HTAi HEALTH TECHNOLOGY ASSESSMENT iNTERNATIONAL

HEALTH AI

A Blueprint for the Assessment of Digital and AI-Enabled Health Technology

DRAFT FOR REVIEW AND CONSULTATION



November 2026

DISCLAIMER – CONSULTATION DRAFT

This HTA Blueprint is a working draft made available for review, discussion and consultation. It does not represent a final or formally endorsed position.

The framework, principles and recommendations presented herein remain subject to further refinement based on stakeholder feedback and consultation outcomes.

Readers are encouraged to provide comments and suggestions to inform the development of the final version. This draft should not be regarded as definitive guidance.

Authors



Rado Andriantsimiavona

Director of AI, HealthAI - The Global Agency for Responsible AI in Health, Switzerland



Iakovos (Jacob) Levi

Health economist, HTAi Affiliate, Italy

Contributors

Grazia Antonacci

Implementation Lead, Imperial College London, United Kingdom

Lior Naamati-Schneider

Head of the Master's Degree Program in Management and Senior Lecturer, Department of Health Systems Management, Jerusalem Multidisciplinary College, Israel

Alisha Morsella

Public Health economist, Environmental sustainability, Italy

Rossella Di Bidino

Co-Chair, HTAi Artificial Intelligence and Data Interest Group, Italy

Iben Fasterholdt

Steering Committee member, HTAi Artificial Intelligence and Data Interest Group, Denmark

Wonjung Choi

National Evidence-based healthcare Collaborating Agency (NECA), Korea

Miyoung Choi

National Evidence-based healthcare Collaborating Agency (NECA), Korea

Clare England

Health Technology Wales (HTW), United Kingdom

Peiling Yap

Chief Scientist, HealthAI - The Global Agency for Responsible AI in Health, Switzerland

Contents

Executive Summary 1
The Blueprint at a Glance 2

1 Introduction 4

1.1 Digital and AI-Enabled Health Technologies/ Interventions and Health Technology Assessment 5
1.2 The Need for an Adapted Approach 8
1.3 The Blueprint 11

2 Assessment of Digital and AI-Enabled Technologies 18

2.1 Stage 1 – Decision framing and technology classification 19
2.2 Stage 2 – Assessment framework 22
2.3 Stage 3 – Assessment domains 26
2.4 Stage 4 – Appraisal and decision 38
2.5 Stage 5 – Monitoring and Reassessment 41

3 Assessment in Resource-Constrained and Emerging-HTA Settings 46

3.1 Rationale and stance 47
3.2 Transferability of external assessments 49
3.3 Repurposing evidence across decision systems 52
3.4 Regional and cross-national collaboration 54
3.5 Minimum assessment core 57
3.6 Profiling the decision context 63

4 Tools and Templates 68

4.1 Tool 1 – Decision-Context Card 70
4.2 Tool 2 – Evidence Gap Matrix 71
4.3 Tool 3 – AIHT Economic Evaluation Reporting Checklist 73
4.4 Tool 4 – AI Readiness Self-Assessment 76
4.5 Tool 5 – Managed Access Agreement 78

5 Governance and Setup of an AI-HTA Unit 82

5.1 Institutional entry models 83
5.2 Workforce and competencies 87
5.3 Institutional deployment 91
5.4 Institutional performance monitoring and review 95

Conclusion and Looking Ahead 101

References 106

Executive Summary

Digital and AI-enabled health technologies (AIHTs) have entered routine clinical use, yet decisions on their funding, procurement and adoption frequently hinge on regulatory approval or reported technical accuracy.

Neither indicates whether a technology yields a net improvement in care, service delivery or the use of resources in the health system or organisation that commissions it. Health technology assessment (HTA) provides that comparative judgement, although its conventional methods accommodate AIHTs imperfectly. Performance drifts after authorisation through retraining and dataset shift, estimates deteriorate on transfer between settings, and evidence rarely demonstrates effects on patient outcomes or resource use beyond accuracy on reference datasets. Generative, foundation-model and agentic systems compound these difficulties, since the unit of evaluation extends to the whole system assembled around the model and open-ended output admits no discrete reference standard.

The Blueprint consolidates existing frameworks and evidence standards into a single lifecycle process, while avoiding supplementing or supplanting any.

It traces the full market access sequence through which a technology progresses from the generation of supporting data to reimbursement, procurement and managed entry, interpreting each juncture in light of the properties of artificial intelligence. Across healthcare systems with varying levels of maturity, the targeted readership consists of those responsible for assessing technologies, issuing funding and adoption decisions, implementing them in clinical practice, and generating the information on which those decisions rest.

Assessment is structured in a consecutive five-stage process, which begins with the decision problem and the risk classification of the technology and concludes with its post-deployment surveillance and reassessment. The third step comprises the multidimensional evaluation, in which nine domains are each examined against predefined admissible evidence, acceptability threshold and endpoint measures. Of these, technical performance, safety, clinical effectiveness, economic value and regulatory compliance are designated core and included by default, whereas ethics, data governance, organisational impact and environmental footprint are classed as peripheral and invoked only where the exposed population, information sensitivity, workflow disruption or computational intensity justifies their inclusion. In an attempt to bolster auditability and comparability, standardised working templates are used to carry information across the stages, from framing the decision context to documenting the terms of managed access.

Specific consideration is directed toward health systems developing HTA infrastructure amidst resource scarcity.

The document sets forth tangible guidance on the transfer of HTA reports produced elsewhere, the reuse of regulatory evidence, and the profiling of the decision context, to which the depth of inquiry is calibrated. The work highlights an irreducible set of domains that every adaptation retains, for which proxy evidence and simplified analysis are admissible under unchanged acceptability standards. As full and condensed HTA alike rely on institutional readiness, the options and sequence of activities for the formation of a unit or agency able to apply the Blueprint in its entirety are further discussed, with reference to legal mandate, the alternative organisational models, the requisite competencies and the indicators by which the performance of the entity is reviewed.

INSTITUTIONAL GOVERNANCE AND CAPABILITY

From decision to reassessment

The Blueprint guides funding, procurement and adoption decisions for digital and AI-enabled health technologies. Assess their value in the setting where they will be used.

THE DECISION QUESTION
Will this technology improve care, service delivery or resource use in our setting?

1 Frame the decision

Define the question and classify risk.

Decision question and risk classification

- **Tool 1** Decision-context card

2 Set the assessment framework

Define PICO, perspective and time horizon.

Agreed assessment scope

- **Tool 2** Evidence gap matrix

Adapt to local context Use existing evidence, proxies or simpler methods where suitable. Retain all five core domains and the same acceptability standards.

3 Assess the domains

Apply the evidence and acceptability criteria for each applicable domain.

Findings, evidence gaps and uncertainty

- **Tool 3** AIHT economic evaluation reporting checklist

5 core domains
Include in every assessment.

See Chapter 02 for the domain names and triggers.

4 peripheral domains
Include when the relevant trigger applies.

Reassess and revisit earlier stages as needed

4 Appraise and decide

Weigh the evidence and specify any conditions

Decision question and risk classification

- **Tool 5** Managed access agreement

For conditional use

5 Monitor and reassess

Track agreed outcomes. Review when due or triggered.

Updated evidence and decision on continued use

- **Tool 4** AI readiness self-assessment

Informs decisions across the pathway

Institutional governance and capability

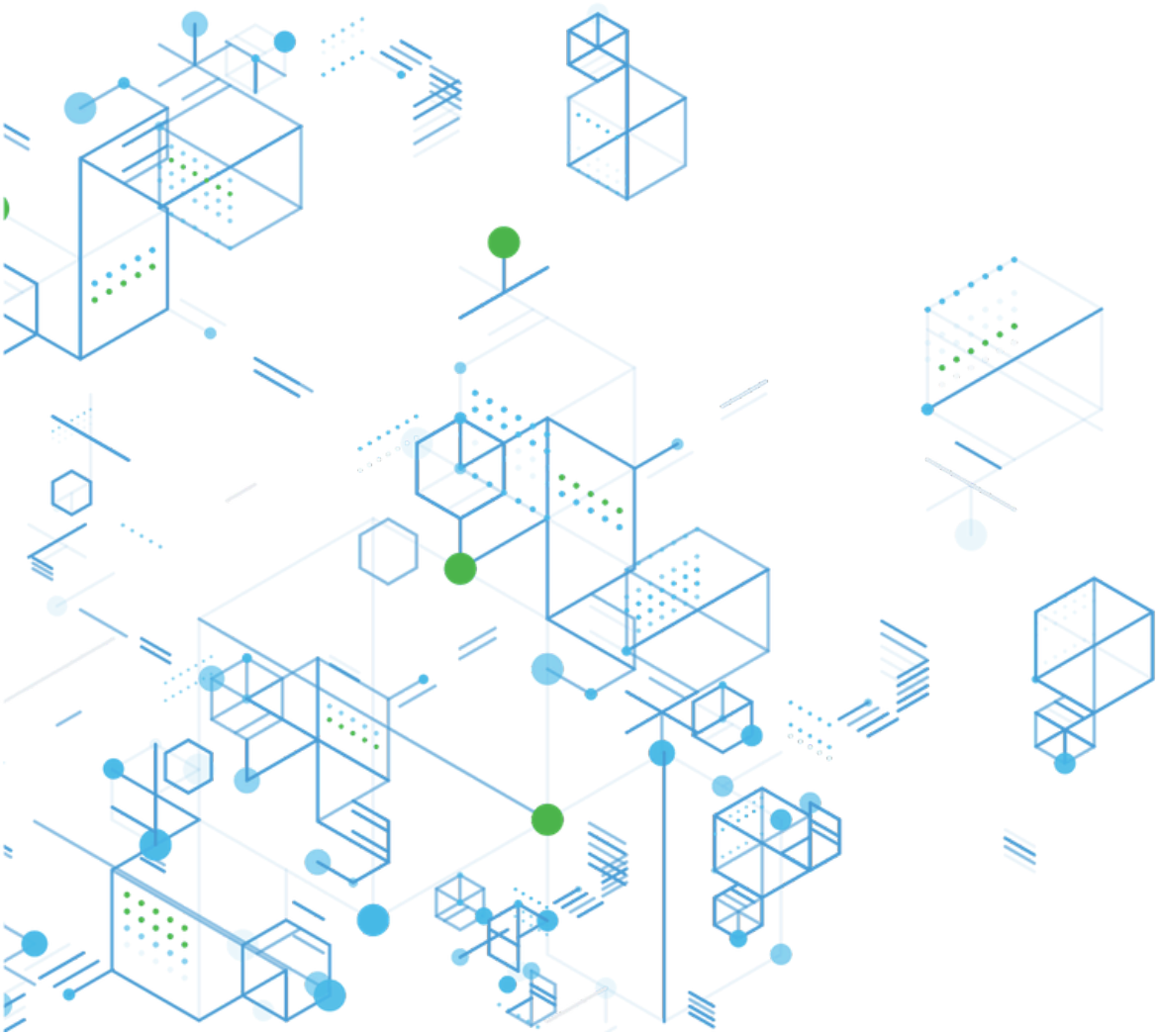
Mandate • Organisational model • Workforce competencies • Performance review

Supports all five stages



1

Introduction



1.1

Digital and AI-Enabled Health Technologies/ Interventions and Health Technology Assessment

Digital health uses information and communications technology to deliver health services and support health-system management. It encompasses eHealth and mobile health, as well as newer computational methods fields such as data science, genomics and artificial intelligence[1]. Within this broad field, the terms digital health technology (DHT) and digital health intervention (DHI) are often used interchangeably. A DHI refers more specifically to a digital function designed to achieve a defined health objective through software, an application or a connected device[2]. These technologies can support patients, the public, health professionals, as well as the management and data functions of health systems. They range from messaging-based reminders to clinical decision-support applications and remote-monitoring platforms[2].

An artificial intelligence-enabled health technology (AIHT) is a DHT whose main function depends on an AI model. For regulatory purposes, AI is defined as a machine-based system that produces predictions, recommendations or decisions capable of influencing real or virtual environments in pursuit of human-defined objectives[3]. When software performs a medical function independently of a hardware device, it may qualify as software as a medical device (SaMD); the regulated product is the software function that uses the model[4]. Two characteristics distinguish AIHTs from many other DHTs. First, their outputs depend on patterns learned from training data rather than solely on explicit rules. Performance therefore depends on the relevance and representativeness of those data and may not transfer across populations or settings. Second, the technology may change over time[5]. A model may remain locked, with its parameters fixed after authorisation, or be updated as new data become available, potentially changing its performance during deployment[6].

AIHTs vary widely in function, and these differences matter for assessment. Narrow AI systems perform a defined task, such as classification or detection in a single data modality, and account for most AI technologies authorised to date[7],[8]. Generative AI produces new content, including text, images or structured outputs. Foundation models are trained through self-supervision on broad, largely unlabelled datasets and can support different downstream tasks, sometimes without task-specific retraining; in medicine, this approach is often described as generalist medical AI[8].



Agentic systems connect models to external tools and allow them to carry out sequences of actions towards a defined goal with limited human intervention. These categories overlap, and a product may combine several of them. Nevertheless, their different functions, evidence requirements and risks must be identified before assessment begins.

These technologies have progressed beyond development into routine use, with 129 countries having adopted a national digital health strategy by 2025[9] and more than 350,000 health applications available worldwide, an increasing proportion of which are directed at the management of specific conditions[10], whilst the number of regulated AI-enabled medical devices continues to rise[7].

Although regulatory findings can inform HTA, regulatory approval is a separate process and cannot replace an assessment of comparative value.

Their growing use makes systematic assessment essential. Evaluation must consider not only what a technology does, but also its effects on patients, clinical services and the systems in which it operates. Health technology assessment (HTA) provides a structured basis for decisions about reimbursement, procurement and adoption. Health Technology Assessment International (HTAi) and the International Network of Agencies for Health Technology Assessment (INAHTA) define HTA as *'a multidisciplinary process that uses explicit methods to determine the value of a health technology at different points in its lifecycle, with the purpose of informing decision-making'*[11]. HTA compares a technology with relevant available alternatives, drawing on clinical, epidemiological and economic evidence. Clinical-, cost-effectiveness and safety are among its principal criteria. HTA is undertaken at national, regional and local levels, including within healthcare facilities[12].

Although regulatory findings can inform HTA, regulatory approval is a separate process and cannot replace an assessment of comparative value. Regulators, including agencies such as the European Medicines Agency and the United States Food and Drug Administration, assess whether a technology is sufficiently safe and performs as intended to enter the market. This benefit–risk judgement does not usually require comparison with existing alternatives[13]. On the other hand, HTA asks whether the technology offers value relative to the care already available in a particular health system, and whether it should be funded, procured or adopted within available resources[13].

Recent European policies have brought elements of the regulatory and HTA evidence bases closer together[14], but the purposes remain distinct: regulation governs market access, while HTA informs reimbursement and use. Where regulatory approval has already established a product characteristic, HTA can use that finding as a starting point and focus on its comparative implications. For safety, this means examining how harms differ from those associated with the comparator pathway and how that difference affects overall value.

The core principles of HTA therefore remain applicable to DHTs and AIHTs, without the need for wholesale methodological change. Several existing initiatives have already shown how established approaches can be adapted to these technologies.



1.2

The Need for an Adapted Approach

HTA assesses value, rather than simply determining whether a technology may enter the market. The question is therefore whether conventional HTA methods can be used to assess AI-enabled technologies without modification. Several characteristics distinguish these technologies from the medicines and fixed-function devices for which many established methods were developed. Each creates specific challenges for evidence generation and appraisal[15].

Assessment becomes a lifecycle activity rather than a one-off judgement.

1.2.1 Lifecycle change and versioning

AIHTs are not always stable interventions. After authorisation, their performance and value may change following retraining, software updates, new data sources, changes in decision thresholds or different levels of human oversight. This challenge is not limited to adaptive models that update their parameters. Even a locked model can lose accuracy when the population or data encountered in practice differ from those used in development, a phenomenon known as *dataset shift*[5],[16]. Performance must therefore be understood in relation to a particular version, setting and point in time.

Frameworks designed for fixed technologies do not readily accommodate these changes. Until recently, digital-health evidence frameworks generally excluded continuously updating algorithms[17]. Regulators have responded with predetermined change-control plans. Under these arrangements, manufacturers set out the changes they intend to make, how they will implement them and how their effects will be evaluated, allowing specified modifications to proceed without a new marketing authorisation[18],[19].

HTA bodies must accordingly record the version assessed, identify changes that could invalidate the supporting evidence and align reassessment triggers with regulatory change-control information. Assessment becomes a lifecycle activity rather than a one-off judgement, as the uncertainty introduced by each change can often be resolved through post-deployment monitoring[20] (§1.2.2).

1.2.2 Evidence and study-design limitations

A second challenge is the nature of the available evidence. AIHT studies often report discrimination or accuracy against a reference dataset without showing whether the technology changes clinical decisions, patient outcomes or resource use[21]. Technical validity is not the same as clinical utility, and establishing the link between them is central to assessing value. Where an AIHT claims to improve patient outcomes, a randomised controlled trial remains an important reference design, particularly for higher-risk or relatively stable technologies[21].

However, conventional trials may be difficult to apply when a system changes during the study period, and they cannot readily accommodate the volume and diversity of available tools. Evidence generation therefore needs to be proportionate to the question and the technology. Early live clinical evaluation can establish feasibility, safety and human-factors performance. Non-randomised comparative and registry studies can examine real-world effects, while randomised trials remain particularly important where patient-outcome claims require robust causal evidence[21]. AI-specific reporting standards support these approaches: SPIRIT-AI covers trial protocols[22] and CONSORT-AI covers trial reporting[23], while DECIDE-AI addresses early-stage clinical evaluation before definitive trials[24].

Many uncertainties can only be resolved once an AIHT is used in practice. Real-world evidence is therefore particularly important to lifecycle assessment. It should capture not only clinical effectiveness in routine populations, but also implementation fidelity, safety signals, subgroup performance, drift and realised costs. HTA agencies increasingly draw on this evidence when reassessing technologies[20].

Transferability should therefore be assessed through evidence from the intended setting, rather than assumed from development-site results alone.

1.2.3 Context and transferability

A third challenge is whether evidence generated in one setting applies to another. AIHT performance depends partly on the populations, sites and workflows represented in the development data. A model may perform less well after transfer because of differences in case-mix, equipment, disease prevalence, coding practices or care pathways, even if the model itself has not changed[16]. Sepsis prediction models, for example, have shown reduced performance when transferred between intensive care units within the same national healthcare system, where patient cohorts, labels and clinical protocols differed[25]. The gap may be greater when technologies developed in high-income settings are introduced into low- and middle-income countries. Reported problems include reduced accuracy, unequal performance and the transfer of biases to populations poorly represented in development data[26]. The viability of the technology in the receiving setting also depends on referral capacity, connectivity, maintenance, data governance and sustainable funding[26]. Transferability should therefore be assessed through evidence from the intended setting, rather than assumed from development-site results alone.



1.2.4 Generative, foundation-model, and agentic systems

Generative, foundation-model and agentic systems introduce a more fundamental question: what exactly is being assessed? A conventional DHT, or a narrow AIHT, can usually be described as a product performing a defined task. By contrast, foundation models support multiple purposes and can be configured in different ways. A single model can be adapted to many tasks without task-specific retraining[8]. Its behaviour depends on the prompts, retrieval sources, connected tools, safeguards, users and degree of autonomy. Different configurations built on the same underlying model may constitute substantially different interventions[27]. Assessment should therefore identify the type and configuration of the system before determining which methods and safeguards are appropriate. Agentic systems add another layer by carrying out sequences of actions rather than producing a single output.

The risks likewise extend beyond those usually associated with earlier AI systems, such as discrimination, privacy breaches and limited transparency. They include fabricated or unreliable content, malicious use, unsafe human-computer interaction, over-reliance, security failures and environmental burden[28].

For these systems, the appropriately defined assessment object is often the full configuration through which the model is used, not the underlying model alone. The configuration determines what the system does, how it acts and the conditions under which its value should be assessed[27].

Even with a clearly defined intervention, conventional performance measures may not apply. Generative systems can produce different responses to the same prompt, open-ended clinical tasks may have no single correct answer, and outputs vary with the phrasing of a query, the user and the setting[29]. Many established evidence standards assume a relatively stable technology with defined outputs; current HTA guidance has not yet fully addressed the assessment of generative AI[29].



1.3

The Blueprint

1.3.1 Purpose and objectives

This blueprint sets out a practical approach to assessing AIHTs so that decisions about funding, procurement and adoption are informed by evidence of value, not regulatory approval or technical performance alone. Because these technologies can change after deployment, assessment cannot be confined to a single decision point. The blueprint brings existing methods into a coherent sequence, explains the question addressed at each stage and identifies where professional judgement is required. It does not introduce a new assessment method or replace those used by national agencies and other institutions. Instead, it provides a common structure for selecting and applying established approaches, with sufficient flexibility to match the technology and the decision being considered.

No single existing framework covers every decision level, technology type, lifecycle stage and setting[30]. The blueprint therefore treats available approaches as complementary. NICE, for example, sets evidence requirements proportionate to a technology's intended purpose and risk[31], while European initiatives such as EDiHTA and ASSESS-DHT are developing more consistent assessment methods across jurisdictions[32],[33]. These resources can be combined within the blueprint without assuming that any one is universally applicable.

The focus is HTA of DHTs and AIHTs, not the use of AI to conduct HTA. AI is increasingly used to support evidence synthesis, data extraction, real-world evidence analysis, economic modelling and the preparation of assessment outputs[34]. Those applications fall outside the scope of this document. Whether an HTA team uses AI does not, in itself, justify favouring or disadvantaging a technology under review. Nor does greater productivity establish that an AI tool can safely replace human judgement in high-stakes scientific assessment, where appropriate human oversight remains essential.



1.3.2 Principles

The blueprint is guided by principles that apply throughout the assessment process. It aims to translate current HTA into a practical and adaptive governance process. Rather than calling for open-ended monitoring, it specifies the evidence to be collected, review intervals, thresholds for action and the scope of reassessment.



Assessment starts with the health-system need and the decision to be made, not with the technology in isolation.

P1

Because adoption changes how care is delivered, the relevant intervention is the AI-enabled care pathway compared with the pathway that would otherwise be used. Its effects depend on more than the algorithm: they also depend on users, interfaces, workflows, infrastructure, data and oversight. The assessment must therefore take into account these sociotechnical conditions. Where essential conditions for implementation are absent, expected value may not be realised. This is why the blueprint also addresses the governance and institutional capacity needed to assess technologies and sustain their safe use, rather than considering an individual appraisal alone.



Assessment requirements should be proportionate to the consequences of the decision

P2

.

The technology's intended use, degree of autonomy and potential for harm determine the scrutiny required. Higher-risk uses warrant more extensive evidence, while lower-risk applications should not face requirements that exceed what the decision needs. Proportionality also applies to resource consumption: computational and material demands should be considered alongside expected benefits.



An assessment can only address the outcomes it chooses to measure.

P3

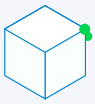
Without patient and public involvement, evaluations may emphasise clinical efficacy while overlooking treatment burden, access and the experience of care. **Patients and affected communities should therefore help define the questions and outcomes, particularly where a technology changes care pathways or affects underserved groups.** Assessors should also examine how benefits and harms are distributed. Training-data representativeness, accessibility and subgroup performance all matter when assessing whether a statistically significant population-level benefit may coexist with disproportionate harms in some subgroups and wider inequalities.



Changes in performance make one-off assessment insufficient.

P4

Each assessment should identify the version evaluated, the changes that would require review and the evidence to be monitored during use. Conclusions should remain open to revision as new information becomes available. Transparency is equally important: the evidence, assumptions, declared interests, use of AI within the assessment and reasons for the decision should be recorded so that others can examine the judgement. Recommendations must also specify practical conditions, accountable owners, timelines and monitoring responsibilities. Where a trusted authority has already completed a sound assessment of the same technology, its work should be reused where appropriate rather than repeated, while the receiving body remains responsible for its own decision.



Reassessment criteria should be defined in advance rather than left to ad hoc judgement.

P5

A recommendation remains valid under stated conditions and should be reopened when any of the following occurs:

i. The technology changes materially, including changes to the model, software, interface, decision threshold, or the retraining or replacement of the underlying model.

ii. The context of use changes, including a new purpose, population or site; greater autonomy; or different data, retrieval sources, connected tools or workflows.

iii. Performance crosses a predefined threshold, or a new signal emerges concerning safety, equity or security.

iv. The basis for the value judgement changes, for example through a new comparator, price or care pathway; evidence that challenges the earlier conclusion; or disruption to contractual supply arrangements.

1.3.3 Intended users

The blueprint addresses the actors involved across the access pathway, from evidence generation and assessment to appraisal, funding, implementation and reassessment. These responsibilities are usually shared among different organisations.

Its primary audience includes authorities responsible for reimbursement, procurement and adoption at national, regional and local levels; the HTA bodies that advise them; and the healthcare providers that implement decisions or conduct local assessments. Developers and manufacturers are also important users, as they generate and submit the evidence on which appraisal depends. Although these groups have different responsibilities, each needs to understand how the value of digital and AI-enabled technologies is established and maintained in practice.

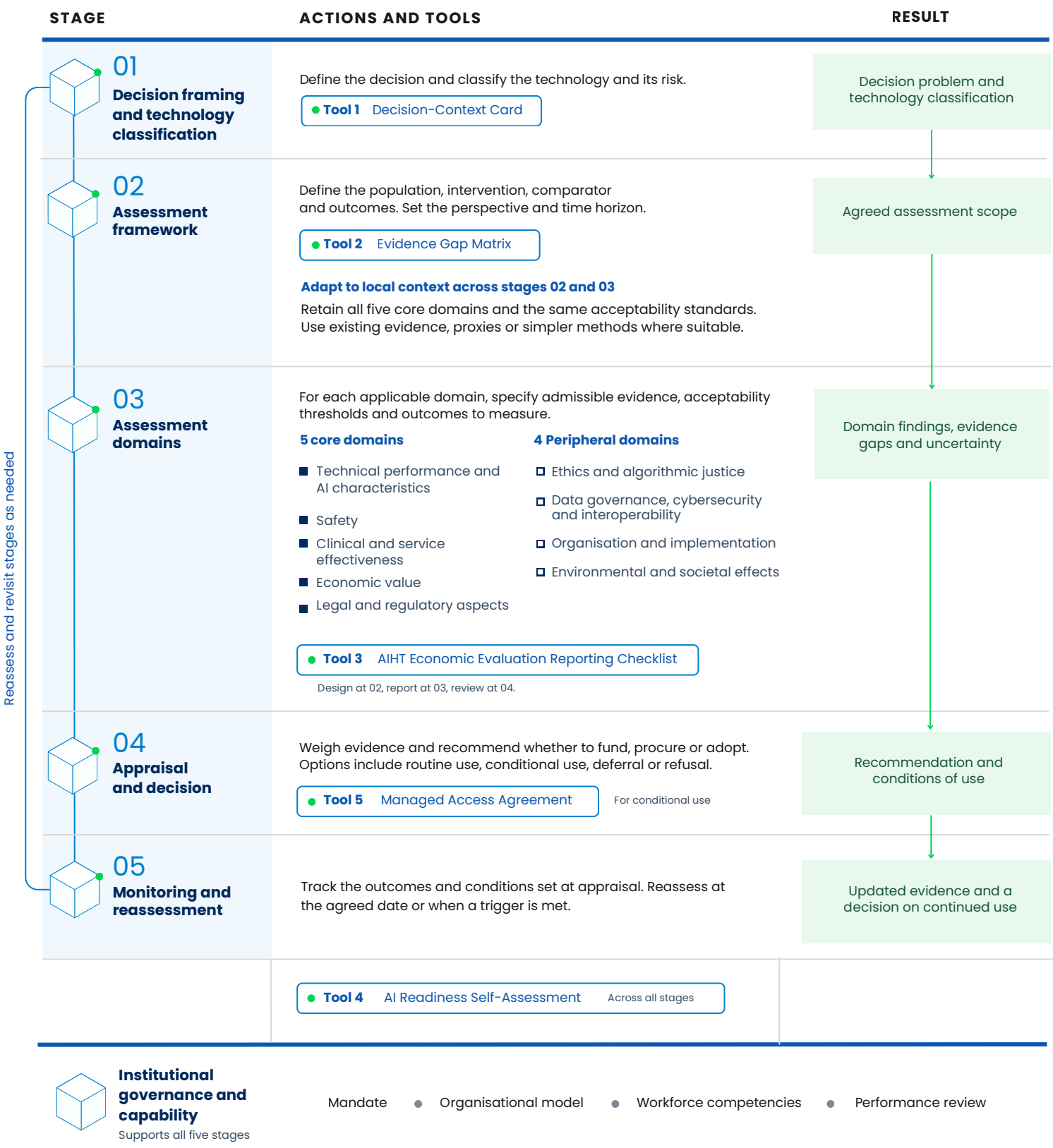
The blueprint is intended for both established HTA agencies and emerging or resource-constrained assessment functions. Its application should reflect the capacity and institutional maturity of the system concerned.HTA maturity describes how far assessment capabilities have been embedded in an institution or health system; it is not a prerequisite for using this blueprint[12]. Chapter 3 explains how methods can be adapted to different levels of capacity while retaining a minimum assessment core.

1.3.4 Architecture

The blueprint is organised around the five-stage assessment process set out in Chapter 2 and illustrated in Figure 1. The remaining chapters explain how the process is adapted, implemented and sustained.

The process begins by defining the decision and deciding whether assessment is needed. It then specifies the technology as used in practice, including the configuration around an AI model, and identifies the characteristics that determine the appropriate level of scrutiny. Evidence is examined across the relevant domains, including clinical and economic performance, ethics and equity, data governance, legal requirements and organisational effects. The findings inform an appraisal and a decision or recommendation, together with the conditions necessary for implementation. Once a technology enters use, monitoring can trigger reassessment when new evidence or predefined signals call its value into question.

Figure 1:
Five-stage process for the assessment of AIHTs and DHTs



Although the process is common to all settings, whatever the stage of HTA institutionalisation, assessment depth can vary. Chapter 3 explains how to adapt the approach to the maturity of the HTA body and the AI governance and infrastructure readiness of the deploying organisation. It also defines the minimum assessment core and the conditions under which evidence or assessments from other jurisdictions may be relied upon rather than repeated locally.

