specific, and credible. Implementation data showed that the average enrolment week decreased from 14.37 to 12.02 weeks, institutional deliveries reached 84.1% among women referred from facilities without previous ultrasound services, and referral completion reached 90.6% among tracked women.

However, ultrasound access alone did not fully transform broader ANC behaviour. Continued care-seeking still depended on counselling quality, household decision-making, transport costs, perceived need, and follow-up support. AI can create a strong moment of engagement, but women still need support to act on the information. Future implementation should therefore connect AI-generated risk information with patient-side support, including family counselling, follow-up calls, and social protection linkages through platforms such as Mamta.

6. AI created a meaningful impact when the intervention was embedded into clinical workflows

One of the strongest implementation lessons from the pilot was that AI created value when teams embedded the scan into the ANC workflow. The operational pathway included EMR or client ID entry, scan completion, AI-generated result, HCP interpretation, counselling, referral slip issuance, maternal health register recording, follow-up prompts, and referral tracking. This converted the device from a standalone technology into part of a maternal care pathway.

This finding is central to the EMBED-AI Framework. AI interventions should not be treated as technical deployments. They should be treated as workflow redesign interventions. The key question is not only whether a provider can use the device, but whether the device fits the way providers already deliver care. If the tool creates parallel documentation, increases administrative burden, or remains disconnected from routine records, use may decline over time.

PPHI's implementation role was central to this workflow integration. PPHI's facility network, district coordination, provider engagement, and operational support allowed the device to enter existing public primary healthcare structures rather than operate as a parallel project. This made the pilot a health-system implementation effort, not only a technology trial.

7. Simple referral tools strengthened the pathway from risk identification to care completion

The pilot showed that identifying a high-risk pregnancy is only the first step. AI creates value when the woman completes the referral, receives care at the referral facility, and has her outcome documented. Referral tracking showed 90.6% referral completion among tracked women. Without this continuity, AI can identify risk without ensuring that the system responds.

The implementation team used several practical mechanisms to strengthen referral completion. HCPs issued referral slips for clinically appropriate red and yellow cases. The pilot also introduced referral stickers with simple instructions in the local language and a WhatsApp contact number. These stickers helped women and families understand that an AI-enabled ultrasound had identified a condition requiring further guidance. They also gave them a practical channel to contact the project team if they needed support in completing the referral process. This type of low-cost implementation tool is important because referral completion often depends on whether the patient clearly understands the next step.

The project further strengthened follow-up through phone calls, app-based follow-up questions, and repeat visit tracking. When women returned to the facility, the app used their EMR number to identify whether they had previously received a yellow or red scan. The HCP could then ask whether they completed the referral, what care they received, and why they did not complete referral if they failed to do so. Referred women also showed better on-time ANC completion than non-referred women, with ANC2 completion at 100% versus 91%, ANC3 completion at 83% versus 79%, and ANC4 completion at 86% versus 83%.

Future implementation should build on this approach by formalising linkages with secondary and tertiary facilities. Referral slips, referral feedback, delivery location tracking, complication reporting, and outcome verification should become standard parts of the AI-CDST pathway. For AI-supported care, a recommendation should not be considered complete until the downstream action and outcome are known.

8. Product adaptation strengthened implementation by turning feedback from the field into app changes

The pilot generated important product-level learning. Providers found BabyChecker promising, but they also identified features that would make the tool more clinically useful, easier to trust, and easier to integrate into routine care. Scan-quality data reinforced these product needs: 59.6% of poor-quality scans were caused by incorrect sweep technique and 25.5% resulted from scanning before the recommended gestational age. Providers also reported reporting delays and connectivity issues. These findings point to the need for fetal heart rate measurement, real-time scan guidance, scan quality feedback, clearer explanations of risk flags, stronger referral prompts, and local language usability.

The implementation team converted some of this field learning into app modification requests. The app was updated to include an 11-digit EMR number after the beneficiary's name so that returning women could be identified more reliably. For red or yellow scans, pop-up questions were added to remind the HCP to issue the referral slip, add the EMR number, explain follow-up steps, and record whether the referral process had been followed. The app also required the HCP to record why a referral slip was not issued when applicable, such as referral slip unavailability, time constraints, beneficiary refusal, or HCP judgment.

A follow-up referral compliance module was also introduced for returning patients. When the app identified a woman with a previous red or yellow scan, it prompted the HCP to ask whether she completed referral, what care she received, and why she did not complete referral if she failed to do so. This moved the app beyond scan generation and turned it into a workflow support tool for referral tracking.

