Healthcare AI Vendor Evaluation Questions

48 questions for understanding what different enterprise stakeholders need before a complex evaluation can advance.

Confident decisions start before the demo.

How to use this guide
Use the questions to prepare evidence routes, not to imply that every answer belongs on a public website. Separate what can be understood publicly from what should be shared after qualification, under NDA or through testing for the proposed use case.

The four evidence-access levels

Visible publicly: The orientation a stakeholder can review before contacting the company.

Shared after a serious enquiry: Detailed information provided after the use case and stakeholder need are qualified.

Confidential under NDA: Sensitive technical, security, legal or commercial material.

Testing for the proposed use case: Evidence that must be produced or confirmed in the buyer's environment.

Clinical leadership

1. What clinical problem is the product intended to address?

Why it matters	A precise problem definition prevents a promising capability from being evaluated outside the need it was designed to solve.
What useful evidence looks like	A bounded intended-use statement, target condition, current workflow problem and non-goals.
Likely access level	Visible publicly
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

2. Who is the intended user, patient population and care setting?

Why it matters	Performance and safety cannot be interpreted without knowing who uses the product, for whom and under which conditions.
What useful evidence looks like	Named users, target population, care setting, inclusion and exclusion boundaries.
Likely access level	Visible publicly
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

3. What role does the AI play in the clinical workflow?

Why it matters	A stakeholder must understand whether the system informs, prioritises, recommends, automates or replaces part of a decision.
What useful evidence looks like	Workflow diagram, intended inputs and outputs, human review point and action supported.
Likely access level	Visible publicly
Reference	https://www.fda.gov/medical-devices/software-

	medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

4. What evidence supports performance, clinical utility and safe use?

Why it matters	A metric alone does not show whether the product improves a relevant decision or outcome in practice.
What useful evidence looks like	Study summaries, evaluation setting, comparators, endpoints, sample characteristics and limitations.
Likely access level	Visible publicly
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

5. What limitations, failure modes and excluded populations are known?

Why it matters	Explicit limits help teams decide where the product should not be relied upon and what safeguards are required.
What useful evidence looks like	Known biases, failure conditions, confidence ranges, data gaps, contraindications and out-of-scope uses.
Likely access level	Visible publicly
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

6. What happens when the AI output is uncertain, incorrect or unavailable?

Why it matters	A safe workflow needs a fallback route before implementation, not after an incident.
What useful evidence looks like	Human escalation, exception handling, override, downtime process and clinical responsibility.
Likely access level	Shared after a serious enquiry

Reference	https://www.nature.com/articles/s41746-025-01900-y

Governance and compliance

1. Who is accountable for approval, oversight and stopping use?

Why it matters	A decision cannot be governed when authority and accountability remain implicit.
What useful evidence looks like	Named decision owner, oversight body, review cadence, escalation authority and stop criteria.
Likely access level	Shared after a serious enquiry
Reference	https://www.nist.gov/itl/ai-risk-management-framework

2. Which decisions remain with humans?

Why it matters	Human oversight must be described at the point where the system influences care, not only as a general principle.
What useful evidence looks like	Decision-rights map, review obligations, override conditions and prohibited automation.
Likely access level	Visible publicly
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

3. How are bias, fairness and population fit assessed?

Why it matters	A health system needs to know whether performance and risk may vary across populations or settings.
What useful evidence looks like	Subgroup analysis, data representativeness, known gaps, mitigation and local-validation plan.
Likely access level	Shared after a serious enquiry
Reference	https://www.who.int/publications/i/item/9789240029200

4. How are model updates, version changes and new claims governed?

Why it matters	A decision made on one version can become invalid if the product changes without a controlled review.
What useful evidence looks like	Version history, change-control process, notification rules, revalidation triggers and claim approval.
Likely access level	Shared after a serious enquiry
Reference	https://www.nist.gov/itl/ai-risk-management-framework

5. How are incidents, complaints and performance concerns escalated?

Why it matters	Teams need to know how issues become visible, investigated, corrected and communicated.
What useful evidence looks like	Incident taxonomy, reporting route, response times, corrective action and stakeholder notification.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01900-y