Chapters 4 and 5 address implementation from two perspectives. Chapter 4 provides practical templates for the different stages, including the decision-context card, evidence gap matrix, economic evaluation reporting checklist, AI readiness self-assessment and managed access agreement. These tools support the process without introducing separate evidentiary requirements. Chapter 5 covers the institution responsible for assessment: its organisational model, legal mandate, links to reimbursement and procurement, minimum competencies, phased development and performance evaluation.

Together, the five chapters set out the assessment method, its adaptation, and the tools and institutions needed to apply it.

Credible appraisal depends on separating the interests of those who generate evidence from those who assess its quality, judge its value and fund the technology.

1.3.5 Governance and decision model

The five-stage process produces a decision or recommendation, depending on the assessing body's legal mandate. To have practical effect, however, that output must identify who is responsible for making and implementing the decision and specify conditions that can be put into operation. The governance model therefore defines both the allocation of responsibilities and the decisions that an appraisal can support.

Governance arrangements vary across jurisdictions, including which actors take part and in what order[12],[35]. They were first developed mainly for pharmaceuticals, but are also used for medical devices and other health technologies[36],[37]. The following description sets out the common roles.

Credible appraisal depends on separating the interests of those who generate evidence from those who assess its quality, judge its value and fund the technology[12],[35]. Manufacturers and developers produce the supporting dossier, either through internal teams or specialist organisations, including those undertaking economic evaluation on their behalf[12],[35]. An assessment body reviews the evidence, and the appraisal process places the findings in the context of the decision-maker's objectives[35],[38]. Assessment and the wider judgement of value should have appropriate safeguards and, where possible, distinct responsibilities[35]. In practice, one public organisation, such as the UK's National Institute for Health and Care Excellence (NICE), may oversee both functions[38]. A favourable recommendation does not automatically secure funding[12],[35]. The payer and manufacturer may still need to agree prices and access terms before the technology is included in funded care[12],[35]. Oversight bodies establish the mandate, priorities and methodological standards; providers then implement the decision and monitor the technology in use[12],[35].

Accountability follows this division of responsibilities. The appraisal body is responsible for the integrity of its methods and conclusions[35],[38], while the payer or procurement authority makes the coverage or purchasing decision within its legal mandate[12]. The manufacturer must maintain current evidence and document successive model versions[19],[39]. In practice, providers often monitor deployed technologies, while appraisal authorities undertake comparatively little routine reassessment[40]. Lifecycle approaches increasingly call for appraisal bodies to take responsibility for reassessment, working with providers to identify changes in value that manufacturers may have little incentive to report[41].

Decisions need not be limited to approval or refusal. The appropriate recommendation depends on the scale of remaining uncertainty about the technology's expected value, whether further evidence can reduce it and the consequences of committing resources before that uncertainty is resolved[36],[37]. Where expected value is positive but important uncertainty remains, provisional funding can allow evidence generation alongside controlled adoption, while preserving the option to reverse the decision[36],[37]. Use may be restricted to a subgroup where benefit is confined to that population[36],[37]. Uncertainty about local implementation or transferability may justify a supervised pilot[37], while unproven clinical utility may require use to remain within research[36]. Negative expected value or unmanageable risk supports refusal[36],[37]; an already funded technology may be withdrawn if changes in evidence, price or the comparator undermine its value[36]. Each decision should state the eligible population, authorised model version, required oversight, implementation conditions and reassessment date. The recommendation must define the terms and duration of use, not simply whether access is granted[36],[39].

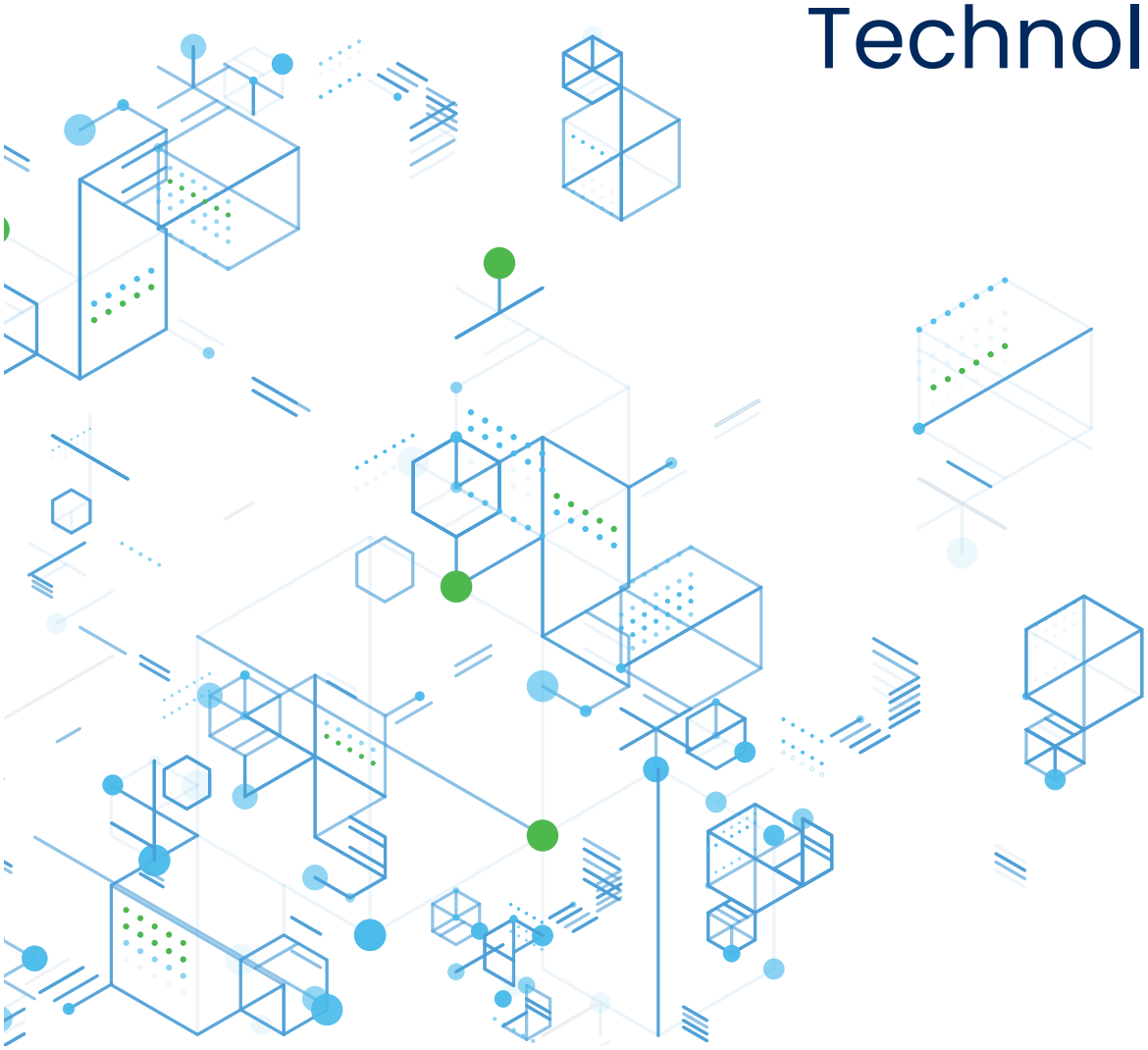
The available decision types and the conditions attached to them are outlined in Stage 4 of the assessment process highlighted in Chapter 2, while Chapter 5 examines the institutional roles and mandates behind the aforementioned governance model.




X-Max
50mm²
EL
WDS
EBSD
Speed
Productivity
Performance
Accuracy

2

Assessment of Digital and AI-Enabled Technologies

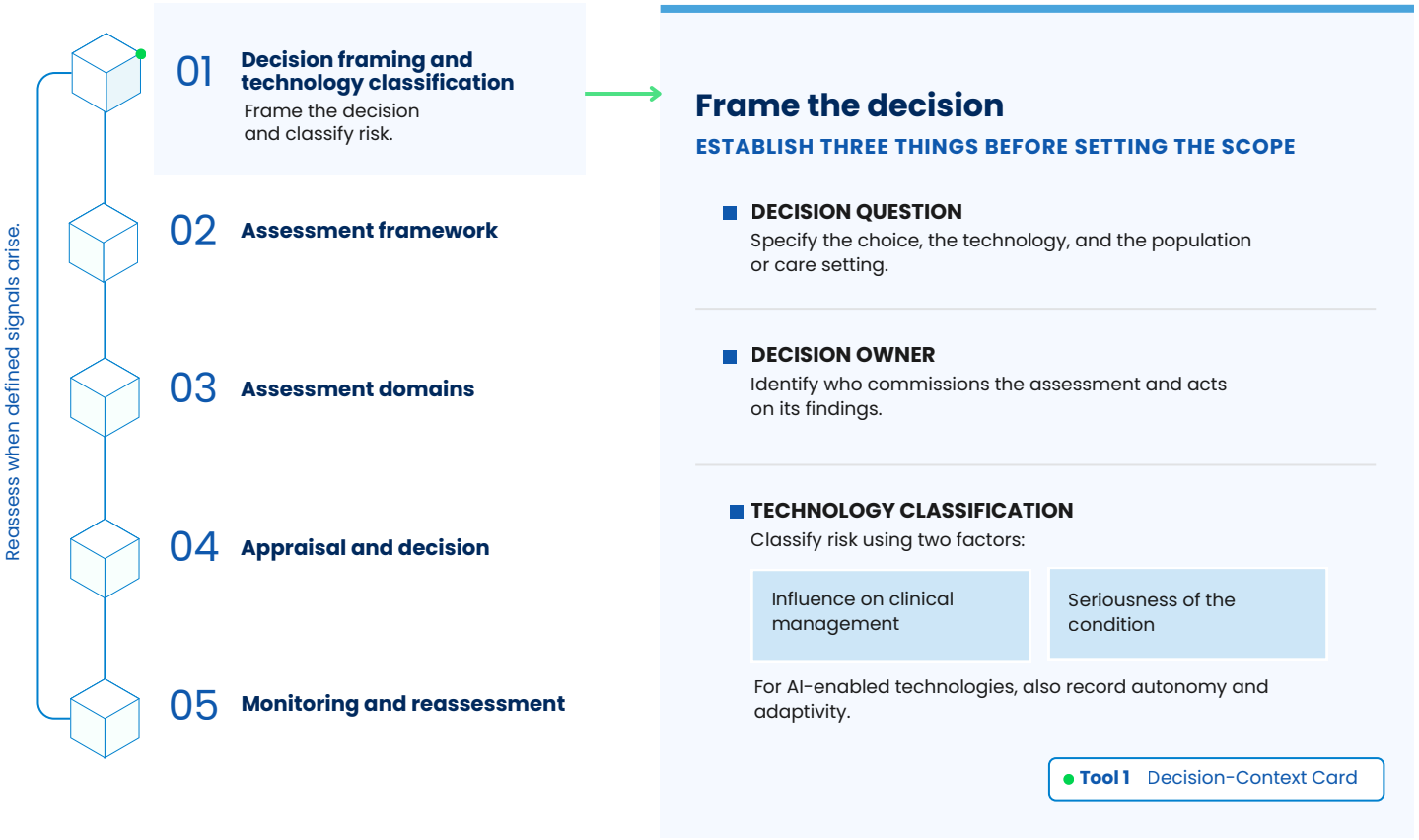


2.1

Stage 1 – Decision framing and technology classification

2.1.1 The decision question

An assessment begins with a specific decision problem, commonly described as the policy interest, which determines its scope[42]. The policy question identifies the decision to be informed and the health-system need behind it[42]. The research question then specifies the choice being considered, the technology and the relevant population or care setting[42]. Decisions may concern marketing authorisation, a controlled pilot, procurement, reimbursement or coverage, local adoption, scale-up or disinvestment. The type of decision determines the appropriate evidence requirements and depth of analysis.





2.1.2 The decision owner

Every assessment should have one accountable decision-maker who commissions the analysis and acts on its findings. This may be a reimbursement authority, a regulator or a healthcare provider making a local procurement decision. The decision-maker's mandate determines which costs and consequences matter. It also helps maintain the distinction between assessment, which examines the evidence, and appraisal, which judges the technology's value.

2.1.3 Classification of the technology

Technology classification comes before assessment because evidence requirements should reflect the risks associated with intended use. The International Medical Device Regulators Forum (IMDRF) approach considers two factors: how strongly the information provided influences clinical management, ranging from informing decisions to directing diagnosis or treatment; and the seriousness of the condition, from non-serious to critical. Together, these determine the risk category and the corresponding level of scrutiny[43]. For AIHTs, two further characteristics matter: autonomy, ranging from assistive to autonomous operation, and adaptivity, distinguishing locked models from those that continue to learn after deployment[19]. Figure 2 shows these classification criteria and their implications for the Stage 3 assessment.

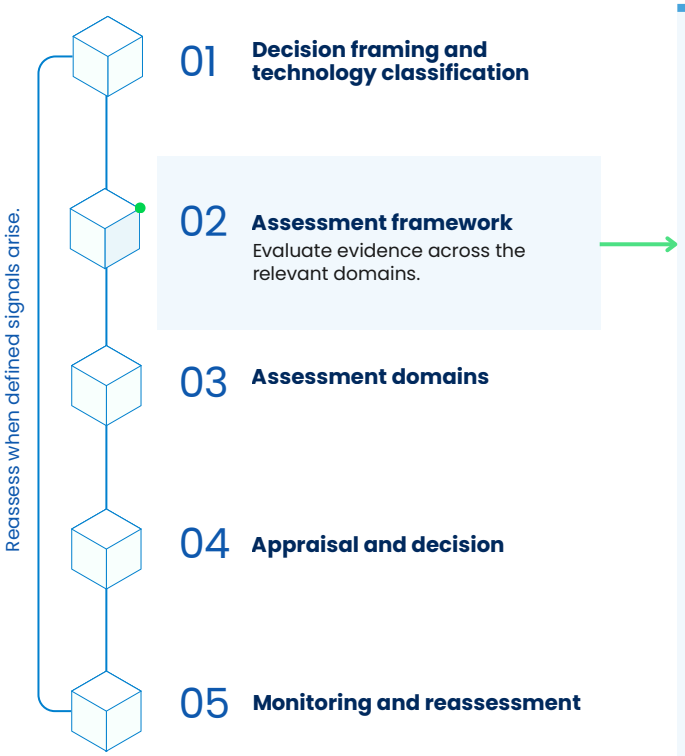
Figure 2:
AIHTs Stratification and assessment implication matrix

CLASSIFICATION AXIS	TIER	ASSESSMENT IMPLICATION
Information significance and condition state *(IMDRF category)*	**Category I** informs in a non-serious or serious condition; drives in a non-serious one	Core domain set. Technical performance and diagnostic accuracy may suffice.
	Category II treats or diagnoses in a non-serious condition; drives in a serious one; informs in a critical one	
	Category III drives in a critical condition, or diagnoses in a serious one	Full domain set. Requires evidence at the patient-outcome level.
	Category IV diagnoses or treats in a critical condition	
Autonomy *(assistive to autonomous)*	**Assistive** informs, the clinician decides	Unit of assessment set at the AI-enabled task, the tool with the immediate human step (ii).
	Augmentative produces output the clinician acts on	
	Autonomous concludes without concurrent clinician involvement	Unit set at the output with its triggered action (i), extending to the full pathway (iii) where the tool acts without a human gate.
Adaptivity (locked or continuously learning)	**Locked** fixed after release	Reassessment at change or on a periodic schedule. No version monitoring.
	Continuously learning updates during use	Continuous version monitoring and a predetermined change-control plan. Higher reassessment frequency.

2.2

Stage 2 – Assessment framework

Stage 2 turns the policy question into an assessment research question using the PICO framework: Population, Intervention, Comparator and Outcomes. It also specifies the analytical perspective and time horizon, which determine whose costs and consequences are included and for how long they are measured[44]. Together, these elements define the scope, boundaries and evidence requirements for the Stage 3 assessment[42]. They should be agreed before reviewing the available evidence. One technology may require several PICOs if it is used in different populations or compared with different care options; each combination represents a separate research question[42].



Set the assessment scope

P I C O

TURN THE POLICY QUESTION INTO A PICO RESEARCH QUESTION

P POPULATION
Who is covered, in which setting, and at what point in care?

I INTERVENTION
What technology and intended use are being assessed?

C COMPARATOR
What would care look like without the technology?

O OUTCOMES
Which endpoints will show its effects?

Perspective	What costs and consequences count?
Time horizon	How long is needed to capture material differences?

Tool 2 Evidence Gap Matrix

2.2.1 Population

The population defines the patients covered by the decision. It should reflect the technology's intended use and the local context, including the condition, care setting and point in the care pathway at which the technology is used[45]. Subgroups should be specified in advance when baseline risk or treatment effect is expected to differ. Defining them after examining the results can introduce bias and should not determine the primary assessment[42],[45].

For AIHTs, the population on which the model operates needs particular attention. The eligible cases encountered in practice may differ from those represented in its training data. Assessors should describe the intended operational population, including cases the model excludes or refers for human review. Such differences matter because a mismatch between development and deployment populations can reduce performance.

2.2.2 Intervention

The terms *intervention* and *technology* are used interchangeably in this blueprint. Both refer to what is being assessed, defined by its intended use and place in the care pathway. Conventional comparative assessments generally specify a distinct intervention against which clinical effects can be measured. Systems-thinking approaches have, however, challenged the assumption that a technology can always be assessed meaningfully in isolation[46],[47].

This distinction is especially important for AIHTs, whose effects may arise at different points in care. Assessing the algorithm alone may miss how its outputs influence users and subsequent clinical actions. An AIHT should therefore be considered as a sociotechnical intervention, with effects shaped by the interaction between the model, its users and the care pathway[48],[49]. The unit of assessment should capture the level at which these effects occur.

For a given research question, the unit of assessment should be defined at one of three nested levels:

i. The model output and the action it prompts, such as a diagnosis, risk score or triage decision.

ii. The AI-enabled task at a particular decision point, including the model and the immediate human action it changes, such as detecting and verifying a finding.

iii. The full AI-enabled care pathway, from patient presentation to outcome, including downstream effects.

The unit should be as narrow as possible while still capturing the relevant health consequences. The technology's autonomy, established during classification, informs which level of assessment is appropriate.

When AI informs a clinician's decision, the relevant comparison is between the clinician working with and without the tool, rather than between the model and the clinician.

2.2.3 Comparator

The comparator is the treatment or care pathway against which the technology is evaluated. It will usually be the current standard of care without the technology in the health system concerned[42],[45]. The assessment should describe that standard in sufficient detail to allow a meaningful comparison[42]. Where current practice involves no active intervention, watchful waiting or best supportive care may be the appropriate comparator[42].

When AI informs a clinician's decision, the relevant comparison is between the clinician working with and without the tool, rather than between the model and the clinician. Standalone accuracy demonstrates technical performance but does not establish clinical utility[48],[49]. For diagnostic technologies, the comparison must also capture downstream tests, treatment and monitoring that may change as a result of the test[45]. The comparator should be defined at the same level as the intervention in Section 2.2.2.

2.2.4 Outcomes

Outcomes are the endpoints used to evaluate a technology's effects, including clinical effectiveness and safety[11]. A final endpoint measures a consequence of direct interest, such as a patient health outcome. A surrogate endpoint measures an intermediate result and should be used to infer a final outcome only when the relationship is sufficiently validated[45]. This distinction is particularly important for DHTs and AIHTs. Measures such as sensitivity and specificity describe technical performance, but do not, by themselves, demonstrate patient benefit. Their relevance to the value assessment depends on an established link to downstream outcomes[45].

Outcomes also differ in how they are expressed. Incremental measures describe the change attributable to the technology relative to its comparator, while proportional measures express the frequency of an event against a defined denominator, such as an error or failure rate across cases[50]. The former describe a difference in effect; the latter describe a rate or probability. Neither should be confused with ratios combining different quantities, such as the incremental cost-effectiveness ratio (ICER)[51].

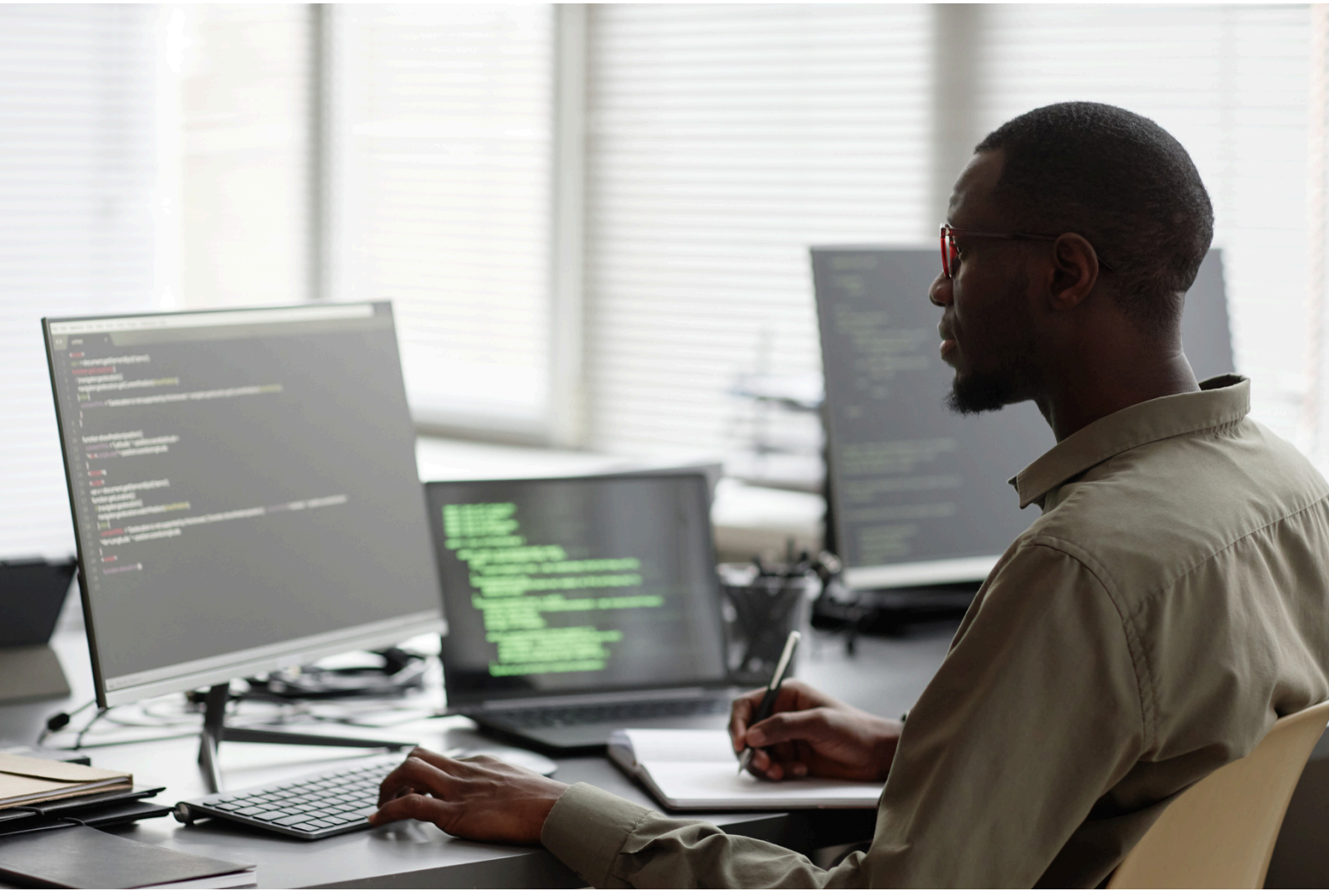
Attributing an outcome to a technology can be difficult when it forms part of a wider change in prevention or treatment. This is especially true when the intervention is an entire AI-enabled pathway influenced by several concurrent changes. Stage 3 matrix sets out the outcomes relevant to each assessment domain.

2.2.5 Perspective and time horizon

The analytical perspective determines whose costs and consequences count in the assessment. Many jurisdictions use the publicly funded healthcare payer or health-system perspective as their reference case[45],[50]. A societal perspective also considers costs and consequences affecting patients, carers and other sectors, often as a secondary analysis when these effects matter to the decision[50].

This blueprint adopts a broader societal perspective because AIHTs can alter workflows, professional roles and equitable access to care in ways that a payer perspective may not fully capture. If a narrower perspective is used, relevant exclusions should be acknowledged[44]. The perspective is often established when the decision is framed at Stage 1.

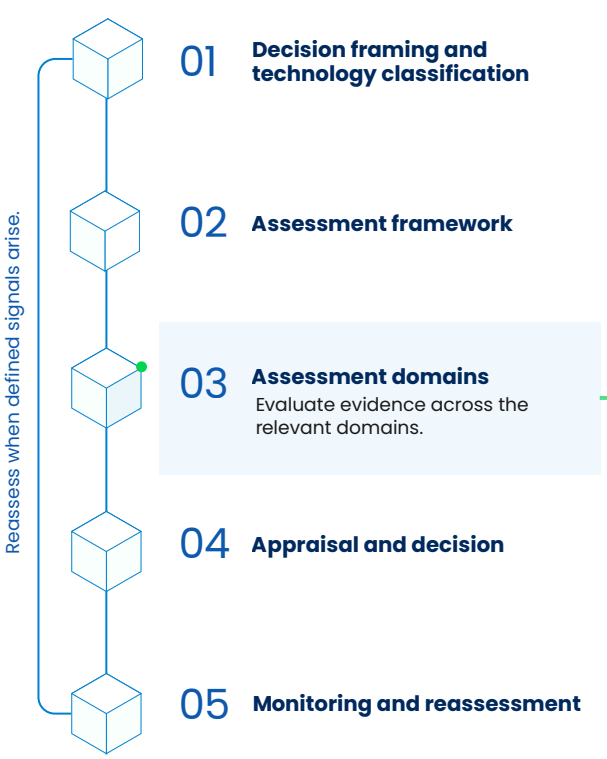
The time horizon should be long enough to capture all material differences in costs and outcomes between the technology and its comparator[44],[45],[50]. It should reflect the condition, care pathway and expected duration of the technology's effects. Choosing a shorter period simply because long-term data are unavailable risks missing consequences that emerge later[44]. This is particularly important for DHTs and AIHTs: performance may change after deployment, while clinical and service effects may extend well beyond initial use. Published economic evaluations often use short horizons, sometimes limited to the trial period or a single year. These have been identified as limitations, not as a methodological standard. The appropriate horizon should therefore be justified for each assessment[52],[53].



2.3

Stage 3 – Assessment domains

Stage 3 evaluates the evidence across the relevant assessment domains. As shown in Figure 3, the assessment matrix records the subdomains, admissible evidence, analytical methods, acceptability rules and outcome metrics for each domain. The same matrix applies across the technology classes defined in Section 2.1.3, although classification determines which subdomains are relevant and how closely they should be examined.



Nine assessment domains

5 CORE DOMAINS

Recommended for every assessment

- Technical performance and AI characteristics
- Safety
- Clinical andservice effectiveness
- Economic value
- Legal and regulatory aspects

4 PERIPHERAL DOMAINS

Triggered by the technology and its context

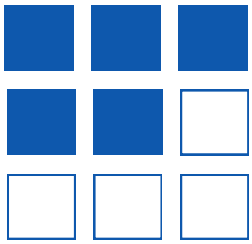
- **Ethics and algorithmic justice**
  Large or vulnerable populations
- **Data governance, cybersecurity and interoperability**
  Sensitive data or extensive integration
- **Organisation and implementation**
  Major workflow or infrastructure changes
- **Environmental andsocietal effects**
  Substantial scale or computing demands

Tool 3 AIHT Economic Evaluation Reporting Checklist

Design at 02, report at 03, review at 04.



Nine assessment domains



5 core domains

4 peripheral domains

The matrix contains nine domains. Five are core and are recommended for every assessment, namely technical performance and AI-specific characteristics; safety; clinical and service effectiveness; economic value; as well as legal and regulatory aspects. Evidence requirements for safety and clinical effectiveness become more demanding as the IMDRF risk category increases. They range from technical validation for tools informing low-risk decisions to patient-outcome evidence for technologies directing care in critical conditions. The other four domains are peripheral in the sense that their assessment is triggered by the technology and its context. Environmental and societal effects become relevant at substantial scale or computational intensity; ethical aspects and algorithmic justice where exposed populations are large or vulnerable; data governance, cybersecurity and interoperability where sensitive information or extensive integration is involved; and organisational and implementation aspects where workflows or infrastructure change materially. All domains should be considered against the comparator pathway. Where regulatory approval has established a product characteristic, the assessment can use that finding and focus on the additional effects and value arising in the care pathway.

The acceptability rule determines how each domain contributes to the overall judgement at Stage 4. Most rules are compensatory: a disadvantage in one domain may be weighed against benefits in another, taking account of opportunity cost. Other subdomains have pass–fail requirements that cannot be offset by gains elsewhere. These may apply where a minimum requirement is essential even without a specific regulatory threshold, including some equity or environmental considerations. Affordability subdomain best illustrates this binary admissibility filter, since a technology rated cost-effective on its incremental ratio may still have its funding deferred and its price reopened through commercial negotiation once its net budget impact exceeds the threshold at which the health system's capacity to absorb it is in question[45].

Figure 3:
AIHTs/DHTs Assessment matrix

5 CORE DOMAINS

TECHNICAL PERFORMANCE AND AI-SPECIFIC CHARACTERISTICS

CORE DOMAIN

How accurately the model does its task, and whether that holds across populations and over time.