The pilot also highlighted the importance of incorporating a human-in-the-loop concept, particularly in distinguishing between quick scan results and verified results. Quick scan results supported immediate counselling and decision-making, while verified results provided additional accuracy and information after upload and review, incorporating a human-in-the-loop layer where trained experts reviewed and validated AI outputs before final confirmation. Providers could also show ultrasound photos or videos to women, which helped build confidence and satisfaction. However, delays in verified results and connectivity issues also created workload and follow-up challenges. Future AI-CDST products should therefore be designed as evolving implementation tools, not fixed digital products.

9. Operational readiness influenced uptake as much as technical performance

The pilot reinforced that routine operational details can determine whether AI tools remain usable. Device charging, internet connectivity, ultrasound gel availability, safe storage, device handling, troubleshooting, spare devices, and theft-risk management directly influenced implementation quality. These issues may appear minor in technical discussions, but in public-sector delivery they often determine whether a tool is used consistently.

Connectivity was especially important. Providers reported that delayed uploads created administrative backlogs, delayed verified results, and increased stress. Delays also made follow-up harder because women might not return to collect results after one or two days. This shows that AI implementation models must plan for weak connectivity, offline functionality, upload queues, and realistic reporting timelines.

Future AI-CDSTs should therefore be accompanied by an operational support package. This package should include SOPs, training materials, device maintenance plans, supply chain arrangements, troubleshooting channels, data-sharing agreements, dashboards, incentives, escalation protocols, and clear accountability mechanisms. A technically strong tool can still become underused if the operational system around it is weak.

10. Provider continuity required onboarding, refresher training, and workload management

Provider turnover and workload emerged as distinct implementation risks. Training cannot be treated as a one-time activity because staff may rotate, leave, or shift responsibilities. New providers must receive onboarding, existing providers need refresher training, and district teams need clear responsibility for maintaining capability over time.

Workload also needs careful management. The device simplified some clinical tasks by making screening easier and more localized. Average time spent obtaining ultrasound reduced from 34.5 minutes to 25.9 minutes, a 24.9% reduction, and field observations found that women typically completed the visit within 20–25 minutes. At the same time, the tool introduced additional documentation, reporting, upload monitoring, referral recording, and follow-up tasks. HCPs may perceive the tool as burdensome if these activities are not integrated into routine systems or if they duplicate existing records.

Future implementation should therefore include a standing provider support system. This should include onboarding materials, refresher sessions, supervision visits, peer champions, a WhatsApp support channel, technical helpdesk support, simplified documentation, and periodic workload review. Support should focus on problem-solving, not only compliance.

11. Data integration and advanced analytics created visibility across the full care pathway from scan to outcome

The pilot showed that AI-enabled ultrasound can generate valuable data for implementation oversight, clinical follow-up, and system learning. However, the full value of the data emerged only when the team triangulated multiple data sources. These included BabyChecker scan data, SSPA administrative data, facility records, baseline/midline/endline surveys, exit surveys, follow-up calls, field observations, qualitative interviews, and validation study data.

The additional analysis strengthened implementation learning in several ways. Monthly Mamta enrollment trends helped distinguish intervention-related changes from seasonal patterns such as winter and Ramadan effects. Interrupted time series analysis helped examine whether observed trends reflected a meaningful shift rather than simple pre-post variation. Linking BabyChecker scan data with SSPA administrative data enabled deeper analysis of referral completion, subsequent ANC visits, delivery location, and outcomes among referred and non-referred high-risk women. Validation data also helped examine false positives and false negatives by comparing BabyChecker outputs with standard ultrasound.

This data work is central for AI-CDST implementation. A dashboard should not only count the number of scans. It should show provider usage, risk categories, referral decisions, reasons for non-referral, referral completion, care received, delivery location, complications, institutional delivery, false flag concerns, upload delays, and device downtime. Monitoring should capture the full pathway from AI output to patient outcome. This allows implementers to distinguish between product performance issues, provider adoption issues, referral system gaps, and patient-side barriers.

12. Strong public-sector ownership and institutional support from PPHI made implementation feasible

The pilot would not have been possible without PPHI's implementation support. PPHI provided the operational infrastructure through which BabyChecker could be tested inside real public primary healthcare settings. Its role in facility access, provider coordination, district-level implementation, operational readiness, and issue resolution was central to making the pilot workable.

This learning has wider relevance for AI scale-up. AI tools cannot be sustainably introduced into public health systems through product partners alone. They require institutional owners, trusted implementation partners, public-sector legitimacy, and routine accountability structures. PPHI's support helped integrate the intervention into existing facilities rather than creating a parallel technology project.

For future scale-up, this support should be formalised through clear roles, governance arrangements, data-sharing protocols, supervision systems, referral accountability, device maintenance responsibilities, and financing plans. Institutionalisation is not only the final stage of implementation; it must be designed from the beginning.

The pilot shows that AI-CDST implementation is a system redesign challenge

Taken together, the pilot generated several cross-cutting lessons for AI-CDST implementation in public health systems.