6. Which regulatory and claim boundaries apply?

Why it matters	Public language should not imply uses, outcomes or assurances beyond the product's current evidence and status.
What useful evidence looks like	Regulatory status, intended-use boundary, approved claims, jurisdiction and legal review route.
Likely access level	Visible publicly
Reference	https://www.gov.uk/data-ethics-guidance/a-buyers-guide-to-ai-in-health-and-care

Security and privacy

1. What data enters, leaves and remains in the system?

Why it matters	A security review starts with an understandable data flow, not a list of certifications.
What useful evidence looks like	Data-flow diagram, data types, storage locations, transfers, outputs and retention.
Likely access level	Shared after a serious enquiry
Reference	https://hhscyber.hhs.gov/cybersecurity-performance-goals.html

2. Which core security controls protect the service?

Why it matters	Stakeholders need evidence of how access, identity, encryption, logging and resilience are managed.
What useful evidence looks like	Security overview, control mapping, independent assurance and current certification scope.
Likely access level	Shared after a serious enquiry
Reference	https://hhscyber.hhs.gov/cybersecurity-performance-goals.html

3. Is customer data used to train or improve models?

Why it matters	The answer affects consent, privacy, contractual rights, data governance and risk.
What useful evidence looks like	Clear data-use statement, opt-in or opt-out conditions, de-identification and training boundaries.
Likely access level	Visible publicly
Reference	https://hhscyber.hhs.gov/cybersecurity-performance-goals.html

4. How are access, logging, retention and subprocessors managed?

Why it matters	Operational control depends on who can access data, how activity is recorded and where third parties are involved.
What useful evidence looks like	Role-based access, audit logs, retention schedule, deletion process and subprocessor list.
Likely access level	Shared after a serious enquiry
Reference	https://hhscyber.hhs.gov/cybersecurity-performance-goals.html

5. How are vulnerabilities disclosed and patched?

Why it matters	A buyer needs to understand how security defects are found, prioritised, communicated and corrected.
What useful evidence looks like	Vulnerability disclosure policy, patch process, severity model and customer notification.
Likely access level	Visible publicly
Reference	https://hhscyber.hhs.gov/cybersecurity-performance-goals.html

6. What happens during downtime, breach or service disruption?

Why it matters	Continuity and recovery affect patient safety, workflow and organisational exposure.
What useful evidence looks like	Business continuity, disaster recovery, recovery targets, breach response and fallback workflow.
Likely access level	Confidential under NDA
Reference	https://hhscyber.hhs.gov/cybersecurity-performance-goals.html

Technical integration and data

1. Which systems, standards and data formats are required?

Why it matters	Integration feasibility depends on the buyer's actual environment, not a generic compatibility claim.
What useful evidence looks like	Supported systems, APIs, standards, formats, infrastructure and dependency list.
Likely access level	Visible publicly
Reference	https://www.nature.com/articles/s41746-025-01900-y

2. What integration work is required from the buyer and vendor?

Why it matters	Hidden technical labour can change the real cost and feasibility of adoption.
What useful evidence looks like	Responsibility matrix, integration sequence, technical prerequisites and effort assumptions.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01900-y

3. Which data-quality assumptions affect performance?

Why it matters	A model may perform differently when local data is incomplete, delayed, inconsistent or outside its training conditions.
What useful evidence looks like	Required fields, quality thresholds, missing-data handling, data provenance and failure conditions.
Likely access level	Shared after a serious enquiry
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

4. What local acceptance testing is required before go-live?

Why it matters	Local testing connects vendor evidence to the buyer's environment and intended use.
What useful evidence looks like	Acceptance-test plan, sample requirements, pass criteria, owners and release decision.
Likely access level	Testing for the proposed use case
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

5. How will drift, degradation and data changes be monitored?

Why it matters	A safe decision must account for performance after deployment, not only before purchase.
What useful evidence looks like	Monitoring metrics, thresholds, alert ownership, investigation and retraining or rollback rules.
Likely access level	Shared after a serious enquiry
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

6. What technical resources and support are required over time?

Why it matters	The operating burden continues after implementation and should be visible before commitment.
What useful evidence looks like	Support model, staffing assumptions, service hours, maintenance, upgrade and escalation responsibilities.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01900-y