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Standalone model performance (pass-fail)	• Model output and reference-standard result per case • Held-out test set from the development sites, independent of model training • Comparator result on the same cases • Demographic and clinical subgroup attributes per case • Predicted probabilities and observed outcomes for each case in an independent validation dataset representative of the intended-use population • Model version and operating threshold	Diagnostic or predictive accuracy study	• Sample size calculated at Stage 2 for the required precision • Lower 95% CI of the difference against the comparator above -Δ in each sample Δ: the largest loss against the comparator the decision tolerates, set and justified at Stage 2.	• Sensitivity, specificity • AUROC • Difference against the comparator, 95% CI
		Calibration analysis	Calibration intercept and slope assessed against their ideal values (0 and 1, respectively), supported by visual inspection of the calibration plot.	Calibration plot, slope and intercept, 95% CI
		Subgroup analysis	Sensitivity, specificity and AUROC in each subgroup within Δ of the overall estimate	Subgroup minus overall difference
		Risk-of-bias and applicability assessment	No domain at high risk of bias	Risk-of-bias rating by domain
External validity and transferability (pass-fail)	• Performance estimates per site, from sites or periods outside development • Case mix, prevalence and input-data profile per site • Equipment, software version and input language per site	Random-effects meta-analysis across sites	Lower bound of the 95% prediction interval above the standalone margin	95% prediction interval of the site-level difference against the comparator
		Transportability assessment	Proportion outside validated scope within a justified Stage 2 limit; no unresolved safety-critical mismatch	Proportion of local cases outside the validated range

Technical performance and AI-specific characteristics (*continued*)

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Human–AI team performance (pass-fail; assistive and augmentative tools only)	• Final decision per case, with and without the tool • Reference-standard result per case • Model output per case, marked correct or incorrect • User experience level	Paired multi-user, multi-case study, with and without the tool	Lower 95% CI of the difference against unaided decisions above -Δ, overall and per experience level	Proportion of correct final decisions with and without the tool, separately in cases with and without the target condition
Lifecycle stability, versioning and drift (pass-fail; failure confines the decision to the assessed locked version)	• Version identifier of the assessed model • Update history, with the type of each change	Version-control review	Assessed version uniquely identified and restorable through a documented rollback	Version identifier and rollback record
	• Performance of each released version on a fixed reference test set • Change-control plan • Drift-monitoring plan • Rollback procedure	Comparison of successive versions on the fixed reference test set	Lower 95% CI of the difference between each update and its predecessor above -Δ	Difference in the primary performance metric between successive versions
		Change-control plan review (adaptive models)	Every permitted change type carries a revalidation protocol and an impact assessment	Proportion of permitted change types with a revalidation protocol
		Drift-monitoring plan review	Alert threshold set at a decline of no more than Δ from the assessed performance	Alert threshold, as decline from the assessed performance
Generative and agentic output performance (pass-fail; generative, foundation-model and agentic systems only)	• Generated output per case on a pre-specified case set • Set-up tested: model version, prompt, retrieval sources, tools, safeguards • Reference answer or source record per case • Comparator output for the same cases • Reworded and repeated prompts per case • Retrieval and tool-call logs • Action logs from simulated scenarios	Blinded rating of outputs against the reference by two independent experts, with disagreements adjudicated	Lower 95% CI of the difference in error-free outputs against the comparator above -Δ	Proportion of outputs free of fabricated or omitted clinically relevant content
		Prompt-variation testing: each case re-run with reworded and repeated prompts	Lower 95% CI of the difference in error-free outputs between varied and reference prompts above -Δ	Difference in error-free outputs between varied and reference prompts
		Retrieval and tool-call audit against the correct source or tool per query	Error rate below the performance goal set at Stage 2	• Proportion of retrievals returning the correct source • Proportion of correct tool calls
		Agentic action audit in simulated scenarios, including requests outside the permitted scope	No action outside the permitted boundary	Actions outside the permitted boundary per scenario set

SAFETY

CORE DOMAIN

The harm the technology can cause when it errs, and whether oversight catches it in time.

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Clinical hazards and residual risk (pass-fail)	• Manufacturer risk-management summary and IFU • Regulatory findings; critical-hazard controls and verification	• Clinical risk-management review • Grade severity and likelihood; verify controls and residual risk	• 0 unresolved unacceptable residual risks • 100% identified critical hazards have documented controls	• Unacceptable residual risks (n) • Critical hazards with verified controls (%)
Observed patient harm	• Serious harm outcomes from comparative studies, including those reporting effectiveness • Event counts with exposure denominators for both arms • Manufacturer vigilance reports and recalls • Regulatory safety alerts • Local incident/corrective-action record	• Comparative safety-event analysis • Absolute risk difference and 95% CI where estimable	Upper 95% CI of excess serious harm < +Δ. Δ: set at Stage 2; no unresolved unacceptable signal	• Serious events per 1,000 uses • Risk difference per 1,000; 95% CI
		• Post-market safety-signal review • Verify applicability, investigation and closure	• 0 unresolved critical alerts affecting intended use • All material signals investigated or use restricted	• Open critical alerts (n) • Unresolved serious incidents (n) • Corrective actions closed (%)
Fail-safe operation and human oversight (pass-fail)	• IFU: required user confirmation, override and downtime procedures • Summative usability validation report (IEC 62366-1)	Simulated incorrect output and missing output and system downtime with intended users, recording detection, response and time to fallback.	• 100% of simulated critical failures detected and overridden or escalated • Fallback to the comparator pathway within the clinically required time • 0 unresolved critical use errors	• Simulated failures safely managed (%) • Time to fallback (minutes) • Critical use errors (n)

CLINICAL AND SERVICE EFFECTIVENESS

CORE DOMAIN

The effect of the AI-enabled pathway on patient outcomes and on the service that adopts it.

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Patient-relevant clinical effectiveness (where clinically relevant)	• Patient-relevant outcome per patient, technology versus the same pathway without it: mortality, morbidity, quality of life or a PROM • Allocation per patient or site, and baseline characteristics per patient • Time from entry to the pathway to outcome or end of follow-up, per patient • Change in detection or management per patient, the published effect of that change on the outcome, and the baseline event rate (where outcomes are not observed)	Comparative outcome analysis, randomised (individual or cluster) Preferred where available	Superiority: lower 95% CI of the difference against the same pathway without it above 0, estimate ≥ M Non-inferiority: lower 95% CI above -Δ M: smallest clinically important difference Δ: largest loss against the comparator the decision tolerates. Both are set and justified at Stage 2	• Absolute difference per 1,000 patients, 95% CI • RR, HR or PROM mean difference, 95% CI
		Comparative outcome analysis, non-randomised (regression, propensity score or interrupted time series) Where no randomised evidence exists	Thresholds as above, applied to the adjusted estimate.	• Adjusted absolute difference per 1,000 patients, 95% CI • Adjusted RR or HR, 95% CI

Clinical and service effectiveness (continued)

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Patient-relevant clinical effectiveness (where clinically relevant)		Linked-evidence analysis: change in detection or management chained to outcome Where outcomes are not observed.	• Each link supported by evidence applicable to the intended-use population • Lower bound of the 95% uncertainty interval meets the thresholds above Reported as inferred benefit.	• Modelled events avoided per 1,000 patients, 95% uncertainty interval

ECONOMIC VALUE

■ CORE DOMAIN

What the pathway costs, whether its benefit justifies that cost, and whether the budget can absorb it.

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Cost-effectiveness	• Cost per patient, technology versus the same pathway without it, including set-up, integration, compute, monitoring and retraining • Health outcome per patient under each, QALYs where estimable, otherwise a clinical endpoint • Resource use and unit costs	• Cost-utility analysis, health in QALYs Preferred where a QALY gain can be estimated.	• ICER below the local cost-effectiveness threshold • Or the technology cheaper and at least as effective	• Cost per QALY gained (ICER) • Net monetary benefit
		• Cost-effectiveness analysis, health in a clinical endpoint Endpoint: a downstream event, case detected that alters management or event avoid.	• Judged against the same pathway without the technology, or an agreed reference value No generic threshold in a clinical unit.	• Cost per unit of the clinical endpoint, per case detected or event avoided
Budget impact and affordability	• Size of the eligible population and expected uptake • Unit price of the technology and of current care • Per-patient resource use under each [source cut off]	• Budget impact analysis Payer perspective. One to five year horizon. Uptake tested as scenarios	• Net cost within the payer's available budget over the horizon Where uncertain, phased or managed entry.	• Net annual and cumulative cost • Cost per patient

LEGAL AND REGULATORY ASPECTS

■ CORE DOMAIN

Whether the technology is authorised for its intended use, who answers when it fails, and its conformity with AI and data law.

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Regulatory status and intended-use conformity (pass-fail)	• CE or public regulatory register entry, certificate, declaration of conformity or documented exemption • IFU: intended purpose, risk class and authorised version	• Comparison of the authorised intended purpose and version with the population, use, setting and version assessed	• All required authorisations valid, or exemption documented • Assessed use within the authorised intended purpose • Assessed version authorised, or within the authorised change-control plan	• Regulatory status (valid/exempt/missing) • Uses outside the authorised intended purpose (n) • Version match (yes/no)
Data protection and patient rights (pass-fail)	• Privacy or data protection impact assessment (e.g. DPIA under GDPR; ISO/IEC 29134), or documented reason none is required • Data processing agreement or equivalent between deploying organisation and vendor • Patient privacy notice; data-flow and hosting location; cross-border transfer basis where required • Contract terms on vendor reuse of patient data, including for model training	• Review of the documentation against the applicable data protection law	• Lawful basis documented for each processing purpose, including any reuse for model training • Agreement and, where applicable, transfer mechanism in place • No unmitigated high risk in the DPIA	• Missing or non-compliant documents (n) • Unmitigated high risks (n)
Legal accountability and AI-specific obligations (pass-fail)	• Contract between deploying organisation and vendor: liability, warranty, incident-reporting and update obligations • Named responsible person or role for clinical use, output oversight and incident reporting • Deployer obligations under applicable AI law, where one exists (e.g. EU AI Act)	• Review of the contract and the allocation of responsibilities against the applicable device and AI law	• Liability for failure allocated in the contract • A named responsible party for each of clinical use, output oversight, incident reporting and updates • Applicable deployer obligations met	• Unassigned responsibilities (n) • Unmet legal obligations (n)

4 PERIPHERAL DOMAINS

ETHICAL ASPECTS AND ALGORITHMIC JUSTICE

☐ PERIPHERAL DOMAIN

Fair distribution of benefit and harm across populations and the terms on which the technology acts on patients (whether users keep agency over its use).

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Distribution of benefit and harm	• Clinical outcome results by subgroup, from published studies or local data • Local population profile by the same subgroups (e.g. census or EHR data)	• Subgroup comparison of benefit and harm against the same pathway without it	• No subgroup disadvantaged beyond Δ~e~ Δ~e~: the largest disadvantage in benefit or harm for any subgroup, relative to the overall population, that the decision tolerates, set and justified at Stage 2.	• Benefit and harm difference by subgroup, 95% CI • Between-group gap (pp)
Equitable access and inclusion (where relevant)	• Eligible population by subgroup, from registry, EHR or census data • Uptake by the same subgroups, from usage logs or service records • Supported languages, accessibility statement (WCAG) and device and connectivity requirements, from the IFU or vendor documentation	Comparison of uptake by subgroup with each subgroup's share of the eligible population, against the same pathway without it	No subgroup's access reduced beyond Δ~a~ Δ~a~: the largest reduction in access for any subgroup, relative to the same pathway without the technology, that the decision tolerates, set and justified at Stage 2.	• Uptake of eligible patients by subgroup (%) • Gap against the population share (pp)
		Review of language, accessibility and device or connectivity requirements against the local population	No eligible subgroup excluded by design	• Subgroups excluded by design (n)
Patient autonomy and informed use	• Patient information on the use of AI (leaflet, notice or in-app text) and opt-out procedure, where offered • Patient survey or user-testing results on understanding of the AI's role	• Review of patient information; comprehension testing with patients	• Patients informed of AI use before or at the point of care • Comprehension at or above the target set at Stage 2	• Patients informed of AI use (%) • Patients correctly understanding the AI's role and their options (%)
	• Procedure for human review of AI-informed decisions, and explanation provided to patients • Complaints and appeal records; sample of contested decisions	• Audit of the review procedure and of a sample of contested decisions	• Human review available for every AI-informed decision with significant effects on the patient • Review completed within the time limit set at Stage 2	• Decisions with human review available (%) • Requests resolved (%); median review time

DATA GOVERNANCE, CYBERSECURITY, AND INTEROPERABILITY

☐ PERIPHERAL DOMAIN

Whether the technology's inputs are accurate, representative and traceable, resist compromise, and connect to the systems already in use.

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Operational data quality and provenance	• Vendor input specification: required fields, formats and handling of missing values • Sample of local input records, with the source system for each field	Comparison of a sample of local input records with the input specification	• Required fields complete and valid within the limits set at Stage 2 • Source system identified for every safety-critical input	• Complete and valid required fields (%) • Safety-critical inputs with identified source (%)
Cybersecurity and operational resilience	• Security certification or report (e.g. ISO/IEC 27001 certificate, SOC 2 report) and penetration-test summary • Software bill of materials, where available, and manufacturer security advisories • Local access-control and security-incident records	Review of vendor security evidence and local records against the deploying organisation's security requirements	• No open critical vulnerability (CVSS ≥ 9.0) affecting the deployed version • High-severity vulnerabilities remediated within the deadline set at Stage 2 or in the contract • Deploying organisation's security requirements met	• Open critical and high vulnerabilities (n) • Median time to patch (days) • Unauthorised access incidents (n)
	• Service-level agreement: availability, recovery time objective (RTO) and recovery point objective (RPO) • Outage history, from the vendor status page or incident log • Backup and restoration-test records	Comparison of observed availability and tested recovery with the service-level agreement and the clinical availability requirement set at Stage 2	• Observed availability at or above the Stage 2 requirement • Tested recovery within the RTO and RPO	• Availability (%) • Recovery time and data-loss window in the last test (min) • Unplanned outages (n)
Interoperability and data-exchange integrity	• Interface specification, with DICOM conformance statement or HL7 FHIR capability statement as applicable. • Per test case: value of each safety-critical field in the source system, as received by the AIHT, and as written back to the patient record • Per test case: patient identifier in the source system, the AIHT and the destination record • Per message: transfer status (sent, received, failed), from the interface log	End-to-end integration testing on a sample of cases: input from the source system to the AIHT, and output back to the patient record	• All safety-critical fields transferred correctly in tested cases • No patient-matching error	• Safety-critical fields transferred correctly (%) • Patient-matching errors (n) • Failed transfers (%)

ORGANISATIONAL AND IMPLEMENTATION ASPECTS

☐ PERIPHERAL DOMAIN

How the technology alters workforce roles, competencies and training needs, workload and workflow within the adopting organisation

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Care delivery processes	• Current and AI-enabled care pathway maps • Process steps, hand-offs and decision points under each pathway • Process and care-pathway time per patient under each pathway • Service throughput per unit of time under each pathway • Workarounds, duplications and delays observed at pilot or comparable sites • Staff perceived usability scores • Usability and human-AI interaction issues affecting care delivery • Referral, escalation or follow-up processes within the assessed care pathway	• Workflow and care-pathway mapping • Time-and-motion analysis comparing task duration with and without the AI feature • Direct observation • Human-factors assessment of human-AI interaction, task allocation, usability, and points in the care process where the technology may alter the risk of error • Process evaluation of changes in routine care delivery, including workflow steps, hand-offs, delays, workarounds, coordination, and service throughput • Qualitative interviews and focus groups • Comparative analysis against the same care pathway without the AI feature	Each selected care-delivery indicator should be assessed against the same care pathway without the AI feature, considering the direction and magnitude of change.	• Change in care-pathway time • Change in number of process steps or hand-offs • Change in delay or duplication rates • Change in workaround or process-deviation rates • Change in workflow-related error or near-miss rates • Change in local service throughput or capacity • Change in referral, escalation or follow-up patterns within the assessed care pathway
Workforce roles and competencies	• Staff time per patient, by professional group, under each pathway • Staff time for output review, oversight and local monitoring of the technology, by professional group • Tasks performed and decisions held, by professional group, and by the technology under each pathway • Training hours required per user, initial and refresher • Competency and AI-literacy scores before and after training • Override and escalation decisions in testing or simulation, and their appropriateness • Ratings of non-technical competencies in simulation (communication, teamwork, shared decision-making) • Workload and cognitive-burden scores under each pathway • Observed or reported instances of over-reliance or deskilling	• Workforce and task analysis • Pre-post competency and AI-literacy assessment • Training evaluation • Simulation or scenario-based assessment • Staff surveys and interviews • Human-factors assessment of workload, reliance on AI, override and escalation behaviour, and changes in professional judgement • Comparative analysis against the same pathway without the AI feature	Each selected workforce indicator should be assessed against the same pathway without the AI feature, considering the direction and magnitude of change. *Safety-critical competency gaps or unclear professional accountability are not acceptable.*	• Change in staff time or FTE requirements • Change in workload or cognitive burden • Change in training hours required per user • AI-literacy or competency assessment score, and proportion of users meeting a defined competency standard where applicable • Documented change in task allocation, decision authority or professional accountability • Change in the frequency with which users override AI recommendations or escalate cases for additional human review • Change in unaided task performance over time

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Effects on the wider health system	• Referral volume, movement of patients between services or levels of care • Distribution of workload across services or organisational units • Activity in adjacent services under each pathway (including additional or avoided consultations, tests or follow-up) • Service demand and capacity across the health system • Service coverage by region and by rural and urban setting • System-level technical support and maintenance requirements attributable to the technology	• Comparative pathway and system mapping • Health-service utilisation analysis • Resource and capacity analysis • Referral and downstream-care analysis • Comparative assessment of changes in system-level technical support and maintenance requirements attributable to the technology • Stakeholder interviews and organisational case-study analysis • Comparative analysis against the same pathway without the AI feature	Each selected health-system indicator should be assessed against the same pathway without the AI feature, considering the direction and magnitude of change.	• Change in referrals between services or levels of care • Change in downstream consultations, tests or procedures • Change in service demand or capacity • Change in workload across services or organisational units • Change in coordination or continuity-of-care indicators • Change in system-level technical support or maintenance requirements • Change in service availability or coverage

ENVIRONMENTAL AND SOCIETAL EFFECTS

☐ PERIPHERAL DOMAIN

Computational and material footprint of the technology and its wider societal consequences.

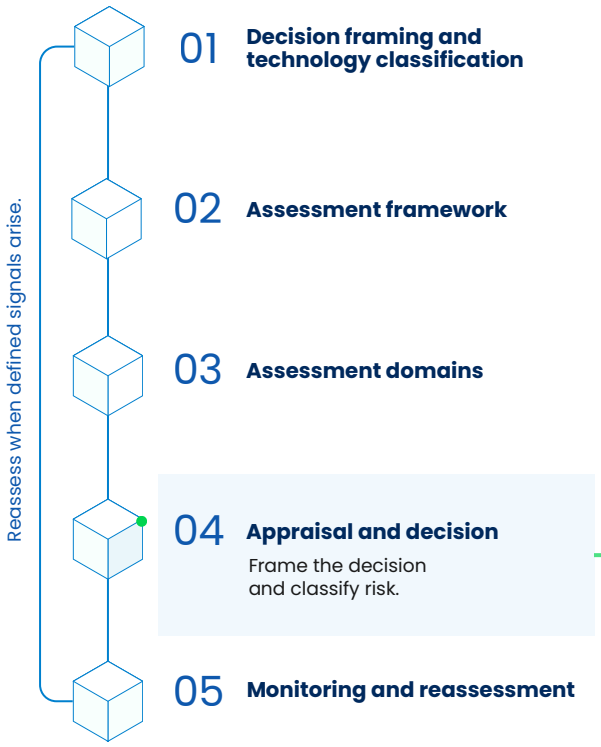
Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Direct impacts Computational and material footprint of the technology normalised to a clinical functional unit (per patient / per averted health outcome)	Attributional LCA • Life Cycle Inventory (LCI) data (bill of materials, energy, water, emissions) • Environmental Product Declarations (EPDs) from hardware manufacturers • Recognised LCA databases (e.g. ecoinvent) for generic ICT/electronics emission factors • Attribution rule + materiality threshold Dynamic/living LCA: • Preferred primary inputs: direct measurement of the energy use, water use and emissions via dedicated tools (e.g. CodeCarbon) Repeated/updated across the lifecycle (re-training, grid decarbonisation) • Real-world use data in clinical setting (volumes/patients/cases) • Share of shared data-centre load attributed to the device (attribution rule + threshold)	Life Cycle Assessment (LCA) - ISO 14040/14044, cradle-to-grave. Specifically: • Attributional LCA (material footprint: production, transport and end-of-life) • Dynamic/living LCA (operations: initial training; inference; re-training) Assessment is comparative to clinical alternative (BAU) per functional unit	Pass-fail based on net environmental footprint compared to BAU, weighted against the health benefit (per patient/per health outcome) No established threshold available	All metrics normalised per functional unit (patient/health outcome) Attributional LCA component: • Output measures of impact: Emissions: kg CO_2eq (GWP) • Input measures with impact: Energy use: electricity (kWh) Water use: litres Land: m²; (+ critical minerals kg/abiotic depletion) Dynamic/living LCA component: • Output measures of impact: Emissions: kg CO_2eq (operational); carbon intensity per query/task • Input measures with impact: Resource use: electricity (kWh; PUE/CUE); Water: litres; WUE, cooling; Land: m² (with data-centre footprint share) Output measures capture emissions released to the environment and input measures the resources withdrawn from it.

Environmental and societal effects (continued)

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Indirect impacts Wider societal consequences	• Care-pathway data: clinical volumes, activity displaced or added (travel avoided - also considering changes in staff composition, tests/admissions avoided or induced) • Behavioural and system-level change (largely qualitative evidence) • Scenario assumptions and uncertainty declared	Consequential LCA/Scenario analysis (cradle-to-grave) • Enabling effects modelled as avoided/induced impact; systemic effects through qualitative scenarios + sensitivity analysis • Contextualised geographically and over time (energy mix)	Compensatory: Net effect (avoided minus induced impact) expressed on a common scale and traded off against other domains (Stage 4) aggregate judgement, against opportunity cost. Flagging of potential behavioural and system-level changes that may reverse the net balance (information gathered qualitatively)	All quantitative metrics normalised per functional unit (per patient/per averted health outcome). Enabling effects: Emissions: kg CO_2eq (avoided/induced) Resource use: electricity, water, land (avoided/induced) Systemic: Qualitative appraisal

2.4

Stage 4 – Appraisal and decision



Possible decisions after appraisal

ROUTINE ADOPTION
Approve for routine use.

CONDITIONAL USE

- **Restricted use**
  Benefit is limited to a defined subgroup.
- **Provisional funding**
  Expected value is positive but more evidence is needed.
- **Supervised pilot**
  Local performance or implementation is uncertain.
- **Research only**
  Clinical utility has yet to be established.

REFUSAL OR WITHDRAWAL
Decline adoption if expected value is negative or risk is unmanageable. Withdraw an existing technology if its value no longer supports continued use.

Tool 5 Managed Access Agreement — For conditional use

Assessment establishes what the technology achieves, while appraisal determines the importance of those findings for the decision.

2.4.1 Appraisal

Appraisal judges the value of the evidence assembled at Stage 3. It is distinct from assessment, which establishes the evidence and its uncertainty. An economic evaluation, for example, estimates the additional costs and effects of a technology relative to its comparator. It does not, by itself, determine whether the resulting value is acceptable. That judgement depends on the decision-maker's objectives and the relevant cost-effectiveness threshold[50]. The same distinction applies across domains: assessment establishes what the technology achieves, while appraisal determines the importance of those findings for the decision.

In many health systems, appraisal is undertaken by a standing multidisciplinary committee that considers clinical benefit alongside the use of limited healthcare resources[45]. Its judgement may take into account the quality and magnitude of benefits and harms, the technology's place in care, disease severity, unmet need, health inequalities and the cost of achieving the expected benefit[45]. The committee should publish its reasons, including the evidence it relied on and any important remaining uncertainty, so that the decision can be scrutinised[45].

Opportunity cost is central to economic appraisal when healthcare budgets are constrained. Funding one technology can displace other care. A cost-effectiveness threshold provides an estimate of the health likely to be forgone per unit of additional expenditure when such displacement occurs[45],[50]. As a technology's incremental cost-effectiveness ratio approaches or exceeds the relevant threshold, other considerations may carry greater weight. The case for adoption may also require stronger evidence as its budget impact increases[45]. Where important uncertainty remains, the committee can recommend further evidence generation as a condition of adoption.

Appraisal methods for DHTs and AIHTs are still evolving across different countries. Published economic evaluations often adapt methods used for medicines and fixed-function devices, use short time horizons and report the analytical perspective inconsistently[53]. Available frameworks do not always specify a perspective or time horizon, and may give limited attention to distributional effects, lifecycle change or transferability[52]. In practice, assessments may cover fewer domains than the domains listed in Stage 3, sometimes omitting economic value altogether. Yet the value of AI-enabled technologies also depends on technical, clinical, human, professional and legal considerations that cannot be reduced to accuracy or cost alone[49].

COUNTRY EXAMPLE

TWO ROUTES TO EARLY ACCESS

Germany DiGA

- Provisional listing
- Generate comparative evidence

France PECAN

- One year of early coverage
- Demonstrate beneift for reimbursement

Some countries have established dedicated assessment and coverage pathways for digital technologies.

- Germany's DiGA procedure requires evidence of positive care effects, defined as medical benefit or patient-relevant improvements in the organisation and delivery of care. It also examines safety, data protection, information security, interoperability and usability[54]. Provisional listing is available for up to 12 months, extendable to 24, to allow comparative evidence to be generated.
- France's PECAN pathway similarly offers one year of early coverage on the basis of expected clinical or organisational benefit. The CNEDiMTS committee of the Haute Autorité de Santé then assesses demonstrated benefit and organisational impact for standard reimbursement[54].

Both pathways consider effectiveness and selected additional domains rather than the full matrix proposed here. In Germany, economic value is not part of the initial listing decision and enters at the subsequent price negotiation[54].

2.4.2 Decision

Another important component of Stage 4 is the conversion of appraisal findings into a decision or recommendation. Possible outcomes include routine adoption, restricted use, provisional funding, a supervised pilot, use only in research, refusal or withdrawal. The choice depends on the extent of remaining uncertainty, whether further evidence can reduce it, and the cost and reversibility of acting before that uncertainty is resolved[36],[37].

A recommendation has practical effect only when its conditions are explicit.

- ☐ The eligible population
- ☐ The model version covered by the decision
- ☐ The human oversight required
- ☐ The competence users need for safe operation
- ☐ The implementation conditions needed to realise the assessed value

Conditional decisions provide alternatives to either full adoption or refusal. If benefit is limited to a defined subgroup, the recommendation should restrict use to that population[45]. Where expected value is positive but evidence remains insufficient for routine reimbursement, provisional funding may support controlled use alongside a specified evidence-generation programme, provided the decision can be reversed[36],[37]. Uncertainty mainly about local performance or implementation may justify a supervised pilot[37]. Where clinical utility has not been established, use may need to remain within research[45]. Negative expected value or an unmanageable risk supports refusal[45]. A funded technology may later be withdrawn if new evidence, a changed price or a better comparator undermines its value[36],[50]. Withdrawal following post-deployment review is addressed in Stage 5.

A recommendation has practical effect only when its conditions are explicit. It should specify:

i. the eligible population[45],[50]
ii. the model version covered by the decision[19],[55]
iii. the human oversight required[19],[55]
iv. the competence users need for safe operation[45]
v. the implementation conditions needed to realise the assessed value[45].

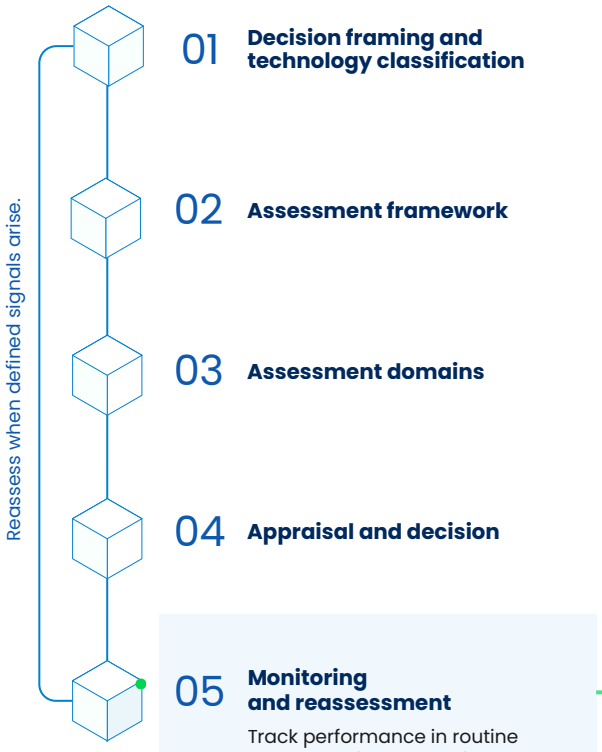
Before routine use, the deploying organisation must demonstrate that it can operate, monitor and maintain the technology safely, using the AI Readiness Self-Assessment (Tool 4, §4.4), which rates its AI governance and infrastructure across seven functional areas on a five-level scale. The recommendation should also specify the indicators to monitor, thresholds for review, relevant commercial or risk-sharing conditions, the reassessment date and any changes that would reopen the decision. The rationale, material uncertainties and any dissenting views within the committee should be made public[45]. Each recommendation is therefore valid under stated conditions and remains open to revision, consistent with Principle P5 in Section 1.3.2.



2.5

Stage 5 – Monitoring and Reassessment

Monitoring is the ongoing observation of a technology in routine use; reassessment is the formal review of its value using the evidence collected[56],[57]. These activities are especially important for DHTs and AIHTs because their performance may change after adoption[56],[57]. Populations, input data and clinical practice evolve, and model accuracy may decline without an obvious failure. The version reviewed at Stage 4 may also differ from the one subsequently deployed. Stage 5 implements the monitoring arrangements to detect changes and determines whether the original decision should stand.



From monitoring to reassessment

MONITOR IN ROUTINE USE

Track the indicators agreed at appraisal against the accepted baseline. Check results by site and subgroup.

TRIGGER A REVIEW

Reassess when a monitoring threshold is breached or a scheduled review date is reached.

REASSESS AND DECIDE

Use a targeted review for a specific change or signal. Reassess fully when the intended use, system design, comparator, or overall value has materially changed.

2.5.1 Monitoring

Monitoring provides continuous or periodic checks between more resource-intensive reassessments[56],[57]. It typically follows a focused set of indicators that reflect the value accepted at appraisal. The baseline includes the level of performance accepted and its expected range of routine variation, sometimes termed common-cause variation[57]. Material departures from that baseline will trigger a reassessment outlined under Section 2.5.2. Monitoring should match the level of evidence on which adoption was justified[58]. A technology accepted on technical performance may be monitored primarily against that measure, whereas one adopted on patient outcomes should be followed using those outcomes or validated surrogates.

