

Health AI Evaluation Guide

Purpose: This guide brings structure to selection and helps you compare vendors practically and consistently. Think of it as a practical scoring tool to identify top contenders. Evaluation helps you narrow the field.

Instructions:

1. Reference your personalized [Health AI Readiness Assessment](#), including your defined use case, domain maturity level, and readiness score are important filters to the matrix below.
2. **Rate each AI tool** across the core criteria (see table below). Use a **simple 1–5 scale** for each category:
 - 5 = Excellent fit
 - 1 = Major concerns or gaps
3. **Assign optional weights** if certain criteria matter more for your organization.
 - Example: Clinical Relevance might be weighted higher than ROI for a patient safety initiative.
4. **Calculate a total score** (simple or weighted average).
5. **Compare across vendors** to build your shortlist — and spotlight risks or gaps early.



PRO TIP

Involve cross-functional reviewers (clinical, IT, procurement) when scoring to capture a holistic organizational perspective.

DESCRIPTION	SAMPLE ASSESSMENT QUESTIONS	RED FLAGS	SCORE
Clinical relevance & capability			
Whether the solution solves a real clinical need and delivers meaningful, actionable results.	<p>[CMIO]: Does this tool solve a real, pressing clinical problem? Are its outputs clinically meaningful and actionable? Is there peer-reviewed evidence in a comparable real-world clinical setting?</p>	Vendor is unable to furnish validation data from a health system of analogous size and complexity.	Select ▾
<i>Notes:</i>			
Validation, bias & evidence base			
The quality, relevance, and fairness of the evidence base and data used to train and test the tool.	<p>[CDO]: What are the demographics of the training datasets? Has the data been audited for bias?</p> <p>[CMIO]: What were the specific performance metrics (AUC, sensitivity, etc.) in real-world patient cohorts?</p>	<p>Vendor assertions are predicated solely on internal pilot data, lacking longitudinal evidence or peer-reviewed studies.</p> <p>Tool lacks stratified validation data for our core patient populations.</p>	Select ▾
<i>Notes:</i>			
Workflow & IT integration			
The tool's ability to fit into existing clinical workflows and integrate with technical systems.	<p>[Clinical Ops Director]: Does the tool reduce or increase clicks and cognitive burden?</p> <p>[Health IT Lead]: Can this integrate seamlessly with our EHR using standard APIs (FHIR, HL7)? What is the required internal resource lift?</p>	The solution functions as a standalone application and does not embed insights directly within the existing clinical workflow.	Select ▾
<i>Notes:</i>			

DESCRIPTION	SAMPLE ASSESSMENT QUESTIONS	RED FLAGS	SCORE
Governance & transparency			
<p>Whether the tool's logic, data use, and compliance posture are clear, explainable, and legally sound.</p>	<p>[Legal & CMIO]: Is the algorithm's logic explainable, or is it an opaque "black box"?</p> <p>[CDO]: What are the vendor's policies regarding data rights, especially for retraining their models?</p> <p>[CIO]: Is it HIPAA compliant and HITRUST or SOC 2 certified?</p>	<p>Vendor provides ambiguous information regarding data ownership, usage rights for model retraining, or intellectual property.</p> <p>A lack of transparency into model logic and data lineage is a major clinical and legal risk.</p>	<p>Select ▾</p>
<p>Notes:</p>			
ROI & financial sustainability			
<p>The solution must be cost-effective, have a viable financial model, and deliver tangible ROI.</p>	<p>[CFO]: What is the complete TCO? What is the projected ROI based on metrics (e.g., reduced length of stay) we can actually track and validate? Is the vendor's business model stable and viable for a long-term partnership?</p>	<p>The pricing structure is overly complex, or the vendor is unable to provide clear ROI case studies from analogous organizations.</p>	<p>Select ▾</p>
<p>Notes:</p>			

DESCRIPTION	SAMPLE ASSESSMENT QUESTIONS	RED FLAGS	SCORE
User adoption & change management			
Whether the solution will be adopted by staff, and what supports are in place for implementation and training.	<p>[Clinical Ops Director]: Is there a formal change management plan and longitudinal training support?</p> <p>[CMIO]: What processes does the vendor have to incorporate clinician feedback into product improvements post-implementation?</p>	The vendor's training plan consists primarily of a single, one-time event and offers minimal post-implementation support.	Select ▾
<i>Notes:</i>			
Clinical context fit & generalizability			
Whether the tool works for your population—including pediatrics, geriatrics, multilingual, and underserved communities.	<p>[CMIO/Clinical Team]: Has the tool been tested in settings or populations that match ours (e.g., pediatric, geriatric, rural, safety-net, multilingual)?</p> <p>[Data Team]: What subgroup performance data is available? Does the tool use exclusionary criteria in its validation that omit your core patient population?</p>	Tool lacks stratified validation data or was not tested in populations relevant to your system.	Select ▾
<i>Notes:</i>			