First, AI creates the most value when it closes a real and underserved clinical access gap. Second, AI-based outputs work best when HCPs use them alongside clinical judgement. Third, provider trust requires validation, transparency, escalation pathways, and ongoing support. Fourth, frontline adoption depends on training, workflow fit, manageable workload, and routine supervision. Fifth, women act on AI-supported information when counselling makes the result clear and actionable. Sixth, referral completion improves when teams translate risk flags into simple tools, follow-up systems, and documented pathways. Seventh, product design must evolve through field learning. Eighth, operational readiness and data integration are as important as technical performance. Finally, public-sector implementation depends on institutional ownership and strong delivery partnerships.

These lessons directly inform the EMBED-AI Framework presented in the next section. The pilot shows that the central implementation challenge is not how to deploy an AI product, but how to embed an AI-supported decision process into routine care. In low-resource health systems, the pathway from AI output to health impact runs through people, workflows, referrals, data systems, product adaptation, operational readiness, and institutions.



Annex 3: Implementation Readiness Matrix

The EMBED-AI Framework is intended to be both conceptual and practical. While the seven stages describe the implementation pathway, decision-makers also need a simple way to assess whether an AI-based clinical decision support tool is ready to move from design to pilot, from pilot to scale, or from scale to long-term institutionalisation. The following Implementation Readiness Matrix translates the framework into a practical assessment tool.

The matrix should not be used as a rigid scoring checklist. Rather, it should be used as a structured reflection tool by governments, implementation partners, product developers, funders, and evaluators. Its purpose is to identify whether the essential conditions for responsible AI implementation are present, partially present, or missing. Where readiness is weak, the appropriate response is not necessarily to reject the intervention, but to pause, redesign, strengthen the implementation model, or sequence rollout more carefully.

How to interpret the matrix?

The matrix should be used to support three types of decisions.

- First, it can guide a go/no-go decision before an AI-CDST is introduced into real-world clinical settings. If the clinical problem is poorly defined, the AI value proposition is unclear, or the system cannot act on AI outputs, implementation should pause until the design is strengthened.
- Second, it can guide a pilot-to-scale decision. A successful pilot should not be judged only by usage numbers or technical outputs. It should demonstrate that providers can adopt the tool, patients can act on its recommendations, referral pathways can respond, data can be used for learning, and the institution can sustain the model.
- Third, it can guide a scale-to-sustainability decision. Once an AI-CDST is embedded in routine services, the system should continue to assess whether it remains safe, useful, affordable, equitable, and aligned with evolving clinical and institutional needs.

The readiness matrix therefore turns EMBED-AI into an applied implementation tool. It helps ensure that AI-CDST is not scaled merely because it is technically promising, but because it is clinically useful, operationally feasible, trusted by providers, connected to care pathways, governed responsibly, and capable of improving health outcomes in the settings where it is needed most.

Core Readiness Question	Minimum Requirement Before Scale-Up
Is the AI tool solving a clearly defined clinical or service delivery problem?	The intervention should have a written problem statement that identifies the target population, the current care gap, the clinical decision being supported, and why AI is justified compared with non-AI alternatives.
Will the AI tool expand access for underserved populations rather than duplicate services for already served?	Facility selection should prioritise locations where the tool fills a real access gap, such as facilities without specialist providers, diagnostics, or timely referral options. Potential exclusion risks should be identified and mitigated.
Is the AI tool aligned with how patients actually move through care?	The patient journey should be mapped from first contact to counselling, referral, follow-up, and outcome verification. Barriers such as transport, cost, family decision-making, fear, and lack of information should be addressed.
Can frontline providers use the tool within routine service delivery without creating unsustainable additional workload?	Provider tasks should be mapped clearly. The tool should be integrated into existing consultations, documentation, counselling, and referral routines wherever possible. Parallel reporting systems should be minimised.
Are the physical, operational, and staffing conditions sufficient for routine use?	Facilities should have minimum readiness in terms of space, privacy, electricity, device storage, supplies, connectivity, staff availability, and supervisory support. Where gaps exist, these should be addressed before rollout.
Are providers trained and supported to use the AI tool correctly and confidently?	A practical provider capability package should be in place, including hands-on training, supervised practice, refresher sessions, job aids, onboarding for new staff, and rapid troubleshooting support.
Are procurement decisions based on health-system value rather than product claims alone?	Procurement should assess clinical utility, evidence quality, usability, interoperability, implementation support, after-sales service, regulatory readiness, data rights, and exit options.