Post-deployment indicators should track the factors that influence outcomes assessed in Stage 3, particularly where those outcomes cannot be measured frequently. An appraisal metric may also be used directly when it can be recalculated reliably at reasonable cost. This is especially important for technical performance, which can decline as input distributions, labels or underlying relationships change, even while model outputs appear confident[57],[59]. Monitoring input data, calibration and errors can identify deterioration before it becomes visible in clinical outcomes, leaving time for corrective action[57],[59]. Results should be examined by relevant subgroups and sites so that favourable overall averages do not hide local failures[56],[57],[59].

Manufacturers of high-risk systems are required to undertake post-market surveillance and collect lifecycle performance information where the applicable regulations require it[39]. Assessors can draw on these technical data, including deployed model versions, notified changes, data drift and error signals. They also need evidence from local clinical use: uptake among eligible patients, adherence to intended use, human-AI interaction and safety incidents. Sources include registries, electronic health records, incident reports and structured patient- and clinician-reported measures[39],[56],[57]. Lifecycle approaches assign responsibility for initiating reassessment to the appraisal body and work with providers rather than relying solely on the manufacturer to identify a loss of value[57].

Repeated trials are rarely feasible for every software update. Real-world evidence therefore has a central role in monitoring and reassessment. The German DiGA pathway offers one example: provisional listing requires the manufacturer to provide evidence of positive care effects within a defined period before longer-term reimbursement is confirmed or refused[54],[60]. New evidence should be judged against the standards used at Stage 3 and the value established at Stage 4. Reassessment should not lower the original evidence requirements.

Figure 4 proposes a minimum monitoring dashboard organised by Stage 3 domain. Each indicator is linked to the aspect of value that a change might affect. Conditional decisions should specify both the timetable for gathering additional evidence and the date for reviewing the recommendation[45],[50],[54],[55]. Monitoring frequency should reflect how quickly each indicator may change[59]. The table provides an indicative schedule.

Figure 4:
Minimum monitoring dashboard

ASSESSMENT DOMAIN	MONITORED PARAMETERS	HELD BY	SOURCE	PROPOSED FREQUENCY
Technical performance and AI-specific characteristics	Version in service and notified changes; input data quality and drift; calibration and error rate; availability and latency	Manufacturer	Post-market monitoring reports, model logs	Continuous or automated, as reported
Safety	Adverse events and near-misses attributable to the tool; incidents; complaints; corrective actions; recalls	Manufacturer and provider	Vigilance and incident reporting	Event-driven, as they arise
Clinical and service effectiveness	Uptake and eligible-population reach; adherence and abandonment; effect on the decision and on downstream care; endpoint on an audited sample	Provider	Electronic records, service audit	Quarterly
Economic value	Realised uptake and unit price; implementation and running cost; resource use; realised budget impact against forecast	Provider and payer	Finance, procurement and activity data	Annual, or on a price or comparator change
Ethical aspects and algorithmic justice	Performance and access by relevant subgroup and site; disparity in error or calibration across groups	Provider and manufacturer	Stratified, anonymised performance data	Annual, or on any retrain
Legal and regulatory aspects	Authorisation and conformity status (CE mark, AI Act classification); change in applicable law; liability events	Manufacturer	Regulatory notices, manufacturer notifications	Event-driven
Data governance, cybersecurity and interoperability	Security incidents and data breaches; access-control and consent failures; interoperability breaks with connected systems	Manufacturer and provider	Incident reporting, IT security logs	Continuous for feed integrity; event-driven for incidents
Organisational and implementation aspects	Workload and cognitive burden; automation bias and appropriate reliance; override rate; training currency; fidelity to intended use; support requests;	Provider	User surveys (PROM, ClinRO, usability), workflow logs	Semi-annual
Environmental and societal effects	Realised compute, energy, water, and material consumption; local externalities of operation against the appraised footprint	Provider and manufacturer	Utilisation and facility data, operator reporting	Annual, or on a change in deployment scale

2.5.2 Reassessment

Reassessment revisits the Stage 4 judgement when a monitoring threshold is breached or a scheduled review date is reached. It considers the technology as deployed in its care pathway, using current evidence against the conditions of the original recommendation[45],[57]. The scope of review should reflect what has changed. A targeted reassessment may examine a new model version, a subgroup, a price change or a specific safety signal. A full reassessment is needed when the intended use, system design, comparator or overall value proposition has changed materially[40],[41]. This approach focuses assessment capacity on changes likely to affect value.

The same decision options apply at reassessment as at initial appraisal. This maintains consistency between the conditions for starting and continuing use. A reassessment can retain or revise a conditional recommendation, narrow the eligible population, introduce new safeguards or return a technology to provisional funding. It need not choose only between continuation and withdrawal[36],[37]. Figure 5 summarises the possible outcomes and the signals that may prompt them.

Disinvestment is an important possible outcome of reassessment because it releases resources for more valuable care[40],[41],[50]. A funded technology should be withdrawn when its benefits no longer justify its costs, for example because new evidence reduces its expected value, a better comparator becomes available or safe operation cannot be maintained[36],[37],[45]. The original funding decision carried an opportunity cost: resources used for the technology were unavailable for other care[50]. Withdrawal can reverse that allocation and make those resources available for higher-value uses. In this way, reassessment completes the economic logic of the HTA lifecycle[50],[60].

Figure 5:
Reassessment outcomes

OUTCOME	MEANING	INDICATIVE TRIGGER
Continue	Use continues on existing terms	Monitoring within the terms the decision fixed
Expand	Extend to further sites, populations, or functions, or reduce oversight	Sustained positive signal, original uncertainty resolved
Modify	Change workflow, threshold, training, interface, monitoring, or contract	Process or human-factor signal
Restrict	Narrow the eligible population or the permitted use, including confinement to a supervised pilot pending local evidence	Subgroup, safety, equity, or implementation signal
Retrain or recalibrate	Restore performance under controlled change and revalidation	Technical drift or performance decline
Suspend	Pause use while an issue/incident is investigated	Unresolved safety or security signal
Retire or disinvest	Withdraw the technology from funded use	Value inadequate, technology superseded, or safe support unsustainable



3

Assessment in Resource-Constrained and Emerging-HTA Settings



3.1

Rationale and stance

3.1 Rationale and stance

Chapter 2 sets out the full five-stage assessment process, with evidence requirements proportionate to the technology's risks and the decision. This chapter explains how to apply that process when HTA capacity, resources or data are limited. Adaptive HTA offers approaches that require less time and data while remaining suitable for the decision[12],[61],[62]. Assessment can be made more efficient by focusing on the most relevant subdomains, simplifying analytical methods, adjusting the order of work, using external evidence or sharing tasks across national bodies. These adaptations reduce the burden of assessment but do not change the questions that must be answered or the standards used to judge whether evidence is sufficient.

Two principles guide this approach. First, national income is not a reliable measure of HTA capacity. Gross National Income (GNI) and Gross Domestic Product (GDP)[63] do not capture the institutional arrangements, workforce, data systems and financing that determine whether assessment can be undertaken[12],[64]. Capacity varies substantially within income groups: some upper-middle-income countries have advanced HTA systems, while some high-income systems remain fragmented[61],[63],[64]. Section 3.6 therefore considers institutional maturity directly rather than using income as a proxy. Second, economic assessment remains essential even when methods are simplified. Opportunity cost links adoption to the health benefits displaced elsewhere in the system[12],[62]; technical performance alone cannot establish whether a technology makes good use of limited resources. Cost-effectiveness and affordability must still be considered[51]. When data are scarce, a budget impact estimate or cost-consequence comparison may be more practical than a full cost-utility model, provided the consequences for other healthcare spending remain clear.

The readiness of the deploying organisation is a separate factor in whether expected value can be realised[65]. Benefits demonstrated in a well-resourced facility may not transfer to a setting without reliable power, connectivity, qualified staff or data governance. Any infrastructure needed to support adoption also increases the technology's total cost of ownership (TCO)[61],[62],[65]. Readiness should therefore be assessed directly, alongside HTA capacity, before the assessment approach and implementation conditions are decided.

Two principles for adapting HTA

1 Assess HTA capacity directly
National income is a poor proxy for HTA capacity. Look at the institutions, workforce, data systems and financing that support assessment.

2 Keep economic assessment essential
Simpler methods can still address cost-effectiveness, affordability and what other care might be displaced.

Prioritisation determines which technologies enter assessment; proportionality determines how thoroughly those selected are assessed.

The central principle of this chapter is that assessment should reflect available resources, HTA maturity and conditions at the point of care. These factors determine how the assessment is carried out and what form of adoption may be supported, regardless of the technology class. A minimum assessment core must nevertheless be retained.

Limited capacity also requires prioritisation before an assessment begins. An HTA unit that cannot evaluate every candidate technology must decide which to assess first. Figure 6 sets out criteria for making that choice. Prioritisation thus determines which technologies enter assessment, whereas the adaptations described above determine how thoroughly those selected are assessed.

Figure 6:
Assessment prioritisation criteria under capacity constraints

CRITERION	CONSIDERATIONS TO BE PRIORITISED
Health need	Disease burden, severity, unmet need, and gaps in access
Potential value	Expected gain in clinical outcome, patient experience, system capacity, efficiency, and equity
Risk	Consequence of error, degree of autonomy, exposure of vulnerable users, and reversibility
Budget and opportunity cost	Acquisition, implementation, and downstream cost, and the care displaced by adoption
Evidence uncertainty	Maturity and generalisability of the evidence, and stability of the model version
System readiness	Data, infrastructure, workforce, interoperability, and governance at the deploying site
Strategic relevance	Alignment with national digital-health, AI, public-health, or service-transformation priorities
Availability of external work	Existing regulatory assessments, HTA reports, standards, and monitoring data open to reliance

3.2

Transferability of external assessments

3.2 Transferability of external assessments

Health systems with limited HTA capacity need not repeat work already completed by a qualified authority elsewhere[12],[66]. They should, however, establish whether that work is reliable, relevant to their decision and applicable locally. This section explains how to make that judgement.

Approaches to using external assessments differ in how much work is accepted and how much judgement remains with the receiving authority[12]. These approaches draw on practices first developed in regulation and now applied to HTA.

Figure 7:
Approaches to the use of external assessments

APPROACH	DESCRIPTION	CONDITION FOR USE
Recognition (direct adoption)	Accepts another authority's decision with little further review[12],[66]	Formal legal basis and a finding that the reference authority applies equivalent standards and quality controls[67]
Reliance	Gives substantial weight to a trusted external assessment, while the receiving authority remains accountable for its own decision	More flexible than recognition, but justified only when the external assessment addresses a sufficiently similar decision and meets defined eligibility criteria[66]
Adaptation	Adjusts external findings to local populations, comparators, care pathways, prices, values or legal requirements[67]	-
Joint assessment	Several authorities produce a common report[67]	-
Work sharing	Authorities divide the work and combine their contributions[67]	-
Reference assessment	Provides a foundation for further analysis without formal reliance status[67]	-

Figure 8 presents a seven-part reliance eligibility test covering the technology and decision question, the evidence and its quality, and the currency and local applicability of the findings. All seven criteria must be met and a material failure in any one will rule out reliance. Passing the eligibility test does not mean that every conclusion can be adopted unchanged. Transferability must be judged separately for each component. The technology description and post-market signals may transfer relatively easily where the version, configuration and observed behaviour are comparable. Evidence on model performance, clinical effects and safety needs closer examination against local populations, disease prevalence, decision thresholds, human factors and workflows. Economic conclusions are usually less transferable because cost-effectiveness, budget impact and affordability depend on local prices, eligible populations, capacity, opportunity cost and available funding.

Figure 8:
Reliance eligibility test

EQUIVALANCE *Parts 1-3*	1	**Object equivalence** Same model and version; identical inputs, outputs, threshold and update status.
	2	**Intended-use equivalence** Same purpose, population, user, setting and autonomy.
	3	**Decision equivalence** Same or closely matching coverage, procurement or adoption question.
EVIDENCE *Parts 4-5*	4	**Population and pathway** Comparable epidemiology, workflow, workforce, equipment, downstream care.
	5	**Methodological acceptability** Transparent methods; evidence tables, bias assessment, uncertainty, stakeholder process.
CURRENCY & LOCAL FIT *Parts 6-7*	6	**Currency** No material change in evidence, comparator, model, safety or price since production.
	7	**Contextual transferability** Local affordability, implementation, equity, legal, data and infrastructure can be met.

Organisational, legal and data-governance findings also depend strongly on the receiving setting. Ethical and equity evidence may be partly transferable: subgroup results can inform other settings, but judgements about fairness depend on local access to care, stakeholder priorities and the distribution of benefits and harms. Figure 9 summarises the expected transferability of each assessment component and the local work still needed. The purpose of reliance is therefore to reuse what remains applicable and focus scarce assessment resources on what must be evaluated locally.

Figure 9:
Transferability of assessment components

ASSESSMENT COMPONENT	EXPECTED TRANSFERABILITY	CONSIDERATIONS
Technology description	High	Confirm the deployed version and set-up match the assessed technology
Model performance	Moderate to high	Revalidate accuracy on local data, case-mix, equipment, and decision threshold
Clinical evidence	Moderate	Re-align to the local comparator and pathway; check subgroup applicability
Safety evidence	Moderate to high	Reassess local human factors, workflow fit, cybersecurity, and fallback
Economic evaluation	Low to moderate	Re-cost using local prices, the opportunity-cost threshold, and perspective
Budget impact and affordability	Low	Recompute for the local eligible population, uptake, price, and fiscal space
Organisation and implementation	Low	Consider local workflow, workforce, infrastructure, and adoption readiness
Ethics and equity	Partial	Consider distributional impact, access, and acceptability for affected groups
Law and data governance	Low	Confirm compliance with local law, data protection, and data residency
Post-market signals	High	Formalise local monitoring, reporting, and pre-set corrective triggers

3.3

Repurposing evidence across decision systems



HTA can use regulatory findings as a starting point and concentrate on questions that remain unresolved: relative effectiveness against local care, cost-effectiveness, budget impact, organisational effects and equity.

An HTA unit can draw on more than assessments undertaken in other jurisdictions. Regulatory reviews may provide a second source of evidence, often from within the same health system[13],[68]. Where national regulatory capacity is limited, the relevant dossier may instead come from a trusted authority abroad[68]. Using this evidence can reduce duplication, provided the unit identifies which questions the regulator has already answered and which remain for HTA.

As already mentioned, regulation and HTA serve different purposes. Regulators assess whether a product meets the safety, performance and quality requirements for market access[13]. HTA considers its comparative benefit and value within the local health system[12]. The evidence bases overlap in areas such as clinical trials and safety, but may differ in comparators, populations and outcomes; patient-reported measures and quality of life, for example, are less commonly primary regulatory endpoints[13],[68],[69]. HTA can use regulatory findings as a starting point and concentrate on questions that remain unresolved: relative effectiveness against local care, cost-effectiveness, budget impact, organisational effects and equity[12],[13],[68].

The receiving unit should decide which regulatory findings it can reuse, component by component, rather than treating the dossier as wholly transferable. Figure 10 maps regulatory outputs to evidence that can inform HTA and to questions requiring local analysis. The distinction is consistent throughout: regulatory evidence establishes important characteristics of the product, while HTA evaluates its comparative consequences and value for the decision at hand.

Figure 10:
Regulatory evidence and residual requirements for local HTA

REGULATORY OUTPUT	REUSED IN HTA	LOCAL EVIDENTIARY REQUIREMENTS
Quality and manufacturing	Accepted in full, not reopened	None
Safety and benefit-risk	The regulator's safety assessment, conducted more thoroughly than an assessment would conduct	Pharmacovigilance and post-market signals in the treated population
Technical and clinical validation (device and AI)	Measured performance, data representativeness and change-control status	Re-check of performance on local data, case-mix, and decision threshold
Efficacy	Trial evidence of efficacy under experimental conditions	Re-assessment as relative effectiveness against the local comparator
Comparator	Little of use, the regulatory comparator mostly being placebo or none	The local active comparator, with an indirect comparison if needed
Endpoints	Trial endpoints as accepted, frequently surrogate or intermediate	Patient-relevant final outcomes and quality-of-life or patient-reported measures
Study population	The trial population	Adjustment to local epidemiology, subgroups, and care pathway
Cost-effectiveness, budget impact, affordability	Nothing, not assessed by the regulator	Produced in full at local prices, opportunity cost, and fiscal space
Organisational, ethical, and equity aspects	Nothing, not assessed by the regulator	Produced in full for the local system
Time horizon	Trial duration only	Extension to the decision-relevant lifetime, by modelling

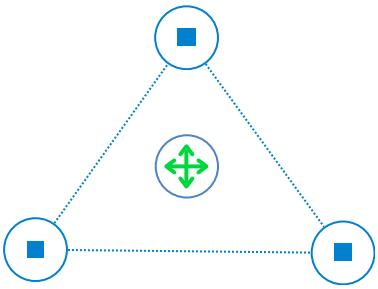
For DHTs and AIHTs, regulators may already have assessed technical and clinical validation, the representativeness of training and test data, human-AI interaction and the arrangements for managing changes over time[43],[55],[56],[70]. These findings provide a useful technical foundation for HTA[30],[54],[68],[71]. Their contribution has limits, however. Regulatory evidence for these products may be less extensive than for medicines[30], and often consists of single-arm or performance studies without an active comparator[70]. The European medical-device framework does not provide a dedicated real-world evidence framework[72], leaving important questions about comparative value unanswered. Short product lifecycles and frequent modifications also mean that regulatory findings may apply to only one model version[70],[71].

3.4

Regional and cross-national collaboration

Collaboration allows HTA bodies to share work while assessments are being planned and produced, rather than relying only on completed reports[14],[40],[67],[73]. In emerging and resource-constrained systems, regional and international networks have helped build skills and distribute assessment costs through joint training, methodological guidance and evidence sharing[61],[63],[74],[75].

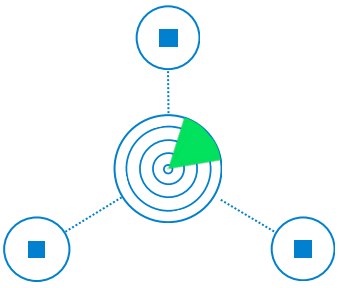
Collaboration ranges from exchanging guidance and tools[14],[74],[75], through working in networks[73],[75],[76], to producing assessments jointly across jurisdictions[14],[73]. Three common forms differ in the resources they require and the work they can save:



Exchange of experience and convergence of method

Exchange of experience and convergence of method

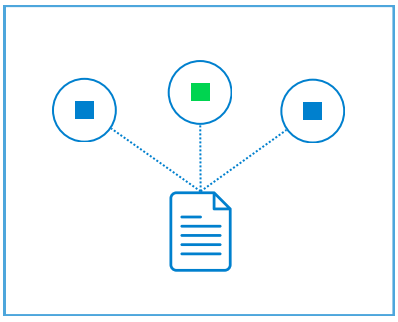
DEGREE OF INTEGRATION
Low



Joint horizon scanning

Exchange of experience and convergence of method

DEGREE OF INTEGRATION
Moderate



Joint assessment

Exchange of experience and convergence of method

DEGREE OF INTEGRATION
High

More integrated forms of collaboration require greater methodological alignment and trust among participating authorities.

Other initiatives include joint price negotiation after assessment, combining purchasing power[73]. More integrated forms of collaboration require greater methodological alignment and trust among participating authorities. They are therefore often developed as institutional capacity grows[61],[67].

In the Asia-Pacific region and the Americas, HTAsiaLink and the Pan American Health Organization's RedETSA mainly support methodological exchange and capacity building, including among low- and middle-income countries. HTAsiaLink promotes joint research and shares assessment experience, while RedETSA maintains a regional database of HTA reports to make existing work easier to find and reuse[74],[75],[76]. The Beneluxa Initiative, involving Belgium, the Netherlands, Luxembourg, Austria and Ireland, operates at a more integrated level. During its first decade, it undertook 15 joint assessments and four joint negotiations, combining horizon scanning, voluntary assessment and price negotiation. Its joint assessments address economic as well as clinical evidence, but each country reaches its own decision under national law[73].

The European Union's joint clinical assessment (JCA) system is the most integrated example and the only binding arrangement described here. It assesses relative clinical effectiveness and safety, while Member States retain responsibility for economic evaluation, appraisal, pricing and reimbursement. A Coordination Group oversees the process. An assessor and co-assessor from two participating countries combine national PICO proposals into a common scope and prepare the report. The developer may check factual accuracy before the Coordination Group endorses the assessment, seeking consensus while recording any differing scientific views. Member States then use the joint report in their own decision processes[14], [77],[78].

Collaboration can also take place between providers. Hospital-based HTA (HB-HTA) units may jointly undertake rapid assessments to inform procurement and local adoption. In Finland, the Managed Uptake of Medical Methods (MUMM) programme brought together 20 hospital districts and the public agency FINOHTA from 2006. The collaboration produced more than 48 reports, mainly on medical devices and procedures, covering effectiveness, safety, organisational effects and running costs. A typical assessment involved two hospital clinicians and three FINOHTA staff, including HTA and information specialists. Hospitals contributed clinical knowledge and made adoption recommendations; FINOHTA provided methodological and literature-search expertise[79]. The AdHopHTA project subsequently described four organisational models for hospital HTA and developed a shared mini-HTA template, although its reported influence on hospital decisions remained modest[80].

Collaboration is particularly valuable for AIHTs because evidence from one site may not generalise to other populations or healthcare settings. Validation across several systems can provide stronger evidence of local performance[81]. Sharing post-market signals also helps identify where model updates or drift affect performance differently across sites[56],[57],[59]. Data-protection and residency rules may restrict transfers of patient-level records across borders[81]. Federated analysis offers an alternative: a model or analytical query is applied to data held locally, with only aggregate findings shared[27],[34],[57],[81],[82]. Common data models support consistent analysis across sites[25]. This enables multi-country evidence generation without transferring individual patient records[27],[34],[57], while preserving local control over the underlying data[27],[81],[82].

A unit with limited capacity may initially use shared horizon-scanning reports, methods, safety information and external assessments, contributing more actively as its own capabilities develop[61],[67]. Collaboration does not transfer accountability for adoption. Each national or institutional decision-maker must consider the findings in its own context, address remaining local evidence needs and ensure that externally produced analyses can be scrutinised. A shared clinical assessment does not, by itself, establish that a technology can be implemented successfully in every setting.



3.5

Minimum assessment core

The minimum assessment core specifies the evidence that every assessment must consider, regardless of the health system's resources or HTA maturity.

- Technical performance and AI characteristics
- Safety
- Clinical and service effectiveness
- Economic value
- Legal and regulatory aspects

The minimum assessment core specifies the evidence that every assessment must consider, regardless of the health system's resources or HTA maturity. It covers the five core domains introduced in Chapter 2, namely technical performance and AI-specific characteristics; safety; clinical and service effectiveness; economic value; as well as legal and regulatory aspects. Together, they establish what a technology does, what harm it may cause, what effect it has on care, what it costs and whether it can be used lawfully. Without these elements, decision-makers cannot make a sound judgement about value and opportunity cost.

The four peripheral domains are not included in the minimum-core matrix (Figure 11) because their detailed assessment depends on the decision context. Ethical aspects and algorithmic justice require particular attention where large or vulnerable populations are affected[15],[30],[31],[32],[55]; data governance, cybersecurity and interoperability where sensitive information or extensive system integration is involved[30],[32],[55]; organisational and implementation aspects where workflows change materially[24],[49],[70]; and environmental and societal effects where deployment scale or computational intensity is substantial[28],[55]. Where these conditions do not apply, detailed assessment may add little to the decision while consuming scarce resources. The same activation rules apply irrespective of how established an assessment body is: a peripheral domain becomes part of the assessment whenever its relevance is established.

Most resource-saving adaptations should take place within Stage 3, where evidence is generated and analysed. The other stages retain the structure set out in Chapter 2, although the subsequent Section 3.6 also allows proportionate adjustments to decisions and reassessment. Stages 1 and 2 define the decision, intervention, comparator, population and analytical perspective. These tasks generally require less resources than evidence generation but are essential for interpreting findings and reusing external work. Stage 4 converts the evidence into a value judgement and recommendation. Stage 5 monitors the uncertainties accepted at appraisal and enables corrective action if performance or value changes. Removing or weakening these stages would only offer limited savings while undermining the reliability and practical relevance of the decision.

Within Stage 3, proportionality can be applied to three elements: the subdomains examined, the evidence accepted and the analytical methods used. An assessment with limited resources can focus on the subdomains most relevant to the decision, use accessible proxy measures where direct data are unavailable and select the simplest method capable of answering the question.

The standards for judging sufficiency of evidence, including whether a requirement is compensatory or pass–fail, should remain the same as in a full assessment. Proxy measures are appropriate only when they meet those standards; any remaining uncertainty must be addressed during appraisal and reflected in the monitoring plan. This allows the minimum-core approach to support decisions against the same standards while requiring less time, data and specialist capacity.



Figure 11:
AIHTs/DHTs Minimum assessment matrix

TECHNICAL PERFORMANCE AND AI-SPECIFIC CHARACTERISTICS				
How accurately the model does its task, and whether that holds across populations and over time.				
Subdomain	**Admissible evidence**	**Methodology**	**Acceptability rule**	**Outcome metric**
Standalone model performance (pass-fail)	• Manufacturer dossier or validation summary and instructions for use: intended purpose, version, test population and reference standard • At least one test set independent of model training/tuning: public regulatory or HTA summary or published external validation where available • If routine records permit: consecutive local cases with AI output, usual clinician result and recorded final diagnosis • For risk scores: predicted risk and observed events in the same records	• Desk review: data separation, reference standard, intended population and risk of bias • Local check on available cases against recorded final diagnosis; for risk scores, comparison of mean predicted with observed event rate	• At least one validation independent of development data, with a defined reference standard; no high risk of bias • Where compared on the same cases, lower 95% CI of the difference against routine clinician performance above -Δ • For risk scores, observed-to-expected ratio compatible with 1; sample size/precision set at Stage 2	• Sensitivity and specificity; AUROC where reported • Difference against clinician/comparator, 95% CI • Observed-to-expected ratio, 95% CI (risk scores)
External validity and transferability (pass-fail)	• Validated populations, sites, language, prevalence, case mix and exclusions from IFU and validation reports • Local case mix/prevalence: national or facility HMIS/RHIS, registries, disease surveillance or national surveys; WHO/GHO or GBD estimates only where local data are absent • Equipment, software, connectivity and working language: facility asset lists, service records or short facility checklist (WHO SARA, where available)	• Side-by-side comparison of the local case profile and facility set-up against the validated range; identification of material mismatches	• No more than 5% of local cases fall outside the validated range • No unresolved safety-critical mismatch in input, equipment, language or care pathway	• Local cases outside validated range (%) • Safety-critical set-up mismatches (n)
Human–AI team performance (pass-fail, assistive and augmentative tools only)	• Published reader/user study of the same product and comparable user cadre, where available. • Instructions for use and training materials; local user cadre/experience and a short set of anonymised routine cases or standard vignettes	• Paired case exercise with intended users: recording of unaided and AI-aided decisions; comparison with the recorded final diagnosis or senior-clinician review	• Evidence includes the intended local user cadre and experience level • Lower 95% CI of the aided-minus-unaided difference above -Δ, overall and by user group • No unresolved safety-critical error	• Correct final decisions with and without AI (%) • Paired difference, 95% CI • Critical use errors (n)

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Version identity, lifecycle stability and drift (pass-fail, locked models only at lower readiness)	• Unique model/product version and configuration from IFU, product label or system information screen • Supplier release notes and update-notification contact; confirmation that the model does not learn or update automatically • One drift signal obtainable from routine records (e.g. observed-to-expected events, override or referral rate); rollback/disable instructions	• Verification of a match between the assessed and deployed versions/configurations • Review of available routine records for the selected drift measure or justified proxy against its baseline (e.g. observed-to-expected events, override or referral rate) • Confirmation of the action trigger, responsible person and rollback/disable route	• Locked, identifiable version; no silent or automatic update • Any update suspends use pending review against the assessed version • Direct performance alert at a decline no greater than Δ; where a proxy is used, its baseline and action threshold are justified at Stage 2 • A proxy signal triggers Stage 5 review; it does not establish model accuracy • Rollback/disable available	• Version/configuration match (yes/no) • Selected drift measure/proxy, baseline and latest value (where routine data permit) • Trigger status (reached/not reached); responsible person documented (yes/no) • Rollback/disable available (yes/no)
Generative and agentic output performance (pass-fail, where applicable)	• Model/configuration version, system prompt, retrieval sources, connected tools and safeguards • Short set of routine cases and out-of-scope prompts in the local working language, with source records or clinical guidance for checking outputs • Sample retrieval/tool-call and action logs, for systems that use external tools	• Independent rating by two clinicians of clinically material accuracy and completeness; resolution of disagreements • Repetition of a subset with reworded prompts; comparison of error-free output rates • Review of sampled retrievals and tool calls against the correct source or intended tool in available logs • Sandbox testing of agentic actions, including out-of-scope requests	• Lower 95% CI of error-free outputs versus current practice above -Δ; no critical fabricated or omitted content • Lower 95% CI of the difference under prompt variation above -Δ • Retrieval/tool-call error below the performance goal set at Stage 2 • No agentic action outside the permitted boundary	• Error-free outputs; critical fabrications or omissions (n) • Difference in error-free outputs under prompt variation, 95% CI • Correct retrievals (%) • Correct tool calls (%) • Out-of-scope actions (n)

SAFETY

The harm the technology can cause when it errs, and whether oversight catches it in time.

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Clinical hazards and residual risk (pass-fail)	• Manufacturer risk-management summary and instructions for use, including critical hazards and controls • Official regulator or manufacturer recalls and critical safety notices affecting the product/version • Local confirmatory test, referral and backup capacity from facility records or a short facility checklist	• Assessment of each critical hazard, its control and residual risk against the local care pathway • Verification of official product notices and of the local implementability of listed controls	• Every critical hazard has a documented control applicable to the local pathway • No unacceptable residual risk • No open recall or critical safety notice affecting the assessed version or intended use	• Critical hazards with applicable controls (%) • Unacceptable residual risks (n) • Open relevant recalls/notices (n)
Fail-safe operation and human oversight (pass-fail)	• Instructions for use: user confirmation, override, escalation and downtime procedures • Manufacturer usability information; local manual comparator pathway and available fallback resources (power, connectivity, staff, confirmatory testing/referral)	• Tabletop walk-through with intended users: incorrect, missing or delayed output and system/power/connectivity failure • Recording of detection, override/escalation and time to return to the comparator pathway	• Every critical scenario detected and safely overridden or escalated • Local fallback available within the clinically required time • No unresolved critical use error	• Scenarios safely managed (%) • Scenarios without local fallback (n) • Time to fallback; critical use errors (n)

CLINICAL AND SERVICE EFFECTIVENESS

The effect of the AI-enabled pathway on patient outcomes and on the service that adopts it.

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Patient-relevant clinical effectiveness (comparative; patient outcomes where relevant)	• Published comparative studies, HTA reports, guidelines or systematic reviews for the same intended use and care pathway • Local baseline event rates: HMIS/RHIS, routine facility registers, disease surveillance or national surveys; WHO/GHO or GBD estimates only if national data are absent • Local referral completion, confirmatory testing and treatment availability: routine service records or facility checklist • Standalone accuracy from the technical domain: not a substitute for outcome evidence	• Application of the published comparative effect to the local baseline rate • Transparent linked-evidence calculation where direct outcome evidence is absent: detection/management change × confirmation and treatment completion × published treatment effect	• Population, comparator and outcome match the Stage 2 PICO • Every link in an inferred-benefit chain is supported by applicable evidence • Lower 95% uncertainty bound meets the Stage 2 superiority/non-inferiority rule • No clinical benefit inferred from accuracy alone; no downstream benefit credited where confirmation or treatment is unavailable	• Relative effect and 95% CI • Localised effect/events per 1,000 patients, 95% uncertainty interval • Referral, confirmation and treatment completion (%)

ECONOMIC VALUE

What the pathway costs, whether its benefit justifies that cost, and whether the budget can absorb it.

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Cost-effectiveness / cost-consequence (comparative)	• Published economic evaluation or HTA of the same technology or a comparable pathway • Supplier quotation/procurement price for licence, hardware, connectivity, compute, integration, training, maintenance and updates • Local salaries, tariffs, procurement prices and service costs; WHO-CHOICE service costs (2010 base-year), updated to current prices, only as a secondary proxy • Localised effect estimate from the clinical domain	• Adaptation of a published evaluation using local prices, resource use and baseline rates; simple ingredients-based costing • Cost-utility analysis only where credible QALY estimates and a relevant local threshold are available; otherwise, cost-consequence analysis in natural units	• Cost-utility: ICER below the applicable local threshold, or the technology is cheaper and at least as effective • Cost-consequence: report incremental costs and each consequence separately for Stage 4; do not infer cost-effectiveness without a valid threshold or import a generic GDP-multiple threshold	• Cost-utility: ICER (cost per QALY gained) and net monetary benefit, where estimable • Cost-consequence: incremental cost per patient and consequences in natural units (e.g. cases detected, referrals completed, events avoided)
Budget impact and affordability (payer perspective)	• Eligible population and current service volumes from HMIS/RHIS, facility registers or national statistics • Expected uptake from programme plans or explicit low/base/high scenarios • Total cost of ownership and local unit costs; available programme/payer budget or procurement allocation	• Payer budget-impact calculation over 1–5 years, using low, base and high uptake scenarios	• Base-case net cost within the available budget over the stated horizon • Where unaffordable, consider phased or managed entry rather than assuming affordability from cost-effectiveness	• Net annual and cumulative cost by uptake scenario • Share of available budget (%) • Cost per eligible/treated patient

LEGAL AND REGULATORY ASPECTS

Whether the technology is authorised for its intended use, who answers when it fails, and its conformity with AI and data law.

Subdomain	Admissible evidence	Methodology	Acceptability rule	Outcome metric
Regulatory status and intended-use conformity (pass-fail)	• National competent-authority register, authorisation or written exemption/determination • Instructions for use: intended purpose, target population, user, setting, risk class and version • Where national records are unavailable, a relevant reference-authority decision may support product evidence; it does not replace the national legal determination	• Comparison of the proposed population, user, setting, function and version with the locally authorised or documented intended use	• Valid local authorisation, written exemption or competent-authority determination • Proposed use and assessed version remain within the documented intended purpose	• Regulatory status (valid/exempt/missing) • Proposed-use elements outside intended purpose (n) • Version match (yes/no)
Data protection, accountability and AI-specific obligations (pass-fail)	• Applicable national data-protection, health, device and AI rules/policies; privacy notice and data-flow description • Data-protection impact assessment where required; otherwise documented screening of privacy and data-protection risks • Supplier contract/data-processing agreement: purpose, access, hosting, retention, sharing, cross-border transfer and reuse for model training • Named persons/roles for clinical use, human oversight, incident reporting, data protection and software updates	• Documentary check against applicable national law and official requirements, including the required impact assessment or documented risk screening • Traceability review of patient data from collection through storage, vendor access, secondary use/training and deletion; verification of responsibility assignments	• Legal basis documented for each processing purpose, including model training and transfer where applicable • Required agreements, permissions and transfer safeguards in place • Impact assessment completed where required; no unmitigated high privacy/data-protection risk • Liability and responsible parties assigned for clinical use, oversight, incidents and updates	• Missing/non-compliant documents or legal bases (n) • Unmitigated high privacy/data-protection risks (n) • Unassigned responsibilities (n)

Application note. This is the Stage 3 adaptation only. Apply the Stage 2 PICO, comparator, perspective, sample-size requirements and Δ/thresholds unchanged. Use national/local sources first; international estimates are gap-fillers, not substitutes for local technical validation or legal authorisation. Check routine-data completeness, timeliness and consistency. Residual uncertainty passes to Stage 4 and Stage 5. This minimum assessment core assumes a locked model; adaptive models require fuller lifecycle controls.

3.6

Profiling the decision context

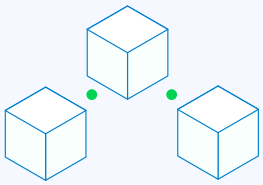
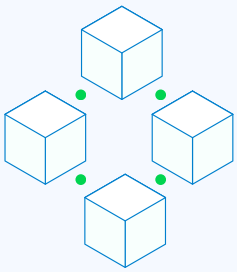
The assessment approach should reflect two separate features of the decision context: the maturity of the HTA function and the readiness of the organisation that will deploy the technology. Neither can be inferred reliably from national income or another single economic indicator[61],[63]. Profiling both dimensions helps determine where detailed local analysis is needed and where external evidence or simpler methods are sufficient. Any adaptation must retain the standards in Chapter 2 and the minimum assessment core in Section 3.5.



3.6.1 HTA maturity

HTA maturity describes how firmly assessment is established within a health system. The WHO framework identifies five components: mandate, legal basis, institutional and governance arrangements, assessment and appraisal procedures, and monitoring and evaluation[12]. Maturity does not follow national income: systems at similar levels of development may have very different capabilities[61]. Figure 12 distinguishes three broad levels. Where HTA is absent, decisions may rely on external assessments and the minimum core. A developing function can adapt existing work, validate it locally and gradually build its own evidence-generation capacity. An established function can undertake fuller assessments, link findings to funding or procurement and support ongoing reassessment.

Figure 12:
Proportionate assessment by HTA maturity

HTA INSTITUTIONAL MATURITY	SETTING AND DECISION BASIS	ASSESSMENT IMPLICATION
Absent 	No formal HTA body, mandate, or dedicated assessment capacity. Without a statal mechanism, allocation defaults to implicit priority setting, ad hoc, decentralised and often sponsor-driven, in which it is unclear how or by whom spending is decided. External references are typically applied without local foundation: a foreign decision adopted reflexively or a borrowed benchmark used without adjustment (i.e. the 1-3 times GDP per capita threshold) [83],[84],[85],[86],[87].	■ Recognition of an external decision where a demonstrated equivalence exists, otherwise reliance on a qualified external assessment that passes the eligibility test (§3.2). ■ Adoption of transferable components, with direction of the scarce local effort onto those that do not transfer. ■ Repurposing of the domestic regulator's dossier, or a foreign reference regulator's where national regulatory capacity is also thin, for the technical-performance and safety domains (§3.3). ■ Regional collaboration enters as either a consumer of shared outputs, horizon-scanning, methodological templates, safety exchanges, and pooled monitoring or a supplier of local epidemiological, price, and case-mix inputs that allow a shared assessment to be calibrated to the local setting (§3.4). ■ Use of the minimum assessment core (§3.5).
Developing 	A formal unit or specialised committee collaboratively undertakes selective tasks, through a nodal agency or on an ad hoc basis, with limited data capacity. It primarily reuses and adapts assessments made elsewhere, de-duplicates the transferable parts, and directs its own effort onto the nontransferable domains of national appraisal[73],[83],[84],[86].	■ Adaptation of the external assessment to the local population, comparator, pathway, and prices, with local revalidation of model performance on local case-mix, prevalence, decision threshold, and subgroups (§3.2). ■ Conditional adoption with managed entry, the residual uncertainty on real-world performance discharged through prospective local collection of performance and safety data (§2.4). ■ A version-monitoring plan for adaptive models and drift surveillance against the deployed version, to build the local monitoring that independent assessment will require (§2.5). ■ Contribution to joint assessment and scientific consultation with partners, to develop method and workforce toward de novo capability (§3.4). ■ Core domains assessed locally, with data governance, cybersecurity, interoperability, and algorithmic-justice domains activated where inputs are sensitive, integration extensive, or exposure broad (§3.5).
Established 	An arrangement with a broad remit and a defined governance, legally integrated into the health policy process, applying assessment and appraisal with stakeholder engagement, and lifecycle integration, with its role extending to pricing and reimbursement[84],[88]. At facility level the function is proportionally carried by hospital-based units for procurement purposes.	■ Full AI-HTA across the complete domain set, with comparative evaluation of the AI-enabled pathway against the current pathway (§2.3). ■ Regulator's technical and safety appraisal repurposed to avoid duplication. ■ Lifecycle reassessment triggered by version change and post-market signal, using provider-level teams for adaptive, high-impact technologies (§2.5). ■ The system contributes to regional work as a reference producer (§3.4).

The Healthcare AI Governance Readiness Assessment (HAIRA) provides a five-level model covering seven domains, from organisational structure and problem formulation to development, validation, deployment, monitoring and maintenance.

3.6.2 Deployment-organisation AI readiness

Before an AIHT or DHT enters routine use, assessors should also consider whether the deploying organisation can operate and govern it safely. Technical effectiveness does not guarantee that a technology is feasible in its intended setting. Readiness should be assessed where the technology will be used, whether a hospital, provider network, department, service or clinical programme. It may vary within a single organisation: for example, radiology or pathology may have more advanced capabilities than other services[89]. HTA maturity largely determines how the assessment is produced[61],[63]; organisational AI governance readiness helps determine which adoption recommendation can be implemented safely.

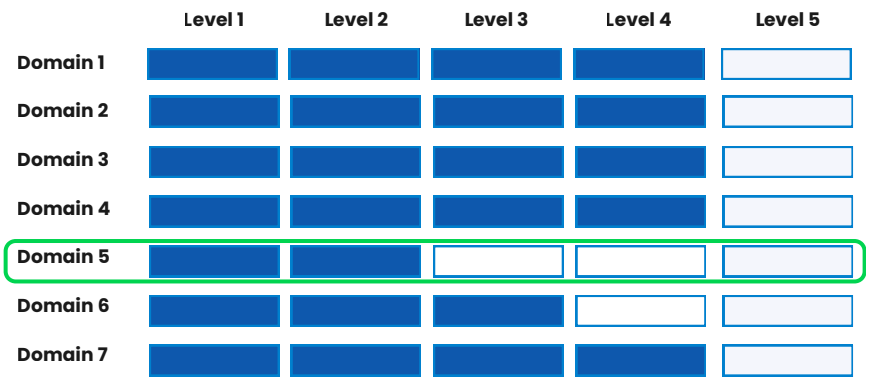
The expected benefits of a technology may not be realised if the provider cannot implement, monitor or maintain it. Deployment readiness therefore shapes both the type of recommendation that can be supported and the conditions attached to it, including safeguards and restrictions on use.

The Healthcare AI Governance Readiness Assessment (HAIRA) provides a five-level model covering seven domains, from organisational structure and problem formulation to development, validation, deployment, monitoring and maintenance[89]. Each level includes enabling requirements for infrastructure and expertise, from basic IT systems and digital skills to advanced data infrastructure and specialist teams[89]. Overall readiness is constrained by the weakest essential domain. For example, a provider without adequate incident reporting, internal validation or clear governance responsibilities cannot be classified at an advanced level solely because it performs well elsewhere[89]. Tool 4 translates this model into a self-assessment for individual providers or wider systems.

For the system-level analysis in this blueprint, HAIRA is applied to the prevailing readiness of healthcare providers collectively rather than to one organisation. The model also allows assessment at health-system and service-line levels[89]. Figure 13 links each readiness level to the types of HTA recommendation and implementation conditions that may be appropriate. This system-level approach assumes broadly similar governance maturity across providers; differences between sites should be considered when applying a recommendation locally.

Readiness depends on the lowest domain rating

ILLUSTRATIVE EXAMPLE



Overall Readiness

Level 2

The lowest domain rating sets the overall readiness

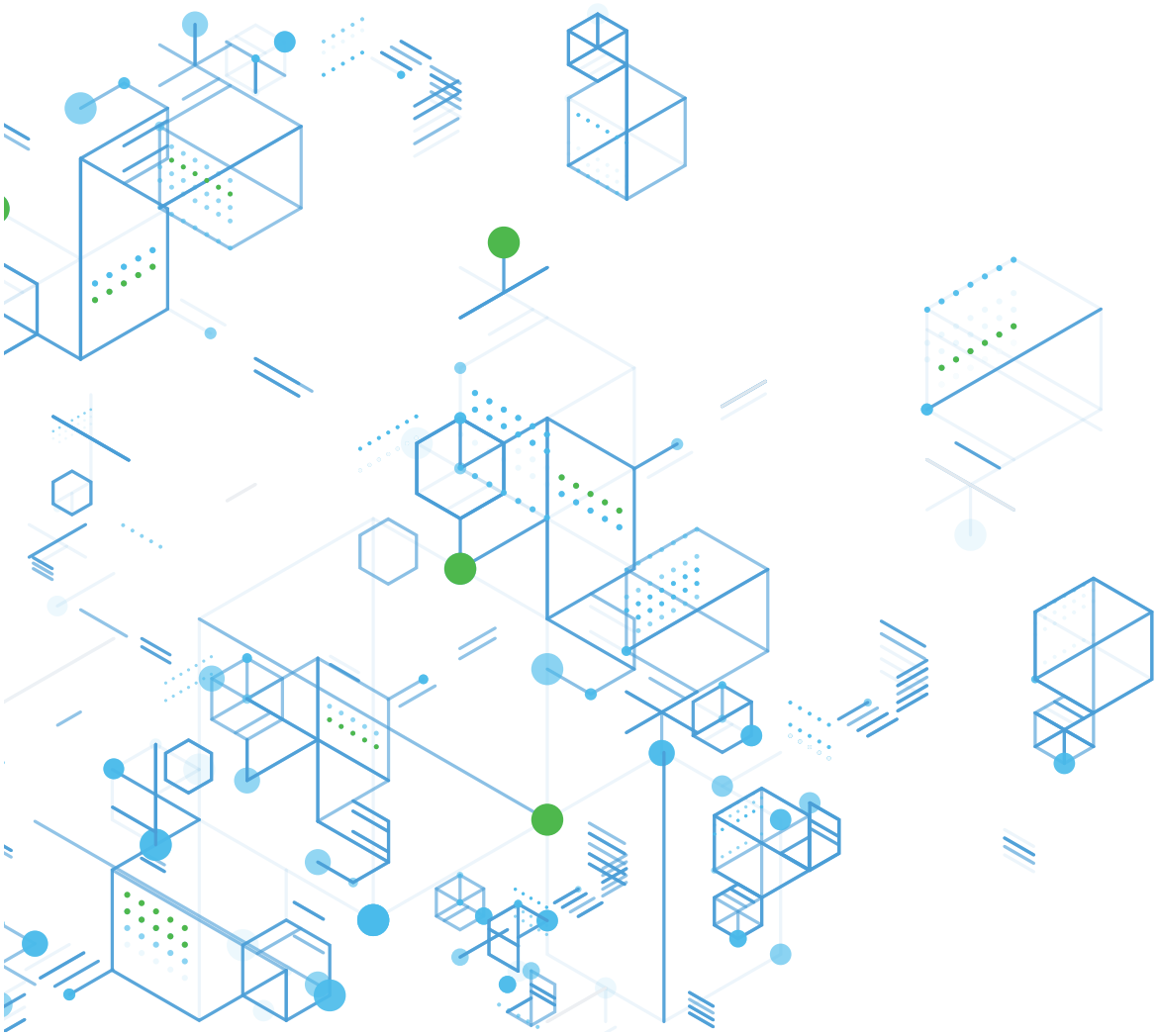
Figure 13:
Recommendation conditions by provider-system AI-governance readiness

AI GOVERNANCE READINESS LEVEL	GOVERNANCE AND INFRASTRUCTURE DESCRIPTION[89]	HTA RECOMMENDATION IMPLICATION
L1. Initial / Ad hoc	No formal AI governance or dedicated technical capability, reliance on vendor claims, confirmatory acceptance testing, manual deployment and monitoring reactive where present	■ Autonomy confined to assistive, human-confirmed use. ■ A positive recommendation is admissible only for lower-risk, regulator-cleared technologies under vendor-supplied validation and the monitoring terms attached to the ruling (Tool 4). ■ Adaptive, generative, and clinically influential systems exceed the organisation's oversight capacity and default to a supervised pilot or research study, or, where the residual organisational risk is intolerable, to refusal.
L2. Defined	Clinical and IT governance committee, documented processes, and structured vendor evaluation, with basic workflow integration and access controls	■ Validation, update-control, and surveillance duties transferred to the supplier, with financial risk carried through warranty or performance-linked payment. ■ Where the assessment itself cannot be resourced, the committee may adopt a qualified external evaluation and fund provisionally against local revalidation (§3.2, §3.6.1). ■ A shorter mandatory review interval applies, with the reassessment date set earliest where local surveillance is weakest.
L3. Established	Multidisciplinary oversight, independent internal validation with bias assessment, and monitoring with intervention thresholds	■ Conditional recommendations become admissible for higher-risk technologies. ■ Expected value insufficiently secure for reimbursement supports provisional funding against an agreed evidence programme, and doubt confined to local performance supports a supervised pilot under shadow deployment before wider commitment. ■ Standard recommendation is reserved for technologies whose monitoring indicators and intervention thresholds the organisation can service (Tool 4, §2.5).
L4. Advanced	Executive AI ownership, domestic development capability, prospective validation, and real-time automated monitoring	■ The full recommendation set is admissible, including routine adoption of adaptive and continuously updated technologies, with monitoring intensity and reassessment triggers scaled in the ruling to model volatility.
L5. Leading	Full lifecycle governance with continuous, predictive monitoring and a research-grade validation and development function	■ The site's own prospective validation and real-time surveillance may supply the evidence on which provisional terms are discharged and living reassessment conducted (§2.5).


Doctor
Administrator

4

Tools and Templates



Tools and templates

The five tools, adapted largely from published frameworks and HTA practice, in this chapter support the assessment process described in Chapters 2 and 3. Each provides a consistent way to record evidence, decisions and remaining uncertainty at a particular point in the access pathway. Standardised templates make assessments easier to review, compare and audit across technologies and institutions. The tools do not introduce new methods, evidence requirements or acceptability thresholds.

The tools may be used as paper forms, spreadsheets, dossier templates or fields in an electronic workflow. Their format can be adapted to local practice, provided the essential information is retained. Not every tool is needed for every assessment: each section explains when it should be used. Where a tool draws on published guidance, its source is identified.

Which tool and template to use?

TOOL	USE IT WHEN YOU NEED TO
Tool 1 Decision-context card	Record the standing characteristics of the assessing body and the system in which it operates.
Tool 2 Evidence gap matrix	Compare the evidence available in a submission with the requirements identified for Stage 3.
Tool 3 AIHT Economic Evaluation Reporting Checklist	Support consistent reporting and review of economic evaluations of technologies whose function depends on AI.
Tool 4 AI readiness self-assessment	Evaluate an organisation's governance and infrastructure capacity to deploy and manage AI safely.
Tool 5 Managed access agreement	Record the obligations attached to conditional use following a Stage 4 decision on provisional funding, a supervised pilot or restricted adoption.

4.1

Tool 1 – Decision-Context Card

The Decision-Context Card records the standing characteristics of the assessing body and the system in which it operates. It sets out the assessing body's mandate, decision rights, reference case, approach to recognition and reliance, and capacity to generate evidence and monitor technologies after adoption. It is suitable for both national agencies and hospital HTA committees, and makes their assessments more transparent and comparable. Unlike Tools 2, 3 and 5, which are completed separately for each technology, the card is prepared once for the assessing body and updated when its mandate, methods or capacity change. It draws together information established in Section 3.6 on HTA maturity and deployment readiness and in Section 3.2 on reliance. At Stage 1, it provides a consistent account of the decision-maker and the context in which each assessment is undertaken.

Figure 14:
Decision-context card

Field	Description	Populated from
Jurisdiction and decision rights	Decision level, the systems legally bound by it, and whether the output is a coverage decision, a procurement decision, or advisory	§2.1
Technology scope	The technology classes the unit or agency is authorised to assess or appraise	Based on assessor's own mandate and policies
HTA maturity	Institutionalisation graded on the WHO five-component framework	§3.6.1
Regulatory interface	Coordination with the medicines/ device regulator and the regulatory dossier available for reuse	§3.3
Reliance position	Recognised reference bodies and the eligibility criteria for reliance and recognition	§3.2
Evidence and RWE infrastructure	Registries, claims, EHR, and common data models for evidence generation and post-market surveillance	Based on assessor's own mandate and policies
Coverage-to-adoption linkage	The mechanism, and any statutory timeline, by which a recommendation takes effect and connects to monitoring	§2.4–2.5
Patient involvement	Patient experts nominated by patient organisations, and written submissions from those organisations on lived experience, unmet need, and the outcomes patients value	Based on assessor's own mandate and policies
AI-in-HTA position	Sanctioned use of AI within the body's own assessment workflow and its validation controls	Standing
Currency and provenance	Date compiled and the source authority for each entry	Standing

Tool 2 – Evidence Gap Matrix

The Evidence Gap Matrix compares the evidence available in a submission with the requirements identified for Stage 3. It shows which items are missing and what further work is needed. The matrix is first completed during Stage 2 (Section 2.2), once the research question is defined, and informs the evidence-development plan and any early scientific advice. Findings from the reliance eligibility test (Section 3.2) and the regulatory evidence review (Section 3.3) identify which requirements have already been addressed by an external assessment or regulatory dossier and which still require local evidence.

The example rows cover common evidence needs for DHTs and AIHTs and relate broadly to Stage 3 domains or subdomains. They may be adapted or extended to suit the technology and methods used. For example, where AUROC and calibration are not applicable to a binary output, sensitivity and specificity can be reported instead. The required-evidence field specifies the data sought and whether an outcome is final or surrogate; the available-evidence field records its source, such as a study, regulatory file or external assessment. Separate columns distinguish relevance to the HTA decision from applicability to local populations, care pathways and prices. The final column records the gap, its severity under the domain's acceptability rule (as per Figure 3), and the action, owner and trigger needed to resolve it. A compensatory gap is considered at appraisal alongside its uncertainty; a pass–fail gap must be addressed before adoption (as per Figure 3).



Figure 15:
Evidence gap matrix

Evidence item and required type	Available evidence	HTA relevance	Local applicability	Gap severity and action
Standalone technical performance Discrimination, as AUROC with sensitivity and specificity, and calibration of the model output, against a defined reference standard, on data independent of model development, reported with uncertainty and by subgroup.				
External validation AUROC and a calibration curve in an independent cohort from a different site or period.				
Human-AI team performance Change in the clinician's sensitivity, specificity, and decision time with the tool against without, in the intended workflow, with automation bias and appropriate reliance reported.				
Safety Adverse events and the clinical consequences of false results against the comparator pathway, with failure modes and fail-safe behaviour documented.				
Clinical and service effectiveness Comparative effect on patient-relevant clinical and service outcomes against the comparator pathway, including patient-reported outcome and experience measures from validated instruments, from a controlled or comparative design.				
Economic value Incremental cost-effectiveness against the comparator on the reference-case method				
Subgroup performance and equity Sensitivity and false-negative rate by key demographic subgroup				
Workflow and implementation Staff time per case and patient throughput in the reconfigured pathway				
Drift and post-market monitoring A monitoring plan with predefined performance indicators, drift-detection thresholds, and revalidation triggers across model versions.				

4.3

Tool 3 – AIHT Economic Evaluation Reporting Checklist



The checklist supports consistent reporting and review of economic evaluations of technologies whose function depends on AI. It draws on the CHEERS-AI framework developed by Elvidge et al. for health economists preparing evaluations and for HTA bodies, decision-makers, editors and peer reviewers assessing published analyses[90]. It is used for the economic-value domain at Stage 3. Analysts preparing reimbursement submissions or local procurement cases should consider its requirements when designing the evaluation at Stage 2 and complete the checklist when reporting findings. Appraisal committees can then use it to judge whether the economic evidence has been reported adequately.

CHEERS-AI comprises 38 reporting items with 28 items from CHEERS 2022, including some with AI-specific considerations, and ten new AI-specific items. The full standard remains relevant; the table below presents the 18 items that explicitly address AI.

The items follow the reporting sequence, from the title and abstract to the methods, results and discussion. They are phrased as questions, with a final column indicating where each is addressed in the submission.

Figure 16:
AIHT economic evaluation reporting checklist, AI-relevant items
(adapted from Elvidge et al.90)

No.	Section/topic	Reporting question	Reported in
Title			
1	**Title**	Does the title indicate that the intervention involves an AI component that is under evaluation?	
Abstract			
2	**Abstract**	Does the abstract specify the purpose of the intervention with an AI component, and the AI technique used?	
Methods			
7	**Comparators**	Are details of the AI component described, including its classification by intended purpose and risk tier, the AI technique used, whether it is locked or adaptive, the version under evaluation, its purpose and potential impact on care, the intended users and how they interact with it, additional requirements to use it, and how it is expected to provide benefit over the standard of care?	
AI 1	**User autonomy**	Is the AI intervention directive, or does the user retain autonomy to make the care decision?	
11	**Selection of outcomes**	Do the measures chosen for the benefits and harms of the AI intervention relate to health outcomes, diagnostic outcomes, process outcomes, or other or multiple outcomes?	
12	**Measurement of outcomes**	For model-based analysis, are the assumptions used to inform the potential benefits and harms of the AI intervention described, and their plausibility justified with supportive evidence?	
AI 2	**Measurement of AI effect**	Are the data sources for the AI intervention's impact on outcomes described?	
AI 3	**Measurement of AI learning over time**	If the AI intervention learns over time, is it explained how this affects its performance at the individual level and how this was measured?	
AI 4	**Development of AI component**	Is it described how the AI component was developed, including the training data used and how errors and biases were identified, or is a source cited?	
AI 5	**Validation of AI component**	Is it described how the AI component and its performance estimates were validated, or is a source cited?	
AI 6	**Health benefit**	Is it described how the AI intervention could directly or indirectly yield a health benefit?	
AI 7	**Population differences**	Are important differences described between the data sources for the AI intervention's impact on outcomes and the dataset used to develop it, the training dataset?	
14	**Measurement and valuation of resources and costs**	Is the purchase cost of the AI intervention and what it is composed of described, along with any additional implementation and maintenance costs?	

No.	Section/topic	Reporting question	Reported in
16	**Rationale and description of model**	Is it described whether the AI component has influenced the choice of health economic model, and why?	
AI 8	**Modelling of AI learning over time**	If the AI intervention learns over time at the individual level, are the assumptions used to model how this learning affects its performance described?	
Results			
AI 9	**Impact of AI uncertainty**	To what extent do features of the AI intervention contribute to increased uncertainty about its cost-effectiveness?	
Discussion			
26	**Study findings, limitations, generalisability, and current knowledge**	Are potential biases associated with the AI intervention, such as algorithmic bias, and the implications for the generalisability and interpretation of results, such as reinforcing existing health inequalities commented on?	
AI 10	**Implementation of AI**	Are the requirements to integrate the AI intervention into practice, and other implementation considerations, commented on, including implications for the interpretation of cost-effectiveness results?	

4.4

Tool 4 – AI Readiness Self-Assessment



How overall readiness is determined

Seven domains assessed

↓

Each rated from initial Leading

↓

Lowest domain rating = overall readiness

ILLUSTRATIVE EXAMPLE:

six domains at Level 4.
One domain at Level 2.

Overall readiness = Level 2

The AI Readiness Self-Assessment evaluates an organisation's governance and infrastructure capacity to deploy and manage AI safely. It places the organisation or provider system on the five-level readiness scale described in Section 3.6.2.

The table assesses seven areas: organisational structure; problem formulation; external algorithm evaluation; algorithm development and model training; model evaluation and validation; deployment and integration; and monitoring and maintenance. Each is rated against five levels, from Initial to Leading. The descriptions specify the governance, skills and infrastructure required at each level, and the final column records the level attained.

The deploying organisation, or a central body familiar with its provider system, compares current capabilities with the benchmarks in each area and records the highest level met. **Overall readiness is determined by the lowest rating across the seven areas.** A weakness in any one area therefore limits the maturity level assigned to the site or system, even when other capabilities are more advanced. Section 3.6.2 explains how the resulting profile informs HTA recommendations.

The tool adapts the Healthcare AI Governance Readiness Assessment (HAIRA) developed by Hussein et al.[89] for use in the HTA process.