Core Readiness Question	Minimum Requirement Before Scale-Up
Do providers understand the tool's purpose, strengths, limitations, and appropriate use?	Providers should be able to explain what the AI tool does, what it does not do, how outputs should be interpreted, and when clinical judgement should override or supplement AI recommendations.
Is there a clear human-in-the-loop decision model?	The implementation model should clarify that AI supports but does not replace clinical judgement. It should define responsibility for final decisions, escalation rules, safety review, and management of uncertain or conflicting outputs.
Has the tool been assessed in the local implementation context?	Local validation should be conducted where feasible, especially for high-risk clinical decisions. False positives, false negatives, disagreement cases, and provider concerns should be reviewed and documented.
Does every AI output trigger a clear next step?	Each AI output or risk category should be linked to a defined clinical action, such as reassurance, counselling, follow-up, repeat assessment, referral, treatment, or emergency escalation.
Can the health system act on high-risk AI outputs?	Referral criteria, referral slips, receiving facilities, communication channels, feedback mechanisms, and follow-up responsibilities should be formalised before expansion. Detection without referral completion should not be treated as impact.
Are patients able to act on AI-supported advice?	Counselling, family engagement, transport support, financial support, community follow-up, or linkage with social protection platforms should be considered where patient-side barriers affect referral completion or continued care-seeking.
Are AI outputs connected to routine records and monitoring systems?	AI-generated data should be linked, where feasible, to existing EMRs, registers, dashboards, referral records, or health information systems. Client identifiers, facility identifiers, timestamps, risk flags, referral decisions, and outcomes should be standardised.
Are the routine systems in place to keep the tool functioning?	Implementation should include supplies, maintenance, charging, connectivity, device security, password management, technical support, issue escalation, and replacement planning. These should be budgeted and assigned to responsible actors.
Does the implementation model recognise provider effort and avoid unnecessary burden?	The tool should not create excessive documentation or unpaid additional work. Provider roles, task-sharing, supervision, recognition, and incentives should be aligned with routine service delivery.
Is there a structured support system beyond initial deployment?	There should be district or programme-level implementation support, including supervision routines, troubleshooting channels, performance review meetings, technical assistance, and feedback loops with product partners.
Is there a mechanism to improve the tool and workflow based on field evidence?	The implementation model should include dashboards, provider feedback, patient feedback, field observations, issue logs, adaptation logs, and regular learning reviews. Product and workflow changes should be documented and governed.
Does the product roadmap respond to real clinical and implementation needs?	Product partners should commit to improving usability, explainability, local language support, decision prompts, data export, interoperability, and clinically important features identified during implementation.
Are implementation outcomes being measured separately from clinical outcomes?	Evaluation should track acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability, in addition to clinical or service delivery outcomes.
Is there a clear public-system owner for the intervention?	The responsible institution should be identified, with clear authority over protocols, training, data governance, supervision, performance review, and scale-up decisions.
Is the long-term cost of implementation understood and funded?	Scale-up decisions should account for total cost of ownership, including procurement, licensing, consumables, maintenance, connectivity, training, supervision, product updates, and replacement.
Are safeguards in place for safety, privacy, accountability, and responsible use?	The implementation model should include consent processes, privacy safeguards, data-sharing agreements, auditability, clinical safety review, and clear rules for product updates or model changes.
Is the intervention ready to expand beyond pilot conditions?	Expansion should proceed only when the intervention has demonstrated acceptable feasibility, provider adoption, workflow fit, safety, referral connectivity, data use, operational support, and institutional ownership.
Is there a plan for pausing, replacing, or retiring the tool if needed?	The programme should define conditions under which the AI tool should be paused, redesigned, replaced, or discontinued, including safety concerns, poor adoption, unsustainable cost, vendor failure, or availability of a better alternative.

Contributors & Reviewers

We'd like to extend our thanks to the following experts who provided feedback and contributed to the development of this Learning Paper. The individuals listed below include reviewers, contributors, and individuals who were associated with the implementation of the AI-enabled ultrasound pilot in Sindh in different capacities. Their inputs, experience, and feedback informed the development of this learning paper. However, inclusion in this list does not imply endorsement of all content, analysis, or conclusions presented in the document. The final interpretation and responsibility for the learning paper rest solely with the authors.

PPHI Team

- **Dr. Zakir Ali Punar**
- **Dr. Kishwer Ali**
- **Dr. Mehreen Iqbal**

SSPA Team

- **Rashid Mehmood**
- **Asma Moomal**
- **Muhammad Sohaib**

CERP Team

- **Imran Zia**
- **Dr. Khizar Nasir**
- **Sidra Bashir**
- **Haseeb Sajid**
- **Muhammad Tahir**
- **Ali Zulkifal Ahmed**

Reviewers

- **Anam Shoaib**

Evidence Matters

29-P Gulberg II, Lahore 54660, Pakistan

www.cerp.org.pk

ali.gardezi@cerp.org.pk

+92 (0) 42 3577 7844 - 45