Figure 17:
AI readiness self-assessment
(adapted from Hussein et al.89)

Domain	L1 Initial	L2 Defined	L3 Established	L4 Advanced	L5 Leading	Attained
Organisational structure	No formal AI governance, decisions rest with existing leadership	AI oversight committee with clinical and IT representation	Multidisciplinary AI governance committee with defined roles	Executive AI officer leading strategy, with specialist subcommittees	AI governance centre of excellence shaping external standards	☐
Problem formulation	Clinical need identified informally, no structured process	Structured needs assessment with basic ROI evaluation	Needs assessed against clinical, technical, and ethical criteria	Documented multi-year AI adoption plan aligned to objectives	Develops novel use cases ahead of the field	☐
External algorithm evaluation	Accepts vendor claims and certifications	Documented criteria for vendor selection	Internal testing of vendor tools before adoption	Real-world validation of vendor tools	Sets external evaluation standards	☐
Algorithm development and model training	No in-house development, commercial solutions only	Limited customisation of commercial solutions	Basic in-house development alongside vendors	Substantial in-house development with research capacity	Leading-edge in-house development	☐
Model evaluation and validation	Acceptance testing only	Structured testing with basic validation	Internal validation including bias assessment	Prospective validation studies	Multi-centre validation studies	☐
Deployment and integration	Manual use, minimal system integration	Integration planned around workflow	Structured rollout with change management	Fully integrated into existing workflows	Continuously optimised deployment	☐
Monitoring and maintenance	Reactive, driven by user reports	Periodic performance review on basic metrics	Proactive monitoring with defined thresholds	Real-time monitoring with automated alerts	Predictive monitoring using advanced analytics	☐

Tool 5 – Managed Access Agreement



The Managed Access Agreement records the obligations attached to conditional use following a Stage 4 decision on provisional funding, a supervised pilot or restricted adoption. It does not apply to routine adoption, refusal or withdrawal. It provides a framework for outcome-based managed entry and coverage with evidence development: use proceeds under defined conditions while further evidence is collected to address the uncertainties identified at appraisal. The agreement sets out the responsibilities of the payer, deploying provider and vendor.

The agreement refers to the eligible population, approved model version and oversight requirements specified in the recommendation rather than restating them. It defines the evidence to be collected, the study design and endpoints, the period allowed and the party responsible for each dataset. These requirements follow the relevant Stage 3 domain standards (Section 2.3). It also covers funding for use and evidence generation, data ownership and governance, interim reports, stop rules and reassessment options. AI-specific provisions require notification and approval of changes or retraining, access to audit records and support for safe exit or transition. Price and financial risk-sharing terms are negotiated separately by the payer and recorded by reference, since the HTA body generally determines conditions of use rather than price.

The agreement is prepared after the Stage 4 decision (Section 2.4) and operates through Stage 5 (Section 2.5). Monitoring indicators and thresholds provide the basis for stopping use or initiating reassessment, which may renew, revise or end the agreement. Deployment is permitted only once the organisation has met the readiness conditions in Tool 4. The HTA recommendation sets the conditions of use; the agreement makes the payer's, provider's and vendor's corresponding responsibilities enforceable.

Figure 18:
Managed access agreement

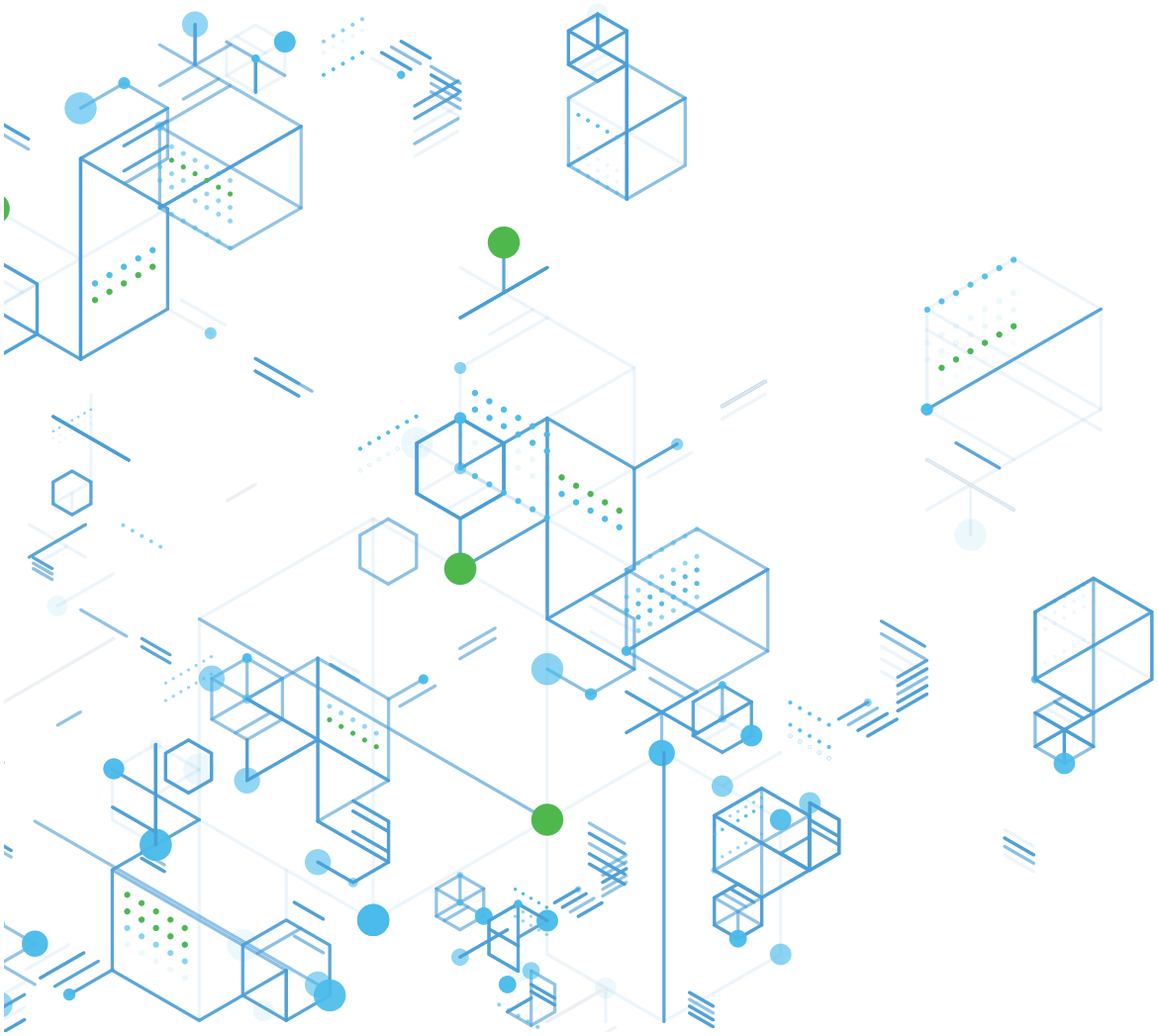
Provision	Detail and responsible party
Scope	
Basis of provisional use	Decision type, provisional funding, supervised pilot, or restricted use, and the residual uncertainty it addresses. Refers to Stage 4.
Eligible population, sites, version	As named in the recommendation, referenced not restated.
Evidence generation	
Evidence questions	The uncertainties the term must resolve, set by the appraising body.
Study design and data source	Prospective registry, real-world data, or interventional trial, with endpoints held to the Stage 3 domain standard. Developer/submitter.
Term	Defined period for the evidence to mature, with an outer limit, indicatively twelve months extendable to twenty-four.
Responsibilities	
Data collection and quality	Which party captures which data, to what completeness standard. Developer/submitter and deploying provider.
Data ownership, access, governance	Custody, permitted use, and data-protection compliance. All parties.
Funding	Cost-bearer for the technology and for evidence generation (including data infrastructure) over the provisional period, and the supply and payment basis: licence or subscription, per-use, hybrid, outcome-linked, or free-of-charge for the term. Payer/Funder, with developer/submitter contribution as agreed.
Control and closure	
Interim reporting	Reporting points against the Stage 5 monitoring dashboard. Developer/submitter and deploying provider.
Stop rules	Signals and thresholds that suspend or discontinue use before term. Appraising body.
Reassessment	Date and decision options, continue, modify, restrict, retrain, retire. Appraising body.

Provision	Detail and responsible party
AI-specific control	
Version and change control	Locked model version and a notification duty on any update. Developer/submitter.
Retraining and replacement	Approval before a changed model operates, on developer/submitter notification. Appraising body and developer/submitter.
Vendor obligations	Model access for audit, performance-log provision, exit and transition duties. Developer/submitter.
Commercial (referenced)	
Price and financial risk-share	Agreed separately by the funder/payer. No appraising body involvement.
Signatories	Appraising body, funder/payer, deploying provider, and developer/submitter.



5

Governance and Setup of an AI-HTA Unit



5.1

Institutional entry models

Before assessing digital and AI-enabled technologies, a government, payer or provider network must decide what legal authority the assessment body will have and how it will be organised. These choices are not unique to AI: existing HTA systems offer established institutional models. They also matter in high-income health systems, where assessment pathways for medical devices and other non-pharmaceutical technologies are often less established than those for medicines[91].

Four institutional models

Statutory agency

Unit or department within ministries of health or national health authorities

Expert advisory committee

Hospital-based unit

This section examines the mandate and four common institutional models, including the settings in which each is suitable and its resource implications.

HTA arrangements differ across countries in four main ways: the legal mandate linking assessment to decisions[88],[91]; the location of the assessment body and its degree of independence from government[84],[91]; the division between evidence assessment and appraisal[84],[91]; and the level of the health system at which decisions are taken[92].

The mandate should be defined before the institutional model. It identifies who may conduct assessments, which decisions require them, including coverage, procurement, pilots and disinvestment, and whether recommendations are advisory or binding. In a WHO survey of 104 countries, 52% reported a legislative requirement to consider assessment results in coverage and benefit-package decisions, rising to 67% among high-income countries, and 31% regarded those results as binding by law[88]. A study of 63 HTA bodies in 32 high-income settings found that 81% issued non-binding recommendations, with limited clarity about their influence on reimbursement negotiations[91]. Legal labels can understate practical effect: positive recommendations from NICE and the Scottish Medicines Consortium, for example, can carry funding obligations for the NHS despite their advisory form[91]. The strength of the link to funding determines whether appraisal findings affect spending and its opportunity costs. If the payer can disregard recommendations, it may fund a technology assessed as poor value or reject one assessed as cost-effective[91],[93].

Once the mandate is established, the appropriate model depends on existing HTA capacity[12], the distribution of budgetary authority[94], available staff and funding[12],[63],[84],[88], and political commitment[95]. Four models recur: three at national level and one within healthcare providers.

i. Statutory agency

Statutory or arm's-length agencies for example operate in England (National Institute for Health and Care Excellence, NICE), France (Haute Autorité de Santé, HAS), Germany (Institute for Quality and Efficiency in Health Care, IQWiG) and Canada (Canada's Drug Agency, CDA)[84]. They generally have broader mandates and larger permanent teams than ministry-based units or advisory committees[84]. Even so, only 12% of 104 countries surveyed reported an HTA entity with more than 100 full-time-equivalent staff[88]. Some agencies organise work both by technology type, such as medicines, devices and vaccines, and by specialist discipline, including health economics, biometry, real-world data and health information[96],[97],[98].

Institutional independence can strengthen scientific credibility, although delegating decisions away from elected governments also raises questions of democratic accountability[84]. Manufacturers usually submit clinical and economic dossiers, while some systems also assess topics proposed by ministries, payers or clinicians[84]. Agency teams review the evidence, appraise its relevance to national priorities and budgets, and publish their findings[84]. Maintaining permanent specialist staff gives this model relatively high fixed costs[88]. AI-enabled technologies also increase the range of methods and expertise established agencies must provide[84].

Where a national agency already assesses medicines, digital and AI-enabled technologies can be handled through a dedicated programme using the existing mandate, committee and reference case[45]. Some agencies have combined previously separate programmes for diagnostics, procedures and medical technologies, explicitly extending them to digital and AI-enabled interventions[99]. Applying a common economic reference case and cost-effectiveness threshold reflects the health displaced when technologies compete for the same healthcare budget.

ii. Unit or department within ministries of health or national health authorities

Examples of HTA functions embedded within government or national health authorities include Brazil's CONITEC, Singapore's Agency for Care Effectiveness and Malaysia's MaHTAS[84], Thailand's Health Intervention and Technology Assessment Program (HITAP)[83] and India's Office of Health Technology Assessment in India (HTAIn)[95]. These bodies may be ministry departments or semi-autonomous research organisations linked to a ministry, health authority or payer[84]. Their findings inform decisions on subsidies, high-cost medicines and benefit packages[83]. Some operate through a secretariat or appraisal board supported by academic resource centres commissioned to prepare assessments[95]. Embedded units generally have smaller permanent teams than statutory agencies[84] and are common in emerging systems.

Six of eight ASEAN countries surveyed located their lead HTA bodies within government[83], and comparative research suggests that an embedded unit linked to actual decisions is often a practical starting point in low- and middle-income countries[84]. Proximity to budget holders may improve policy relevance[84], although some authors favour eventual organisational independence to protect scientific credibility[100]. When one programme both assesses evidence and appraises value, conflict-of-interest safeguards and transparent reasoning are essential. Government location may also protect a developing unit from undue influence by external partners[95].

Across 104 countries, ministries of health or national health services nominate assessment topics in 64% of cases[88]. In embedded arrangements, the host ministry may exert particularly strong control over the assessment agenda[101],[102],[103]. Emerging HTA bodies may also receive few or no industry submissions[61],[83],[104],[105].

For AIHTs, voluntary submission can leave assessors without developer dossiers, while real-world evaluations and cost-effectiveness evidence remain limited[26]. In these circumstances, assessors may need to draw on routine data from public facilities that have already deployed the technology[106].

iii. Expert advisory committee

Expert advisory committees are used in Australia, including the Pharmaceutical Benefits Advisory Committee (PBAC) and Medical Services Advisory Committee (MSAC)[84], and in parts of the Swiss system, including its federal committees for pharmaceuticals, services and medical aids[107]. This model is uncommon: a WHO survey found only three of 111 countries with a committee or university-affiliated unit outside the ministry[64]. Committees generally comprise independent, part-time members[84],[108] appointed under legislation or government decision[109],[110],[111]. They appraise evidence prepared elsewhere and advise public authorities on funding[84],[110],[112]. Membership commonly includes clinicians, other health professionals, economists and consumer representatives, sometimes organised by technology type[107],[109].

The legal effect of their advice varies. A positive recommendation may be necessary for listing without obliging a government to fund it[113],[114]; elsewhere, the final decision remains entirely with the responsible public authority[84],[91],[112]. Evidence assessment is often contracted to independent teams, either for most new submissions[115] or for reassessment of existing services[116]. Teams may be selected through standing procurement panels or commissioned for each topic[115],[116], while the ministry may still check dossier completeness and identify key issues[112]. Contracting converts some permanent staffing costs into per-assessment expenditure and allows access to specialist AI expertise. Its effectiveness depends on an adequate supply of qualified independent teams and safeguards against conflicts of interest.

iv. Hospital-based unit

Hospital-based HTA is used in provider organisations in several countries, including Hospital Clínic de Barcelona in Spain, Odense University Hospital in Denmark and Centre Hospitalier Universitaire Vaudois in Switzerland[117], McGill University Health Centre in Canada[118] and NATS units in Brazil[76]. It supports local management decisions on equipment, devices, medicines, procedures, organisational changes and digital health technologies[117].

The literature identifies four models: independent groups, integrated-essential, stand-alone and integrated-specialised units, distinguished by their formal status, expertise and relationship with other HTA bodies[117],[119]. Small teams drawing on universities and research centres were the most frequently reported configuration, representing 40% of units surveyed in 2023[119]. The presence of a national or regional agency can influence whether a hospital unit develops as an integrated or independent function[117].

Units are most common in hospitals with more than 500 beds. They have a median staffing level of three full-time or two part-time members, and only 37% report a dedicated budget[119]. In Latin America and the Caribbean, hospital HTA is concentrated in public teaching hospitals, often relies on part-time staff and faces shortages of specialist training[76]. Assessment is optional in 61% of hospitals with an HTA function[119]. Department heads usually initiate requests, while chief executives make the final decision in 68% of cases[119]. Hospital decision-makers often give greater weight to budget impact than to societal cost-effectiveness[80]. Units commonly assess clinical need, the appropriate care setting, the available market and procurement options[119]. Because national reports may not adequately cover organisational implications[117], hospital managers often need additional local evidence[80]. A typical output is a mini-HTA report addressing around 15–25 clinical, safety, economic and organisational questions, drawing on published evidence and expert judgement[117],[119].

Hospital HTA is especially relevant when technologies are too new for comprehensive agency review or have not been assessed nationally[117]. More than a third of hospital units already assess digital health interventions, typically producing one to four such reports annually[119]. For AIHTs, retrospective accuracy may not hold in a new care setting, and data or concept drift can change performance over time[26]. Local assessment can examine the hospital's own infrastructure, patient population, clinical alternatives and learning curve[117]. Four in five surveyed units use data from hospital databases[119]. Their close connection to hospital management[119] and shorter production times of 2–12 weeks[117], compared with months or years for some national assessments[76], also make repeated review feasible. Among surveyed units, 47% occasionally reassessed technologies and 9% did so consistently, generally 12–36 months after initial evaluation[119].



5.2

Workforce and competencies

HTA is multidisciplinary, so an effective team needs complementary expertise rather than expecting every individual to cover all assessment domains[120]. Appraisal bodies are commonly staffed by clinicians, public health specialists, government officials and economists[88]. However, skilled staff remain scarce. Human-resource shortages were among the main barriers reported in a survey of 104 countries[88]; all but one ASEAN respondent reported inadequate local technical capacity[83]; and 47.1% of experts in an international survey reported recruitment difficulties[121].

AI adds to these capability gaps, including for regulators. The US Food and Drug Administration has called for stronger expertise in AI evaluation and highlighted the limited capacity of developers and clinicians to assess systems repeatedly in local settings throughout their lifecycle[122]. Hospitals that develop their own AI tools are more likely to test predictive models locally for accuracy and bias, probably because development and evaluation draw on related technical skills[123]. Some newer HTA competency frameworks address AI, but mainly as a tool for conducting HTA activities such as literature reviews and data analysis, rather than as a technology being assessed[124].

Capacity-building programmes also identify a need for stronger assessment skills beyond pharmaceuticals, particularly for medical devices[125]. In a multidisciplinary Delphi study, experts identified 14 of 20 additional considerations as important for AIHT assessment and judged the EUnetHTA Core Model insufficient on its own[126].



The eight competency areas below combine established HTA skills with the additional capabilities needed for AIHTs. They are not ranked by importance; collectively, they define the expertise that an assessment organisation should be able to access, regardless of its institutional model.

Eight competency areas

1	**iHTA methodology and evidence synthesis**	5	**AI, data science and health informatics**
2	**Health economics**	6	**Ethics, equity and law**
3	**Clinical safety and patient perspective**	7	**Organisation and implementation**
4	**Epidemiology and biostatistics**	8	**Organisation and implementation**

i. HTA methodology and evidence synthesis

Core HTA skills include prioritisation, scoping and PICO formulation, systematic literature searching, critical appraisal, evidence synthesis, interpretation of meta-analyses and communication of findings[120]. Assessors must also be able to prepare and interpret full, rapid and mini-HTA reports[124], consistent with the proportionate approaches described in Chapter 3. Real-world data may be needed when randomised trial evidence is limited or absent[124]. Using those data reliably requires expertise in study design, confounding, missing data and the limits of causal inference[124].

ii. Health economics

Health economics and cost-effectiveness analysis (CEA) are among the highest-rated competencies for HTA experts[121]. The incremental cost-effectiveness ratio compares the additional cost of an intervention with its additional health effect relative to a comparator[127]. Health effects are often measured in quality-adjusted life years (QALYs), combining duration and quality of life[128]. Budget impact analysis (BIA) estimates the financial consequences of adoption using the eligible population, expected treatment mix and associated costs[129]. CEA addresses value for money, while BIA addresses affordability[121]; both are relevant to decisions involving opportunity cost. Teams need to identify and value healthcare costs and resource use, apply discounting, interpret clinical evidence and use decision models and sensitivity analyses to examine uncertainty. Training in these methods can require substantial time, with some capacity-building programmes extending their original schedules[125].

iii. Clinical safety and patient perspective

Clinical assessment requires the ability to find, interpret and apply evidence on effectiveness and safety in routine practice, including adverse events and other patient harms[120]. A recurring weakness is limited expertise in judging relative-effectiveness methods[125]. This matters when a submitted comparator differs from local care or when assessors must choose between adjusted indirect comparisons and naïve comparisons[125]. Results must also be clinically interpretable: interpretability concerns the meaning that can be assigned to a measured score or change in score [130].

Patient safety was the second-highest priority in a Delphi study of AIHT assessment topics, with 95.65% agreement[126]. Patient involvement also helps determine which outcomes matter. Patient-relevant measures cover mortality and morbidity as well as patients' experience, preferences, needs and functioning[130]. People living with a condition should contribute to selecting meaningful endpoints. The minimal important difference describes the smallest change that patients regard as important[130]. Quality standards for patient involvement recommend training HTA staff, reviewers and committee members; providing a named staff contact to support patient participation; and documenting how patient input influenced conclusions and decisions[131].

iv. Epidemiology and biostatistics

HTA experts rank epidemiology third and biostatistics fifth among 16 relevant subjects[121]. Training programmes have addressed gaps in epidemiological methods, including regression, confounding and survival analysis[125]. Assessors must verify disease incidence, prevalence, progression and the numbers of undiagnosed or undertreated patients[129], because these determine the population eligible for treatment and influence budget impact[129]. They must also assess statistical uncertainty and the reliability of subgroup, sensitivity, post hoc and multiple-testing analyses[132]. Survival extrapolation is another important skill: economic evaluations may project outcomes beyond trial follow-up, and the choice of parametric model can materially affect the results[133].

v. AI, data science and health informatics

Digital and AI-enabled technologies require expertise not always present in conventional HTA teams[122], drawing on methods also used to develop AI systems[134]. Assessors should understand the quality and representativeness of training and validation datasets, whether a model is suitable for its intended task, and how its accuracy and bias have been evaluated[134]. Discrimination and calibration are key measures for predictive models[135]. Assessing their limitations may require knowledge of machine learning and deep-learning methods[70]. Experts identified model accuracy and data bias among the ten most important AIHT assessment topics[126], while traditional HTA practitioners may lack familiarity with model training and learning processes[136],[137]. Data programmers, engineers and cybersecurity specialists were therefore included among the technical experts in the Delphi study[126].

vi. Ethics, equity and law

This area examines whether an AIHT meets ethical and human-rights principles, avoids discriminatory effects and complies with the laws governing its use[55],[120],[126],[138]. Ethics was the highest-rated domain in one AIHT Delphi study, with data bias a prominent concern[126]. Assessors need to identify the ethical issues raised by a technology and apply suitable frameworks[120], including principlism, IEEE standards or Value Sensitive Design[138]. They must understand what these approaches can establish and where their limitations lie. Equity assessment requires examining group representation in datasets, sampling bias and differences in performance or consequences across populations[55],[126],[138]. Legal assessment requires interpreting requirements for data governance, human oversight and, where applicable, fundamental-rights impact assessment for high-risk systems[55],[120],[126],[138]. Teams therefore need access to legal expertise capable of reviewing compliance arrangements and liability in the intended care setting.

vii. Organisation and implementation

Organisational effects have received less attention in AIHT assessment: one expert panel gave the strongest agreement in this domain to process-related costs, at 75.56%[126], while HTA experts ranked management near the bottom of 16 competency subjects[121]. Yet assessors must understand how technologies affect workflows, care pathways, management and implementation[120]. Organisational sustainability also includes affordability, resource use and workforce planning, responding to decision-makers' need for evidence on staff workload and performance[124]. Published organisational assessments have relied mainly on literature reviews, with some case studies, interviews, surveys, observation and document analysis. Findings are often reported narratively[139], while some agencies address organisational effects only indirectly through clinical or economic evaluation[140]. Competence in qualitative research, observational study designs, workflow mapping, change management and resource analysis is therefore important.

viii. Environmental sustainability

Environmental sustainability is an emerging area of HTA practice. Agencies in the Netherlands, the UK and Canada have begun to incorporate environmental considerations, and HTAi has established a working group on the subject[124]. Some HTA training programmes have not yet covered it[120]. An AIHT Delphi panel did not reach consensus on environmental topics, which its authors attributed partly to limited experience. The same study cited an estimate of approximately 300,000 kg of carbon dioxide equivalent for training one large language model[126]. WHO guidance calls for AI systems that minimise environmental impacts and improve energy efficiency[138].

Environmental life cycle assessment (eLCA) is a principal quantitative method for examining these effects[141]. It defines a functional unit, such as a patient visit or a population served over time, and accounts for impacts across production, use, maintenance and disposal. The associated life cycle inventory draws on measured activity data and established background databases[142],[143],[144].

Contribution, sensitivity and Monte Carlo analyses help identify the main sources of environmental impact and compare them with potential savings, such as avoided patient and staff travel[143],[144]. Results may feed into multi-criteria appraisal or be valued using a carbon shadow price in economic evaluation[141],[142]. Assessment teams therefore need specialist expertise in hardware materials, energy use, greenhouse-gas accounting, activity-data collection, life cycle databases, uncertainty analysis and valuation of environmental effects.

5.3

Institutional deployment

Four deployment phases

1 **Establish the mandate**

2 **Set methods and procedures**

3 **Test through pilot assessments**

4 **Embed in routine decisions**

Establishing an HTA unit takes more than a single legal or administrative decision. Institutional development typically proceeds from the mandate and legal basis to organisational arrangements, assessment and appraisal procedures, and monitoring of the HTA function itself[12],[88]. Most systems begin with some existing activity. Reviewing available expertise, data, funding and links to reimbursement helps identify a suitable model[12]. Initial studies may survey health-economics capacity in academic institutions[145], examine the policy and political context[74],[146] or combine a needs assessment with a SWOT analysis[125]. Development partners have also assessed political commitment, institutional capacity and financing to identify where support would be effective[85]. Roadmaps are generally developed with government, stakeholders and technical partners[145],[146]. This blueprint groups the work into four phases, each ending with a verifiable milestone.

Phase 1

Phase 1 establishes the mandate and links the future HTA function to those responsible for reimbursement and procurement. Decision-makers should be involved from the outset, since perceived loss of authority can undermine engagement. Stakeholder mapping should cover evidence producers, those who make and implement decisions, and the groups affected by them. The team should also identify available data on healthcare use, prices, populations and disease patterns, along with their custodians. Before choosing the operating structure, it should assess risks concerning data access, financial and political support, organisational culture, stakeholder acceptance and coordination, and agree actions to address them. Funding must be secured.

Established HTA agencies have reported annual budgets of approximately 0.01–0.06% of health or statutory insurance expenditure, funded through dedicated appropriations or submission fees[12]. The mandate should define responsibilities, review procedures, decision criteria, appeal rights and how assessment findings inform policy. A temporary taskforce or collaborating departments can coordinate the initial work and agree roles, processes and timelines[146]. For an embedded unit, an existing public institution may act as the focal point and work with universities, especially where academic expertise has limited direct access to decision-makers[74]. Phase 1 ends when the mandate is formally adopted.

Phase 2

Phase 2 establishes written methods, operating procedures and an appraisal committee. Evidence assessment should remain distinct from appraisal, with conflict-of-interest safeguards in place before the first pilot, including in small or embedded units. Established appraisal committees have between 9 and 39 members from several disciplines and disclose their members' names, affiliations and interests[12]. National guidance serves different needs as capacity develops: initially raising awareness of HTA, then supporting consistency and timely decisions. Procedures can be tested in pilot assessments before formal publication[147].

Decision criteria can be drawn from national policy documents and refined through stakeholder consultation[146]. Where local capacity is limited, an economic reference case may be adapted from a comparable jurisdiction[125]. A conflict-of-interest policy and public access to methods, evidence, models and reasons for decisions help establish trust[145]. Workforce development must address individual skills, organisational processes and system-wide capability[125]. It may combine academic partnerships, in-country training[145] and exchanges with experienced agencies[125]. Training schedules should allow for substantial learning needs: some programmes in health economics and epidemiology have needed twice their planned duration[125].

Phase 3

Phase 3 tests the new arrangements through pilot or shadow assessments before full implementation. These exercises examine methods, procedures, data, staff capacity and the route from evidence to policy[125],[145]. They are common outputs of international capacity-building programmes[74]. Starting with a manageable set of topics, such as high-budget-impact technologies or disinvestment opportunities, can demonstrate the value of HTA early[146]. A modest unit that grows with resources and demand is often more sustainable than establishing a large agency immediately[12]. Administrative delays and coordination problems should be tracked and managed[74], with an interim review to identify weaknesses before formal launch[12].

Phase 4

Phase 4 embeds HTA in routine decision-making and has no fixed end date. Lasting institutionalisation requires political support, stakeholder participation and scientific independence, not only trained staff and completed pilots[74]. The assessment agenda should expand beyond initial priorities as referral and funding pathways become established. Where appropriate, coverage can be made dependent on an HTA recommendation[146].

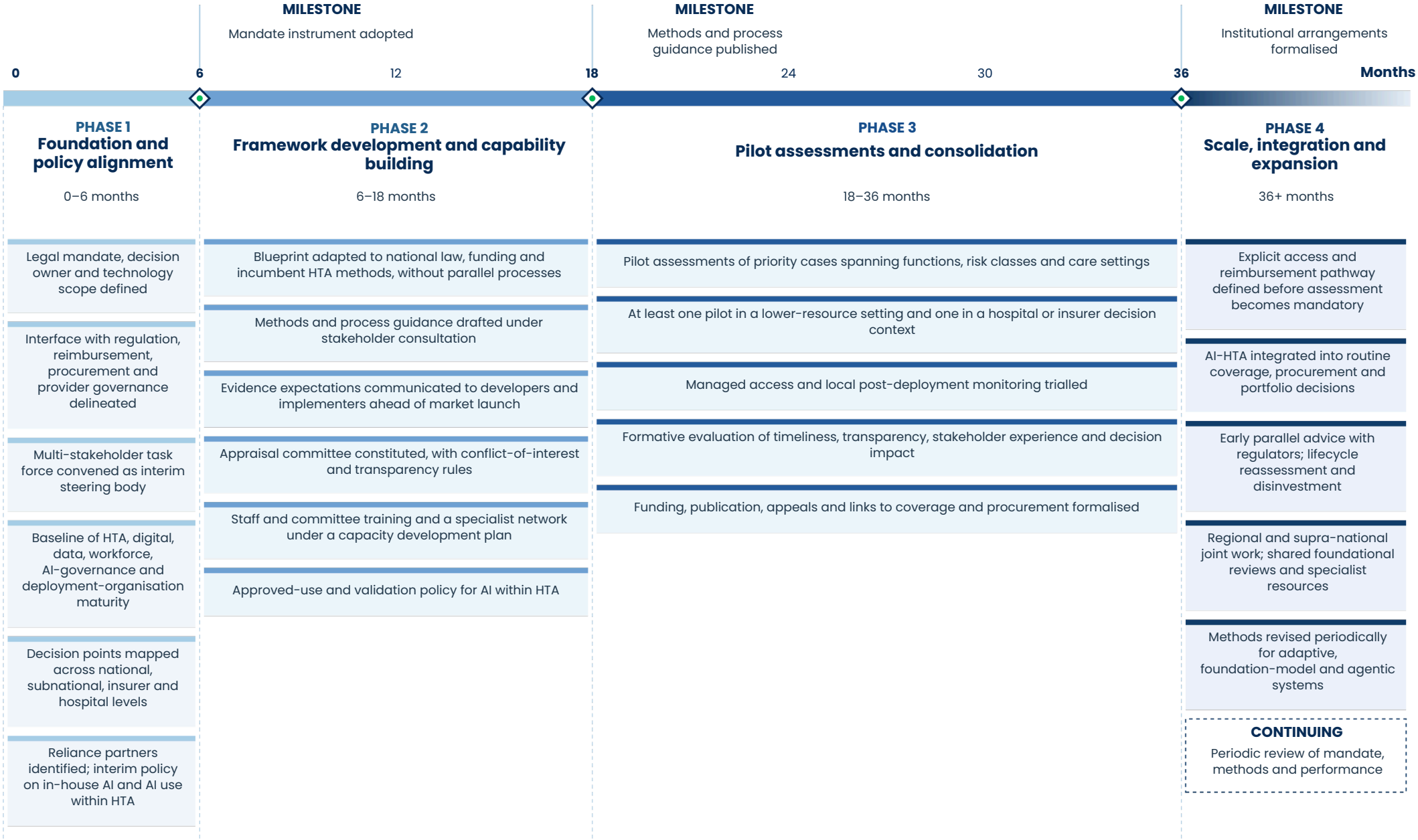
Economic criteria have less influence when they are formally recognised but optional in law[125]. Existing recommendations should be reviewed as costs, epidemiology and other evidence change[12], with additional reassessment when the triggers in Section 2.5.2 are met. Such reviews can support disinvestment. One benefit-package review projected annual savings of USD 147 million by removing inappropriate use of costly medicines and devices without detectable loss of health benefit[74].

Each phase should begin once the previous milestone is met, with intermediate targets used to assess progress over several years[12]. The baseline established in Phase 1 also supports the subsequent evaluation of the HTA unit under Section 5.4. Because institutional weaknesses and staffing shortages are common obstacles[88], stakeholder engagement and training must continue throughout implementation. Timelines should remain flexible: the scale of training needs may become clear only as staff begin assessments[125]. Figure 19 summarises the phases, milestones and activities over an indicative period of 36 months and beyond.



Figure 19:

Tiered AI-HTA unit implementation roadmap in months from inception



Each milestone conditions entry to the following phase. Stakeholder engagement, workforce development and monitoring against the Phase 1 baseline continue throughout.

5.4

Institutional performance monitoring and review

HTA units use resources that might otherwise support patient care[148]. They should therefore demonstrate that their own activities justify their cost[149]. This requires monitoring progress over time and evaluating whether the unit achieves its wider objectives. Monitoring tracks activities and results against agreed indicators; evaluation judges the programme's overall contribution, ideally considering what would have happened without it[12]. Established approaches examine the number, timeliness and quality of assessments; independence, transparency and stakeholder involvement; effects on policy, practice, spending and health; institutional credibility; and the efficiency with which resources are used[12],[148],[150],[151]. These dimensions remain relevant to AI-HTA. The main additional requirements concern methodological quality, including whether the assessed model version, local transferability and post-deployment drift are addressed, and whether decisions translate into appropriate conditions for use. Standard measures of timeliness and procedural quality need little change.

Indicators should reflect the unit's mandate and reasons for being established. A function created to control spending, for example, should examine changes in expenditure after its recommendations are implemented. One intended to improve access should track the time taken to assess technologies, make coverage decisions and implement them. Baseline information collected when the unit is established allows progress to be compared over time and helps assess whether observed changes may be attributable to its work[12].

Figure 20 groups indicators into seven areas. Six address the performance of HTA organisations: production and timeliness, methodological quality, decision influence, operational efficiency, stakeholder participation and institutional learning[12],[148],[150]. The seventh, economic contribution, considers whether the function represents value for money. The indicators are a proposed synthesis, not a mandatory minimum set. Some are drawn from published practice; others are adapted for AIHT assessment. They measure the performance of the HTA unit, not the performance of individual technologies, which is assessed and monitored under Chapter 2. For example, the frequency of model drift is a technology-level indicator, while the proportion of recommendations specifying drift thresholds and reassessment dates measures the quality of the unit's work. Similarly, stakeholder participation and organisational equity are distinct from the distributional effects of an individual technology. Figure 19 also distinguishes indicators available from routine administrative records, such as submission and decision registers, from those needing surveys, interviews, document review or modelling[12],[151].

Figure 20:
Indicators of AI-HTA unit performance

	APPLICABLE TO ANY HTA UNIT		SPECIFIC TO AI-ENABLED TECHNOLOGIES	
	Routine records	Additional evaluation	Routine records	Additional evaluation
Production and timelines	Assessments completed, by assessment type and AI system class Median time from intake to recommendation, by assessment type Assessments outstanding beyond the target interval Median time from finalisation to publication, and completed assessments made publicly available where publication is permitted	–	Median time from a notified model change or reassessment trigger to the revised ruling	–
Methodological quality	–	External review findings and material corrections after publication	Rulings recording the appraised model version, its locked or adaptive status and the changes that reopen the decision Assessments comparing the AI-enabled pathway with the human pathway, beyond standalone accuracy Assessments examining performance on local or external data and across subgroups	–
Decision influence	Recommendations considered by the decision-maker, with documented response and stated reasons	Graded influence on decisions, from minimal to major Clinical guidelines, policy documents and research priorities citing or incorporating the assessments of the unit Change in coverage, procurement or use following recommendations, against a counterfactual	Conditional rulings whose terms (version, oversight, monitoring indicators, reassessment date) enter the funding, procurement or managed access instrument	–
Operational efficiency and reliance	Assessments drawing on external HTA or regulatory work, by component reused Staff time per assessment, by assessment type and AI system class	Duplicated effort avoided, and defects later found in reused evidence	Reliance decisions with a documented version-equivalence and currency check	

	APPLICABLE TO ANY HTA UNIT		SPECIFIC TO AI-ENABLED TECHNOLOGIES	
	Routine records	Additional evaluation	Routine records	Additional evaluation
Equity and stakeholder participation	Patient, clinician and end-user involvement in scoping and appraisal, by group Assessment topics prioritised by burden of disease and disparities in access among underserved groups Declarations of interest by assessors and appraisal members completed before appraisal, and identified conflicts with documented management Appeals lodged, upheld, and resolved within the set term	–	Assessments examining training-data representativeness and subgroup error where the ethical domain is activated	–
Institutional learning	–	Concordance of appraised with realised performance, uptake and cost at reassessment –	Rulings specifying drift indicators, intervention thresholds and a reassessment date Reassessments initiated on trigger and concluded within the term set in the ruling Method revisions prompted by reassessment findings or by new system classes	–
Economic contribution	Annual operating cost, fixed and per assessment by type	Gross savings, expenditure avoided and resources redirected, each reported separately Net health benefit attributable to informed decisions, against an explicit counterfactual and with its uncertainty	–	Resources released by restriction or withdrawal at reassessment, valued at their alternative use



HTA can affect knowledge, policy and practice in ways that are not captured by adoption rates.

The proportion of recommendations followed by decision-makers is often used to measure HTA influence[151],[152], but it is not sufficient on its own. HTA can affect knowledge, policy and practice in ways that are not captured by adoption rates[153]. Its influence may be symbolic, reinforcing a position already taken; conceptual, improving understanding; or instrumental, changing policy or care delivery[150]. A recommendation may reasonably be rejected because of budget constraints, new evidence or other considerations within the decision-maker's remit. Research on the adoption of new interventional procedures found that decisions were never based solely on HTA findings[152]. Conversely, following a recommendation does not guarantee sound decision-making or realised benefit. One hospital-financing body made decisions that diverged from HTA advice in both directions, permitting some technologies that were not recommended and restricting others more than advised[153]. Performance review should therefore document whether an assessment informed the decision, what was decided and why, and how implementation progressed. Influence can extend from awareness and understanding to policy change, clinical practice and patient outcomes. At each successive level, more factors affect the result and attribution to HTA becomes harder[152]. For conditional decisions, evaluation should also check whether the approved model version, oversight requirements, monitoring indicators and reassessment date were reflected in the funding, procurement or managed access agreement (Tool 5).

The unit should record its total operating expenditure from the outset, distinguishing fixed organisational costs from the cost of individual assessments[154]. These figures describe resource use and productivity, but not whether the unit produces sufficient value. Evaluating that value requires examining improvements in the health system attributable to the decisions it informs and comparing them with the cost of the HTA function[155]. One measure is net health benefit: the additional health generated by a decision minus the health that could have been produced by using the same resources elsewhere[155]. A technology adds value only if its benefits exceed the health displaced by its additional costs[154].

Lower expenditure alone does not establish that an HTA programme adds value.

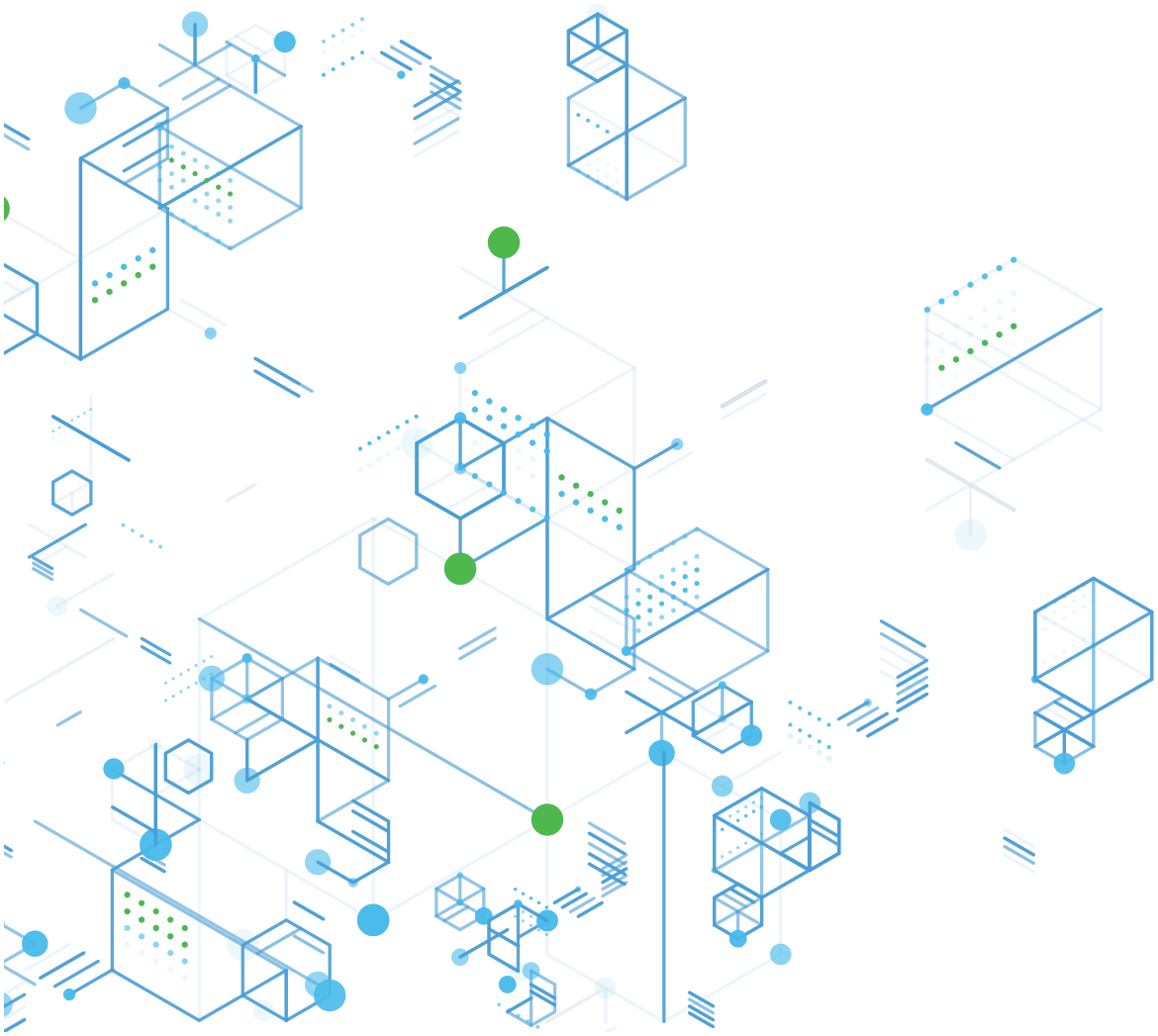
Gross financial savings, avoided future spending and resources redirected to other services should be reported separately. Lower expenditure alone does not establish that an HTA programme adds value[154]. Some programme audits have documented reductions in the use and cost of overprovided technologies without establishing whether spending was contained overall or redirected to higher-value care[153]. HTA generally influences how a fixed budget is spent rather than the size of that budget[155]. Resources released therefore matter chiefly because of the health they can generate elsewhere. Estimating the benefit attributable to HTA requires comparison with a plausible counterfactual: what adoption and outcomes would have occurred without the assessment function[154]. Historical uptake trends, international benchmarks or synthetic controls may provide such comparisons; assuming there would otherwise have been no adoption is rarely sufficient[155]. Studies from low- and middle-income settings illustrate these approaches[154],[156]. In India, a retrospective assessment of HTAIn estimated the net health benefit of three positive recommendations. It converted incremental adoption costs into the health forgone elsewhere using a national cost-effectiveness threshold, then scaled benefits to the population reached, and compared them with agency and assessment costs expressed in QALYs. Estimated returns ranged from 5:1 to 40:1 at observed implementation levels[154]. Such estimates often rely on strong assumptions, including full attribution of benefit to HTA or an uncertain counterfactual, and should not be generalised without qualification. Results should be expressed in measures meaningful to the jurisdiction, such as return on investment or cost–benefit analysis[154].

For AIHT assessment, attributable value also depends on whether the technology is used under the conditions and with the model version appraised. Reassessment can add value by restricting or withdrawing systems whose performance has declined, releasing resources for other care. Avoiding duplicate assessments also saves resources[154], particularly when trusted external evidence is reused as described in Section 3.2.

Indicators and targets should be reviewed regularly as the unit gains experience and its mandate evolves[12]. Findings should inform improvements in methods, evidence reuse, assessment timelines and scientific quality.



Conclusion and Looking Ahead



Conclusion

Digital and artificial intelligence-enabled health technologies present new opportunities for healthcare, but their value cannot be assumed from their technical capabilities. Whether they improve care depends on the populations they serve, the decisions they support, the resources they require and the conditions under which they are used. Their capacity to change over time also means that evidence supporting initial adoption may not remain valid throughout their lifecycle.

The central argument of this blueprint is that these challenges do not require a new discipline of HTA, but a more adaptive application of its established principles and methods. Assessment must remain comparative, evidence-based and focused on the decisions facing health systems, while accounting for the distinctive characteristics of digital and AI-enabled technologies.



The central argument of this blueprint is that these challenges do not require a new discipline of HTA, but a more adaptive application of its established principles and methods.

The five chapters translate this principle into a practical approach:

- Chapter 1 establishes the rationale and methodological principles, including the need to assess AI-enabled interventions within their care pathways rather than as isolated algorithms.
- Chapter 2 provides a five-stage process, from decision framing and evidence assessment to appraisal, monitoring and reassessment. It distinguishes technical performance from clinical utility, incorporates economic value and opportunity cost, and links recommendations to explicit conditions for adoption and continued use.
- Chapter 3 extends this approach to health systems with different levels of HTA maturity and resources. It shows how proportionate methods, reliance on qualified external assessments and international collaboration can reduce duplication without compromising essential assessment requirements. It also distinguishes the capacity to conduct HTA from the AI governance readiness of the organisation that will deploy the technology.
- Chapter 4 provides practical tools to support consistent application.
- Chapter 5 sets out the institutional mandates, organisational models, workforce capabilities and implementation steps needed to establish an effective HTA function.

Together, these contributions address two closely related challenges: determining whether a technology offers sufficient comparative value, and ensuring that the conditions needed to realise and sustain that value are in place. The first requires sound evidence and transparent appraisal. The second requires clear responsibilities, implementation capability, monitoring and the ability to revise decisions as technologies and healthcare needs change.

The blueprint is designed to be applicable across health systems without prescribing a single institutional model or assessment method. Its common principles provide consistency, while its proportional approach recognises differences in resources, governance and local priorities. Ultimately, the success of a HTA unit should be judged not by the number of assessments produced, but by its contribution to better decisions, improved health outcomes and more effective use of healthcare resources.

Looking ahead: Extending the Blueprint



This blueprint is intended as a living document. Its methods, tools and examples will be updated as DHT and AIHT assessment practices, technologies, evidence standards and policy requirements evolve. Future editions will draw on new research, experience from implementation and contributions from the HTA community.

We invite members of the HealthAI and HTAi community, including policymakers, assessors, regulators, healthcare providers, developers, researchers and patient representatives, to comment on this edition, share practical experience and suggest topics or country cases for future development. Such contributions will help prioritise updates and keep the blueprint useful as the field evolves.

As for now, three priorities are proposed below.

1. Country Framework Cards: learning from international practice

A future chapter will examine how different jurisdictions assess digital and AI-enabled technologies and translate their findings into reimbursement, procurement and adoption decisions. Building on the initial cases developed in the current blueprint, including France, Germany, the Republic of Korea, the United Kingdom and Brazil, the chapter will provide a structured comparison of institutional mandates, assessment methods, evidence requirements, decision pathways and implementation arrangements.

The purpose is not to identify a single preferred national model, but to understand which approaches work under different conditions and which elements can be adapted or reused elsewhere. The framework cards will provide a growing, referenced body of practical experience, supported by country expertise and updated as national policies and assessment practices evolve.

2. Making continuous and periodic monitoring operational

Chapter 2 sets out what to monitor, indicative review intervals and when reassessment is required. The next challenge is to make those arrangements workable across institutions. A future module will set out a proportionate framework for sharing monitoring responsibilities, distinguishing continuous or automated checks for time-sensitive technical and safety signals from periodic reviews of clinical outcomes, equity, costs and continued value.

It will define the complementary responsibilities of four groups:

- Manufacturers and developers: maintain version and change records; monitor technical performance and safety; report material changes and emerging risks; and provide data and documentation needed for independent review.
- Healthcare providers and deployers: collect local evidence on use, patient outcomes, safety, workflow and costs; check performance across relevant groups; respond to alerts; and escalate material findings.
- Sub-national, regional and hospital HTA units: interpret evidence in the local care context, coordinate monitoring across sites, identify changes in clinical or economic value, and refer significant findings for wider reassessment.
- National HTA units: establish proportionate monitoring expectations and common indicators, review evidence across settings, initiate or coordinate reassessment within their mandate, and advise decision-makers on continuation, restriction or withdrawal.

The framework will specify the five PPTO components: people, processes, policies, technologies/infrastructure and operations. This includes named owners and skills; data-quality checks, reporting intervals and escalation pathways; enforceable access, notification and data-sharing arrangements; secure, interoperable records and monitoring systems; and sustained funding, support and capacity for corrective action. It will also clarify how HTA review connects with regulatory vigilance, payer decisions and provider governance. The aim is to establish who collects, validates and reviews each signal, who is authorised to act on it and how the work is maintained over time.

3. Responsible use of AI within HTA: a complementary area for future development

While this edition focuses on the HTA of digital and AI-enabled technologies, AI is also beginning to change how HTA itself is conducted. Applications in horizon scanning, literature screening, evidence extraction, real-world data analysis, economic modelling and report preparation may improve the efficiency and timeliness of assessments.

These opportunities raise a separate set of methodological and governance questions. Future work should define which HTA tasks can benefit from AI assistance, what evidence is needed to validate its use, and which safeguards are required to maintain scientific quality, reproducibility, transparency and accountability. A risk-based approach should distinguish routine administrative support from analytical tasks and high-stakes scientific judgements, for which human responsibility remains essential.



References

[1] World Health Organization. Global strategy on digital health 2020-2025. Geneva: World Health Organization; 2021.

[2] World Health Organization. Classification of digital interventions, services and applications in health: a shared language to describe the uses of digital technology for health. 2nd ed. Geneva: World Health Organization; 2023.

[3] 15 U.S. Code § 9401. Definitions [Internet]. Ithaca (NY): Legal Information Institute, Cornell Law School; [cited 2026 Aug 26]. Available from: https://www.law.cornell.edu/uscode/text/15/9401

[4] U.S. Food and Drug Administration, Center for Devices and Radiological Health, Center for Biologics Evaluation and Research, Center for Drug Evaluation and Research, Office of Combination Products. Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations. Draft Guidance for Industry and Food and Drug Administration Staff. Silver Spring (MD): U.S. Food and Drug Administration; 2025 [cited 2026 Aug 26]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/artificial-intelligence-enabled-device-software-functions-lifecycle-management-and-marketing

[5] Smith A, Severn M. An overview of continuous learning artificial intelligence-enabled medical devices. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; 2022 May. (CADTH Horizon Scan; vol. 2, no. 5). ISSN: 2563-6596.

[6] U.S. Food and Drug Administration, Center for Devices and Radiological Health. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. Silver Spring (MD): U.S. Food and Drug Administration; 2021.

[7] Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. Lancet Digit Health. 2021;3(3):e195-e203. doi:10.1016/S2589-7500(20)30292-2.

[8] Moor M, Banerjee O, Shakeri Hossein Abad Z, Krumholz HM, Leskovec J, Topol EJ, Rajpurkar P. Foundation models for generalist medical artificial intelligence. Nature. 2023;616(7956):259-265. doi:10.1038/s41586-023-05881-4

[9] World Health Organization. The World Health Assembly endorses the extension of the Global Strategy on Digital Health to 2027 and approves the next phase for 2028–2033 [Internet]. Geneva: World Health Organization; 2025 May 23 [cited 2026 Aug 26]. Available from: https://www.who.int/news/item/23-05-2025-world-health-assembly-endorses-extension-of-the-global-digital-health-strategy-to-2027

[10] IQVIA Institute for Human Data Science. Digital Health Trends 2021: Innovation, Evidence, Regulation, and Adoption. July 2021.

[11] O'Rourke B, Oortwijn W, Schuller T, International Joint Task Group. The new definition of health technology assessment: a milestone in international collaboration. Int J Technol Assess Health Care. 2020;36:187-190. doi:10.1017/S0266462320000215.

[12] Bertram M, Dhaene G, Tan-Torres Edejer T, editors. Institutionalizing health technology assessment mechanisms: a how to guide. Geneva: World Health Organization; 2021.

[13] Ofori-Asenso R, Hallgreen CE, De Bruin ML. Improving interactions between health technology assessment bodies and regulatory agencies: a systematic review and cross-sectional survey on processes, progress, outcomes, and challenges. Front Med. 2020;7:582634. doi:10.3389/fmed.2020.582634.

[14] European Parliament, Council of the European Union. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU. Off J Eur Union. 2021 Dec 22;L458:1-32. Available from: http://data.europa.eu/eli/reg/2021/2282/oj

[15] Bélisle-Pipon JC, Couture V, Roy MC, Ganache I, Goetghebeur M, Cohen IG. What Makes Artificial Intelligence Exceptional in Health Technology Assessment? Front Artif Intell. 2021 Nov 2;4:736697. doi: 10.3389/frai.2021.736697.

[16] Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med. 2021 Jul 15;385(3):283–286. doi:10.1056/NEJMc2104626.

[17] Unsworth H, Wolfram V, Dillon B, Salmon M, Greaves F, Liu X, et al. Building an evidence standards framework for artificial intelligence-enabled digital health technologies. Lancet Digit Health. 2022 Jun;4(6):e390-e391. doi:10.1016/S2589-7500(22)00078-4

[18] U.S. Food and Drug Administration, Health Canada, Medicines and Healthcare products Regulatory Agency. Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices: Guiding Principles. Silver Spring (MD): FDA; 2023 Oct [cited 2026 Aug 29]. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/predetermined-change-control-plans-machine-learning-enabled-medical-devices-guiding-principles

[19] U.S. Food and Drug Administration, Center for Devices and Radiological Health, Center for Biologics Evaluation and Research, Center for Drug Evaluation and Research, Office of Combination Products. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions: Guidance for Industry and Food and Drug Administration Staff. Rockville (MD): FDA; 2025 Aug 18 [cited 2026 Aug 29]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence

[20] Jaksa A, Arena PJ, Hanisch M, Marsico M. Use of real-world evidence in health technology reassessments across 6 health technology assessment agencies. Value Health. 2025;28(6):898–906.

[21] Park SH, Choi J-I, Fournier L, Vasey B. Randomized clinical trials of artificial intelligence in medicine: why, when, and how? Korean J Radiol. 2022;23(12):1119–1125. doi:10.3348/kjr.2022.0834.

[22] Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ, SPIRIT-AI and CONSORT-AI Working Group, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 2020 Oct;26(10):1351–1363.doi:10.1038/s41591-020-1037-7.

[23] Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health. 2020 Oct;2(10):e537–e548. doi:10.1016/S2589-7500(20)30218-1.

[24] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022 May;28(5):924–933. doi:10.1038/s41591-022-01772-9.

[25] Tranchellini F, Farag Y, Jutzeler C, Meegahapola L. Evaluating deep learning sepsis prediction models in ICUs under distribution shift: a multi-centre retrospective cohort study. npj Digit Med. 2026;9:306. doi:10.1038/s41746-026-02364-4.

[26] Ciecierski-Holmes T, Singh R, Axt M, Brenner S, Barteit S. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. npj Digit Med. 2022;5:162. doi:10.1038/s41746-022-00700-y.

[27] Khoshravan Azar A, Draghi B, Rotalinti Y, Myles P, Tucker A. Addressing 6 challenges in generative AI for digital health: a scoping review. PLOS Digit Health. 2024;3(5):e0000503. doi:10.1371/journal.pdig.0000503.

[28] Slattery P, Saeri AK, Grundy EAC, Graham J, Noetel M, Uuk R, et al. The AI risk repository: a meta-review, database, and taxonomy of risks from artificial intelligence. Patterns. 2026;7(5):101517. doi:10.1016/j.patter.2026.101517.

[29] van Kessel R, Schmidt J, Winitsky S, Wharton G, Carter AW, Guo C, et al. Critical gaps in health technology assessment guidance for generative AI across six health systems. NPJ Digit Med. 2026 Oct 8. doi:10.1038/s41746-026-03365-z.

[30] Segur-Ferrer J, Moltó-Puigmartí C, Pastells-Peiró R, Vivanco-Hidalgo RM. Methodological Frameworks and Dimensions to Be Considered in Digital Health Technology Assessment: Scoping Review and Thematic Analysis. J Med Internet Res. 2024 Apr 10;26:e48694. doi:10.2196/48694. PMID:38598288.

[31] National Institute for Health and Care Excellence. Evidence standards framework for digital health technologies [Internet]. London: National Institute for Health and Care Excellence; 2018 [updated 2022 Aug 9; cited 2026 Aug 31]. Available from: https://www.nice.org.uk/corporate/ecd7

[32] Dietz J, Garrett Z, Ayiku L, Battaglia L, Lipska I, Oortwijn W, et al. Domains in health technology assessment of digital health technologies: a mapping review. Value Health. 2026 Jul 14. doi:10.1016/j.jval.2026.06.017.

[33] European Commission. ASSESS-DHT: development and harmonisation of methodologies for assessing digital health technologies in Europe [Internet]. Brussels: European Commission; 2023 [cited 2026 Aug 30]. Available from: https://cordis.europa.eu/project/id/101137347. doi:10.3030/101137347

[34] Fleurence RL, Bian J, Wang X, Xu H, Dawoud D, Higashi M, et al. Generative artificial intelligence for health technology assessment: opportunities, challenges, and policy considerations: an ISPOR Working Group report. Value Health. 2025;28(2):175–183. doi:10.1016/j.jval.2024.10.3846.

[35] Drummond MF, Schwartz JS, Jönsson B, Luce BR, Neumann PJ, Siebert U, et al. Key principles for the improved conduct of health technology assessments for resource allocation decisions. Int J Technol Assess Health Care. 2008;24(3):244-258. doi:10.1017/S0266462308080343.

[36] Walker S, Sculpher M, Claxton K, Palmer S. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. Value Health. 2012;15(3):570-579. doi:10.1016/j.jval.2011.12.013.

[37] Klemp M, Frønsdal KB, Facey K; HTAi Policy Forum. What principles should govern the use of managed entry agreements? Int J Technol Assess Health Care. 2011;27(1):77-83. doi:10.1017/S0266462310001297.

[38] Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. BMJ. 2004;329(7459):224-227. doi:10.1136/bmj.329.7459.224.

[39] European Parliament, Council of the European Union. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. Off J Eur Union. 2017 May 5;L117:1-175. Available from: http://data.europa.eu/eli/reg/2017/745/oj

[40] Trowman R, Migliore A, Ollendorf DA. Health technology assessment 2025 and beyond: lifecycle approaches to promote engagement and efficiency in health technology assessment. Int J Technol Assess Health Care. 2023;39(1):e15. doi:10.1017/S0266462323000090.

[41] Pichler FB, Boysen M, Mittmann N, Gilardino R, Bruce A, Bond K, Vreman RA, Largeron N, Banhazi J, Ollendorf DA, Jain M, Upadhyaya S, Goettsch WG. An operationalization framework for lifecycle health technology assessment: an HTAi Global Policy Forum Task Force report. Int J Technol Assess Health Care. 2024;40(1):e45. doi:10.1017/S0266462324000199.

[42] Member State Coordination Group on Health Technology Assessment (HTA CG). Guidance on the scoping process. V1.0. 2024 Nov 13. Available from: https://health.ec.europa.eu/publications/guidance-scoping-process_en

[43] International Medical Device Regulators Forum. "Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations. IMDRF/SaMD WG/N12FINAL:2014. 2014 Sep 18.

[44] Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices—Modeling Studies. Value Health. 2003;6(1):9-17. doi:10.1046/j.1524-4733.2003.00234.x.

[45] National Institute for Health and Care Excellence. NICE technology appraisal and highly specialised technologies guidance: the manual. London: National Institute for Health and Care Excellence; 2022. Last updated 2026 Mar 31.

[46] Richardson M, Sander B, Daneman N, Mighton C, Miller FA. Moving from intervention management to disease management: a qualitative study exploring a systems approach to health technology assessment in Canada. Int J Technol Assess Health Care. 2023;39(1):e67. doi:10.1017/S0266462323002696.

[47] Richardson M, Ramsay LC, Bielecki JM, Berta W, Sander B. Systems thinking in health technology assessment: a scoping review. Int J Technol Assess Health Care. 2021;37:e71:1-8. doi:10.1017/S0266462321000428.

[48] Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Evaluating artificial intelligence in medicine: phases of clinical research. JAMIA Open. 2020;3(3):326-331. doi:10.1093/jamiaopen/ooaa033.

[49] Alami H, Lehoux P, Auclair Y, de Guise M, Gagnon MP, Shaw J, et al. Artificial intelligence and health technology assessment: anticipating a new level of complexity. J Med Internet Res. 2020;22(7):e17707. doi:10.2196/17707.

[50] Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. 4th ed. Ottawa: CADTH; 2017 Mar.

[51] Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 4th ed. Oxford: Oxford University Press; 2015.

[52] Santos AR, von Hafe F, Sampaio F, Londral AR, Perelman J. Mapping methodologies for economic evaluation of digital health technologies: a scoping review. Value Health. 2026;29(2):311-323. doi:10.1016/j.jval.2025.09.012.

[53] Campione A, Zetzsche M, Rombey T, Rödiger H, Henschke C. Health economic evaluations of digital health technologies—a rapid review of applied methods. Front Digit Health. 2026;8:1816757. doi:10.3389/fdgth.2026.1816757.

[54] Tarricone R, Petracca F, Weller HM. Towards harmonizing assessment and reimbursement of digital medical devices in the EU through mutual learning. npj Digit Med. 2024;7(1):268. doi:10.1038/s41746-024-01263-w.

[55] European Parliament, Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Off J Eur Union. 2024 Jul 12;L 2024/1689:1-144. Available from: http://data.europa.eu/eli/reg/2024/1689/oj

[56] Gilbert S, Pimenta A, Stratton-Powell A, Welzel C, Melvin T. Continuous improvement of digital health applications linked to real-world performance monitoring: safe moving targets? Mayo Clin Proc Digit Health. 2023;1(3):276-287. doi:10.1016/j.mcpdig.2023.05.010.

[57] Feng J, Phillips RV, Malenica I, Bishara A, Hubbard AE, Celi LA, et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. npj Digit Med. 2022;5(1):66. doi:10.1038/s41746-022-00611-y.

[58] Rothery C, Claxton K, Palmer S, Epstein D, Tarricone R, Sculpher M. Characterising uncertainty in the assessment of medical devices and determining future research needs. Health Econ. 2017;26(Suppl 1):109-123. doi:10.1002/hec.3467.

[59] Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. J Am Med Inform Assoc. 2017;24(6):1052-1061. doi:10.1093/jamia/ocx030.

[60] Brönneke JB, Herr A, Reif S, Stern AD. Dynamic HTA for digital health solutions: opportunities and challenges for patient-centered evaluation. Int J Technol Assess Health Care. 2023;39(1):e72,1-11. doi:10.1017/S0266462323002726.

[61] Nemzoff C, Ruiz F, Chalkidou K, Mehndiratta A, Guinness L, Cluzeau F, et al. Adaptive health technology assessment to facilitate priority setting in low- and middle-income countries. BMJ Glob Health. 2021 Apr;6(4):e004549. doi:10.1136/bmjgh-2020-004549.

[62] Wilkinson T, Sculpher MJ, Claxton K, Revill P, Briggs A, Cairns JA, et al. The International Decision Support Initiative Reference Case for Economic Evaluation: an aid to thought. Value Health. 2016 Dec;19(8):921-928. doi:10.1016/j.jval.2016.04.015.

[63] Mbau R, Vassall A, Gilson L, Barasa E. Factors influencing the institutionalization of health technology assessment: a scoping literature review. Health Syst Reform. 2023;9(3):2360315. doi:10.1080/23288604.2024.2360315.

[64] World Health Organization. 2015 Global Survey on Health Technology Assessment by National Authorities: Main findings. Geneva: World Health Organization; 2015.

[65] John D, Dutta Majumdar A, Pillai RN, Khatoon S, Bhattacharya P, Mukherjee N, et al. Health technology assessment for digital health technologies in India: a framework for action. Int J Technol Assess Health Care. 2024;40(1):e70,1-9. doi:10.1017/S0266462324000345.

[66] Welte R, Feenstra T, Jager H, Leidl R. A decision chart for assessing and improving the transferability of economic evaluation results between countries. Pharmacoeconomics. 2004;22(13):857-876.

[67] World Health Organization. Good reliance practices in the regulation of medical products: high level principles and considerations. In: WHO Expert Committee on Specifications for Pharmaceutical Preparations: fifty-fifth report. Geneva: World Health Organization; 2021:Annex 10. (WHO Technical Report Series, No. 1033).

[68] Yang WW, Juan YC, Wu GHM, Pwu RF. The critical intersect of regulations, health technology assessment, and drug safety assessments. Drug Saf. 2024 Apr;47(4):289-299. doi:10.1007/s40264-023-01386-1.

[69] Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13. BMJ. 2017;359:j4530. doi:10.1136/bmj.j4530.

[70] Farah L, Borget I, Martelli N, Vallee A. Suitability of the current health technology assessment of innovative artificial intelligence-based medical devices: scoping literature review. J Med Internet Res. 2024;26:e51514. doi:10.2196/51514.

[71] Rödiger H, Wittich LFM, Jeindl R, Hidaka Y, Bowie D, Rejón-Parrilla JC, et al. Frameworks for assessing digital health technologies: a scoping review. Int J Technol Assess Health Care. 2026;42(1):e24. doi:10.1017/S0266462326103523.

[72] Cresswell K, Claire R, Issa L, Kubehl L, Bilal O, Purcell R, et al. Best practices and gaps in current regulatory and health technology assessment real-world evidence policies for medicines and medical devices: current state of play and next steps. Clin Transl Sci. 2026;19:e70507. doi:10.1111/cts.70507.

[73] Vogler S, Knoll V, Sonemann A, Quanten A, Adams R, de Jager T, et al. Collaborating for affordable patient access to medicines: review of the first decade of the Beneluxa Initiative in pharmaceutical policy. Appl Health Econ Health Policy. 2026. doi:10.1007/s40258-026-01057-z.

[74] Tantivess S, Chalkidou K, Tritasavit N, Teerawattananon Y. Health technology assessment capacity development in low- and middle-income countries: experiences from the international units of HITAP and NICE. F1000Res. 2017;6:2119. doi:10.12688/f1000research.13180.1.

[75] Teerawattananon Y, Luz K, Yothasmutra C, Pwu RF, Ahn J, Shafie AA, et al. Historical development of the HTAsiaLink network and its key determinants of success. Int J Technol Assess Health Care. 2018;34(3):260-266. doi:10.1017/S0266462318000223.

[76] Itaborahy A, Santos VCC, Salomon FCR, Bregman C, Mateus I, Santos M. Hospital-based health technology assessment (HTA) in Latin America and the Caribbean: a scoping review. Rev Panam Salud Publica. 2025;49:e8. doi:10.26633/RPSP.2025.8.

[77] Member State Coordination Group on Health Technology Assessment (HTA CG). Procedural guidance for JCA medicinal products. V1.0. 2024 Nov 13 [cited 2026 Sep 13]. Available from: https://health.ec.europa.eu/publications/procedural-guidance-jca-medicinal-products_en

[78] Member State Coordination Group on Health Technology Assessment (HTA CG). Annual Work Programme 2026. 2025 Nov 28 [cited 2026 Sep 13]. Available from: https://health.ec.europa.eu/publications/2026-work-programme-member-state-coordination-group-hta-htacg_en

[79] Halmesmäki E, Pasternack I, Roine R. Hospital-based health technology assessment (HTA) in Finland: a case study on collaboration between hospitals and the national HTA unit. Health Res Policy Syst. 2016;14:25. doi:10.1186/s12961-016-0095-2.

[80] European Commission. Adopting Hospital Based Health Technology Assessment in EU (ADHOPHTA): Final report summary. Brussels: European Commission; 2016 [cited 2026 Sep 13]. Available from: https://cordis.europa.eu/project/id/305018/reporting

[81] Sarma KV, Harmon S, Sanford T, Roth HR, Xu Z, Tetreault J, et al. Federated learning improves site performance in multicenter deep learning without data sharing. J Am Med Inform Assoc. 2021;28(6):1259-1264. doi:10.1093/jamia/ocaa341.

[82] Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. npj Digit Med. 2020;3:119. doi:10.1038/s41746-020-00323-1.

[83] Sharma M, Teerawattananon Y, Dabak SV, Isaranuwatchai W, Pearce F, Pilasant S, et al. A landscape analysis of health technology assessment capacity in the Association of South-East Asian Nations region. Health Res Policy Syst. 2021;19(1):19. doi:10.1186/s12961-020-00647-0.

[84] Dai Z, Cheung EKY, Wang T, Bishai DM, Si L, Yip BHK, et al. Comparative review of selected health technology assessment agencies: features and insights for health technology assessment development and improvement. Pharmacoecon Policy. 2026;2(2):60-68. doi:10.1016/j.pharp.2026.06.007.

[85] Li R, Hernandez-Villafuerte K, Towse A, Vlad I, Chalkidou K. Mapping priority setting in health in 17 countries across Asia, Latin America, and sub-Saharan Africa. Health Syst Reform. 2016;2(1):71-83. doi:10.1080/23288604.2015.1123338.

[86] Németh B, Kaló Z. European cooperation in health technology assessment implementation: the perspective of Central and Eastern European countries. J Comp Eff Res. 2020;9(9):599-602. doi:10.2217/cer-2020-0062.

[87] Leech AA, Kim DD, Cohen JT, Neumann PJ. Use and misuse of cost-effectiveness analysis thresholds in low- and middle-income countries: trends in cost-per-DALY studies. Value Health. 2018;21(7):759-761. doi:10.1016/j.jval.2017.12.016.

[88] Mirelman AJ, Goel K, Tan-Torres Edejer T. The global landscape of country-level health technology assessment processes: a survey among 104 countries. Health Policy Open. 2025;8:100138. doi:10.1016/j.hpopen.2025.100138.

[89] Hussein R, Zink A, Ramadan B, Howard FM, Hightower M, Shah S, et al. Advancing healthcare AI governance through a comprehensive maturity model based on systematic review. npj Digit Med. 2026;9(1):236. doi:10.1038/s41746-026-02418-7.

[90] Elvidge J, Hawksworth C, Avşar TS, Zemplenyi A, Chalkidou A, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence (CHEERS-AI). Value Health. 2024;27(9):1196-1205. doi:10.1016/j.jval.2024.05.006.

[91] Fontrier AM, Visintin E, Kanavos P. Similarities and Differences in Health Technology Assessment Systems and Implications for Coverage Decisions: Evidence from 32 Countries. PharmacoEconomics Open. 2022;6(3):315-328. doi:10.1007/s41669-021-00311-5.

[92] Drummond MF, Augustovski F, Bhattacharyya D, Campbell J, Chaiyakunapruk N, Chen Y, et al. Challenges of Health Technology Assessment in Pluralistic Healthcare Systems: An ISPOR Council Report. Value Health. 2022;25(8):1257-1267. doi:10.1016/j.jval.2022.02.006.

[93] Fontrier AM, Kanavos P. Do reimbursement recommendations by the Canadian Agency for Drugs and Technology in Health translate into coverage decisions for orphan drugs in the Canadian province of Ontario? Value Health. 2023;26(7):1011-1021. doi:10.1016/j.jval.2023.02.013.

[94] Ciani O, Tarricone R, Torbica A. Diffusion and use of health technology assessment in policy making: what lessons for decentralised healthcare systems? Health Policy. 2012;108(2-3):194-202. doi:10.1016/j.healthpol.2012.09.017.

[95] Fan VY, Mehndiratta A, Ahazie J, Guzman J, Prinja S, Sundararaman T, et al. Organizational formation for priority setting: historical perspectives and thematic analysis of India's Health Technology Assessment Agency. Health Syst Reform. 2023;9(3):2327414. doi:10.1080/23288604.2024.2327414.

[96] Nasser M, Sawicki PT. Institute for Quality and Efficiency in Health Care: Germany. New York (NY): The Commonwealth Fund; 2009 Jul. Issue Brief No. 57.

[97] Institute for Quality and Efficiency in Health Care. General methods. Version 3.0. Cologne: Institute for Quality and Efficiency in Health Care; 2008 May 27.

[98] Weill C, Banta D. Development of health technology assessment in France. Int J Technol Assess Health Care. 2009 Jul;25 Suppl 1:108-111. doi:10.1017/S0266462309090503.

[99] National Institute for Health and Care Excellence. NICE HealthTech programme manual: NICE process and methods [Internet]. London: National Institute for Health and Care Excellence; 2025 Jul 14 [updated 2025 Dec 17; cited 2026 Oct 3]. Available from: https://www.nice.org.uk/process/pmg48

[100]Novaes HMD, Soárez PC. Health technology assessment (HTA) organizations: dimensions of the institutional and political framework. Cad Saude Publica. 2016;32 Suppl 2:e00022315. doi:10.1590/0102-311X00022315.

[101] Shafie AA, Chandriah H, Yong YV, Puteh SEW. Health technology assessment and its use in drug policy in Malaysia. Value Health Reg Issues. 2019;18:145-150. doi:10.1016/j.vhri.2019.03.003.

[102] Lertpitakpong C, Chaikledkaew U, Thavorncharoensap M, Tantivess S, Praditsitthikorn N, Youngkong S, et al. A determination of topics for health technology assessment in Thailand: making decision makers involved. J Med Assoc Thai. 2008;91 Suppl 2:S100-109.

[103] Ministry of Health, Singapore. Medical technologies evaluation methods and process guide [Internet]. Version 3.0. Singapore: Ministry of Health; 2024 Mar [cited 2026 Oct 4]. Available from: https://isomer-user-content.by.gov.sg/68/53adb002-9122-496b-8e44-394bd4a8a935/ace-mte-methods-and-process-guide-(version-3-01).pdf

[104] Lee HY, Nguyen TTT, Park S, Hoang VM, Kim WH. Health technology assessment development in Vietnam: a qualitative study of current progress, barriers, facilitators, and future strategies. Int J Environ Res Public Health. 2021;18(16):8846. doi:10.3390/ijerph18168846.

[105] Babigumira JB, Jenny AM, Bartlein R, Stergachis A, Garrison LP Jr. Health technology assessment in low- and middle-income countries: a landscape assessment. J Pharm Health Serv Res. 2016;7(1):37-42. doi:10.1111/jphs.12120.

[106] Raval D, Parmar D, Saha S, Sarkar R, Wadhwa M, Pandya A, et al. Cost-effectiveness analysis of AI-assisted chest X-ray interpretation tools for TB screening: a rapid HTA. Front Digit Health. 2025;7:1629127. doi:10.3389/fdgth.2025.1629127.

[107] De Pietro C, Camenzind P, Sturny I, Crivelli L, Edwards-Garavoglia S, Spranger A, Wittenbecher F, Quentin W. Switzerland: health system review. Health Syst Transit. 2015;17(4):1-288.

[108] Swiss Federal Chancellery. Ausserparlamentarische Kommissionen, Leitungsorgane und Bundesvertretungen [Internet]. Bern: Swiss Confederation; [cited 2026 Oct 5]. Available from: https://www.admin.ch/de/ausserparlamentarische-kommissionen

[109] Medical Services Advisory Committee. About us [Internet]. Canberra: Australian Government Department of Health and Aged Care; 2024 Oct 22 [cited 2026 Oct 5]. Available from: https://www.msac.gov.au/about-us

[110] Hailey D. The history of health technology assessment in Australia. Int J Technol Assess Health Care. 2009;25(Suppl 1):61-67. doi:10.1017/S0266462309090436.

[111] Bundesamt für Gesundheit. Eidgenössische Kommission für allgemeine Leistungen und Grundsatzfragen (ELGK) [Internet]. Bern: Bundesamt für Gesundheit; [cited 2026 Oct 5]. Available from: https://www.bag.admin.ch/de/eidgenoessische-kommission-fuer-allgemeine-leistungen-und-grundsatzfragen-elgk

[112] Brügger U, Horisberger B, Ruckstuhl A, Plessow R, Eichler K, Gratwohl A. Health technology assessment in Switzerland: a descriptive analysis of "Coverage with Evidence Development" decisions from 1996 to 2013. BMJ Open. 2015;5:e007021. doi:10.1136/bmjopen-2014-007021.

[113] Australian Government Department of Health and Aged Care. Pharmaceutical Benefits Advisory Committee (PBAC) membership [Internet]. Canberra: Australian Government Department of Health and Aged Care; 2026 Oct 1 [cited 2026 Oct 5]. Available from: https://www.pbs.gov.au/industry/listing/participants/pbac

[114] Australian Government Department of Health and Aged Care. PBAC outcomes explained [Internet]. Canberra: Australian Government Department of Health and Aged Care; 2015 Jul 22 [cited 2026 Oct 5]. Available from: https://pbs.gov.au/industry/listing/elements/pbac-meetings/pbac-outcomes-explained

[115] Kim H, Byrnes J, Goodall S. Health technology assessment in Australia: the Pharmaceutical Benefits Advisory Committee and Medical Services Advisory Committee. Value Health Reg Issues. 2021;24:6-11. doi:10.1016/j.vhri.2020.09.001.

[116] Schweizerische Gesellschaft der Vertrauens- und Versicherungsärzte. HTA für neue Leistungen in der OKP nach KVG [HTA for new services in statutory health insurance under the Health Insurance Act] [Internet]. Winterthur: Schweizerische Gesellschaft der Vertrauens- und Versicherungsärzte; [cited 2026 Oct 5]. Available from: https://www.vertrauensaerzte.ch/manual/4/hta/okp/

[117] Sampietro-Colom L, Lach K, Cicchetti A, Kidholm K, Pasternack I, Fure B, Rosenmöller M, Wild C, Kahveci R, Wasserfallen JB, Kiivet RA, et al. The AdHopHTA handbook: a handbook of hospital-based Health Technology Assessment (HB-HTA). Public deliverable. The AdHopHTA Project (FP7/2007-13 grant agreement No. 305018); 2015. Available from: http://www.adhophta.eu/handbook

[118] Gagnon MP, Desmartis M, Poder T, Witteman W. Effects and repercussions of local/hospital-based health technology assessment (HTA): a systematic review. Syst Rev. 2014;3:129. doi:10.1186/2046-4053-3-129.

[119] Di Bidino R, Lipska I, Kukla M, von Pinoci M, Papavero SC, Marchetti M, Sampietro-Colom L, Cicchetti A. The potential of the hospital-based Health Technology Assessment: results of a world-wide survey. Int J Technol Assess Health Care. 2025;41(1):e19:1-8. doi:10.1017/S0266462325000108.

[120] Mueller D, Gutierrez-Ibarluzea I, Chiumente M, Oortwijn W. Toward a common understanding of competencies for health technology assessment: enhancing educational and training programs around the globe. Int J Technol Assess Health Care. 2021;37:e29:1-10. doi:10.1017/S0266462320001919.

[121] Wu YS, Chen C, Wang LC, Jian LS, Ko Y. Talent cultivation in health technology assessment: an expert survey. BMC Med Educ. 2022;22:157. doi:10.1186/s12909-022-03214-z.

[122] Warraich HJ, Tazbaz T, Califf RM. FDA perspective on the regulation of artificial intelligence in health care and biomedicine. JAMA. 2025;333(3):241-247. doi:10.1001/jama.2024.21451.

[123] Nong P, Adler-Milstein J, Apathy NC, Holmgren AJ, Everson J. Current use and evaluation of artificial intelligence and predictive models in US hospitals. Health Aff (Millwood). 2025;44(1):90-98. doi:10.1377/hlthaff.2024.00842.

[124] van Mun L, Kievit W, Oortwijn W. Competencies for health technology assessment professionals: a scoping review and content analysis. Int J Technol Assess Health Care. 2026;42(1):e54:1-8. doi:10.1017/S0266462326103808.

[125] Abraham K, Kvamme I, Magrin Sammut S, de Vries S, Formosa T, Dupree R, et al. A blueprint for health technology assessment capacity building: lessons learned from Malta. Int J Technol Assess Health Care. 2024;40(1):e11:1-6. doi:10.1017/S0266462324000072.

[126] Di Bidino R, Daugbjerg S, Papavero SC, Haraldsen IH, Cicchetti A, Sacchini D. Health technology assessment framework for artificial intelligence-based technologies. Int J Technol Assess Health Care. 2024;40(1):e61:1-9. doi:10.1017/S0266462324000308.

[127] York Health Economics Consortium. Incremental cost-effectiveness ratio (ICER) [Internet]. York (UK): York Health Economics Consortium; 2016 Oct [updated 2025 Oct; cited 2026 Oct 7]. Available from: https://www.yhec.co.uk/glossary-term/incremental-cost-effectiveness-ratio-icer/

[128] National Institute for Health and Care Excellence. Glossary: Q [Internet]. [cited 2026 Oct 7]. Available from: https://www.nice.org.uk/glossary?letter=q

[129] Sullivan SD, Mauskopf JA, Augustovski F, Caro JJ, Lee KM, Minchin M, et al. Budget impact analysis—principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value Health. 2014;17(1):5-14. doi:10.1016/j.jval.2013.08.2291.

[130] Member State Coordination Group on Health Technology Assessment (HTA CG). Guidance on outcomes for joint clinical assessments. Brussels: European Commission; 2024 Jun 10. Available from: https://health.ec.europa.eu/publications/guidance-outcomes-joint-clinical-assessments_en

[131] HTAi Interest Sub-Group for Patient/Citizen Involvement in HTA. Values and quality standards for patient involvement in HTA [Internet]. Health Technology Assessment international; 2014 Jun [cited 2026 Oct 7]. Available from: https://htai.org/interest-group/patient-citizen-involvement/

[132] Member State Coordination Group on Health Technology Assessment (HTA CG). Guidance on reporting requirements for multiplicity issues and subgroup, sensitivity and post hoc analyses in joint clinical assessments. 2024 Jun 13. Available from: https://health.ec.europa.eu/publications/guidance-reporting-requirements-multiplicity-issues-and-subgroup-sensitivity-and-post-hoc-analyses_en

[133] Latimer NR. Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. Med Decis Making. 2013;33(6):743-754. doi:10.1177/0272989X12472398.

[134] Russell RG, Novak LL, Patel M, Garvey KV, Craig KJT, Jackson GP, et al. Competencies for the use of artificial intelligence-based tools by health care professionals. Acad Med. 2023;98(3):348-356. doi:10.1097/ACM.0000000000004963.

[135] de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med. 2022;5:2. doi:10.1038/s41746-021-00549-7.

[136] Tachkov K, Zemplenyi A, Kamusheva M, Dimitrova M, Siirtola P, Pontén J, et al. Barriers to use artificial intelligence methodologies in health technology assessment in Central and East European countries. Front Public Health. 2022;10:921226. doi:10.3389/fpubh.2022.921226.

[137] Zemplényi A, Tachkov K, Balkanyi L, Németh B, Petykó ZI, Petrova G, et al. Recommendations to overcome barriers to the use of artificial intelligence-driven evidence in health technology assessment. Front Public Health. 2023;11:1088121. doi:10.3389/fpubh.2023.1088121.

[138] World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021 Jun 28. Available from: https://www.who.int/publications/i/item/9789240029200

[139] Lee A, Sinding LS. A review of organisational and patient-related assessments in HTAs published by INAHTA members. Copenhagen: National Board of Health, Danish Centre for Health Technology Assessment; 2007. Danish Health Technology Assessment 2007;9(2). Available from: https://www.sdu.dk/da/om-sdu/institutter-centre/cast/forskning/castrapdownload/pub2007

[140] Mathy C, Pascal C, Bongiovanni-Delarozière I, Favez L. Proposing a health-specific organizational impact framework to evaluate organizational impacts in health technology assessments. Int J Technol Assess Health Care. 2023;39(1):e60:1-8. doi:10.1017/S0266462323000508.

[141] Pegg M, Gregg E, Bishop E, Bouttell J. Methods for evaluating environmental sustainability in health technology assessment: views of stakeholders. Environ Res Health. 2026;4(1):015009. doi:10.1088/2752-5309/ae476f.

[142] Di Bidino R, Dutta Majumdar A, Pegg M, Mahon R, Papavero SC, Mueller D. Implications of incorporating environmental sustainability into health technology assessment for digital health technologies. Value Health. 2026;29(4):595-599. doi:10.1016/j.jval.2026.01.019.

[143] Tarpani RRZ, Gallego-Schmid A. Environmental impacts of a digital health and well-being service in elderly living schemes. Clean Environ Syst. 2024;12:100161. doi:10.1016/j.cesys.2023.100161.

[144] van Bree EM, Snijder LE, Ossebaard HC, Brakema EA. Environmental impact of physical visits and telemedicine in nursing care at home: comparative life cycle assessment. J Med Internet Res. 2025;27:e67538. doi:10.2196/67538.

[145] Downey LE, Mehndiratta A, Grover A, Gauba V, Sheikh K, Prinja S, et al. Institutionalising health technology assessment: establishing the Medical Technology Assessment Board in India. BMJ Glob Health. 2017;2:e000259. doi:10.1136/bmjgh-2016-000259.

[146] Bijlmakers L, Egeli P, Al Saeedi AI, Sadoon B, Richter D, Oortwijn W. Road toward institutionalizing health technology assessment in the Emirate of Abu Dhabi: the role of evidence-informed deliberative processes. Int J Technol Assess Health Care. 2024;40(1):e80:1-6. doi:10.1017/S0266462324004744.

[147] Botwright S, Sittimart M, Chavarina KK, Bayani DBS, Merlin T, Surgey G, et al. Good practices for health technology assessment guideline development: a report of the Health Technology Assessment International, HTAsiaLink, and ISPOR Special Task Force. Int J Technol Assess Health Care. 2024;40(1):e74:1-14. doi:10.1017/S0266462324004719.

[148] Millar R, Morton A, Bufali MV, Engels S, Dabak SV, Isaranuwatchai W, et al. Assessing the performance of health technology assessment (HTA) agencies: developing a multi-country, multi-stakeholder, and multi-dimensional framework to explore mechanisms of impact. Cost Eff Resour Alloc. 2021;19:37. doi:10.1186/s12962-021-00290-8.

[149] Trowman R, Migliore A, Ollendorf DA. The value and impact of health technology assessment: discussions and recommendations from the 2023 Health Technology Assessment International Global Policy Forum. Int J Technol Assess Health Care. 2023;39(1):e75. doi:10.1017/S0266462323002763.

[150] Lafortune L, Farand L, Mondou I, Sicotte C, Battista R. Assessing the performance of health technology assessment organizations: a framework. Int J Technol Assess Health Care. 2008;24(1):76-86. doi:10.1017/S0266462307080105.

[151] Berndt N, Schuller T. HTA impact assessment study. Part I: Practices of HTA impact assessment in INAHTA member agencies [Internet]. Edmonton (AB): International Network of Agencies for Health Technology Assessment; 2020 Feb [cited 2026 Oct 13]. Available from: https://www.inahta.org/hta-tools-resources/hta-impact-influence/

[152] International Network of Agencies for Health Technology Assessment (INAHTA). Published evidence on the influence of health technology assessment: a systematic review [Internet]. Edmonton (AB): INAHTA; 2014 Sep [cited 2026 Oct 14]. Available from: https://www.inahta.org/hta-tools-resources/hta-impact-influence/

[153] Zechmeister I, Schumacher I. The impact of health technology assessment reports on decision making in Austria. Int J Technol Assess Health Care. 2012;28(1):77-84. doi:10.1017/S0266462311000729.

[154] Bahuguna P, Baker PA, Briggs A, Gulliver S, Hesselgreaves H, Mehndiratta A, et al. Is health technology assessment value for money? Estimating the return on investment of health technology assessment in India (HTAIn). BMJ Evid Based Med. 2025;30(Suppl 2):s29-s37. doi:10.1136/bmjebm-2023-112487.

[155] Murphy P, Walker S, Putri S, Sculpher M, Griffin S. What is the value of developing an HTA process? PharmacoEconomics. 2026;44:791-798. doi:10.1007/s40273-026-01613-4.

[156] Kingkaew P, Budtarad N, Khuntha S, Barlow E, Morton A, Isaranuwatchai W, et al. A model-based study to estimate the health and economic impact of health technology assessment in Thailand. Int J Technol Assess Health Care. 2022;38(1):e45. doi:10.1017/S02664623220

HTAi HEALTH TECHNOLOGY ASSESSMENT INTERNATIONAL

HEALTH AI

Driving Responsible AI for a Healthier Future

About HealthAI

HealthAI - The Global Agency for Responsible AI in Health, is a Geneva-based, independent nonprofit driving equitable access to AI-powered health innovations. Collaborating with governments, international organizations, and global health leaders, HealthAI focuses on AI governance and regulation to ensure AI transforms healthcare for all. Positioned at the intersection of healthcare, artificial intelligence, and policy, HealthAI aims to bridge the gap between emerging technologies and accessibility, particularly in low- and middle-income countries. By fostering trust, governance, and collaboration, HealthAI is shaping a future where health technologies benefit every population worldwide.



How HealthAI Can Support Member States

HealthAI can support governments and ministries of health in navigating the implementation gap identified across the regulatory landscape. The organization's established track record in capacity strengthening and regulatory implementation across its Global Regulatory Network provides a foundation for practical engagement with Member States facing similar structural challenges.

HealthAI's support encompasses:

- **Implementation readiness assessments** to identify AI in health governance gaps and priorities
- **Capacity strengthening for regulators and policymakers** through technical assistance and tailored training with a focus on EU AI Act implementation
- **Peer learning facilitation** connecting Member States to share regulatory tools and operational solutions
- **Access to governance resources,** including implementation guidance and validation instruments developed through multi-country collaboration

This approach recognizes that successful implementation requires operational capacity, institutional coordination, and sustained knowledge exchange beyond policy design alone.

HealthAI – The Global Agency for Responsible AI in Health
Rue de Varembé 7, 1202 Geneva, Switzerland
contact@healthai.agency
www.healthai.agency
+41 (0) 22 552 00 34