Operations and implementation

1. What changes in the current workflow?

Why it matters	Operational fit depends on how tasks, handoffs, exceptions and accountability change in practice.
What useful evidence looks like	Before-and-after workflow, role changes, handoffs, exception routes and decision points.
Likely access level	Visible publicly
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

2. What new work, training and exception handling will be required?

Why it matters	A productivity claim can hide new review, correction, monitoring or coordination work.
What useful evidence looks like	Training plan, review burden, exception volume assumptions, support and ownership.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01900-y

3. What is the implementation sequence, timeline and ownership model?

Why it matters	A credible transition requires more than a destination. Teams need to see how the change will be made survivable.
What useful evidence looks like	Phases, milestones, dependencies, internal owner, vendor owner and readiness criteria.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01900-y

4. How will a pilot be designed and judged?

Why it matters	A pilot without success, failure and exit criteria can consume resources without producing a decision.
What useful evidence looks like	Pilot question, scope, baseline, measures, responsibilities, success criteria and stop rule.
Likely access level	Testing for the proposed use case
Reference	https://www.nature.com/articles/s41746-025-01767-z

5. What happens when the system does not perform as expected?

Why it matters	Exception and fallback design protects workflow continuity and clarifies responsibility.
What useful evidence looks like	Fallback workflow, manual alternative, issue ownership, communication and correction path.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01900-y

6. How will adoption, feedback and refinement be managed?

Why it matters	A product can technically launch while failing to become usable or accepted in the real workflow.
What useful evidence looks like	Adoption measures, feedback route, review cadence, training refresh and improvement ownership.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01900-y

Finance and procurement

1. What current problem, cost or constrained capacity is the investment intended to change?

Why it matters	A value case should start with the existing condition, not the product's feature list.
What useful evidence looks like	Baseline problem, affected volume, current cost, resource constraint and consequence of doing nothing.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01767-z

2. What costs exist beyond the licence price?

Why it matters	Implementation labour, integration, training, governance, monitoring and switching costs form part of the effective price.
What useful evidence looks like	Total-cost model covering internal labour, technical work, training, support and ongoing oversight.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01767-z

3. Which outcomes will be measured, over what period and by whom?

Why it matters	A measurable value case needs defined outcomes, timing, data sources and accountable owners.
What useful evidence looks like	Outcome definitions, baseline, measurement plan, time horizon and reporting owner.
Likely access level	Testing for the proposed use case
Reference	https://www.nature.com/articles/s41746-025-01767-z

4. What evidence and assumptions support the projected value?

Why it matters	A projection should separate observed evidence from assumptions and buyer-specific estimates.
What useful evidence looks like	Assumption register, source evidence, ranges, dependencies and sensitivity analysis.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01767-z

5. What contract, service, liability and exit conditions apply?

Why it matters	Commercial terms can change operational risk, accountability and the cost of reversing the decision.
What useful evidence looks like	Service levels, support, data return, termination, liability, transition and continuity terms.
Likely access level	Confidential under NDA
Reference	https://www.gov.uk/data-ethics-guidance/a-buyers-guide-to-ai-in-health-and-care

6. What would justify scaling, renegotiating, pausing or stopping?

Why it matters	A decision is more defensible when continuation and exit conditions are agreed before commitment.
What useful evidence looks like	Decision gates, thresholds, review dates, escalation and exit criteria.
Likely access level	Testing for the proposed use case
Reference	https://www.nature.com/articles/s41746-025-01767-z

Executive sponsor

1. How does this opportunity support a current organisational priority?

Why it matters	Strategic relevance determines whether the opportunity deserves scarce attention, resources and political support.
What useful evidence looks like	Named priority, affected objective, executive owner and expected organisational contribution.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01767-z

2. Why is AI the appropriate approach compared with non-AI alternatives?

Why it matters	A defensible decision considers substitutes, process redesign and the option of doing nothing.
What useful evidence looks like	Alternative analysis, comparative advantages, constraints and conditions where AI is not preferred.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01767-z

3. What benefits, risks and opportunity costs are concentrated across the organisation?

Why it matters	Value and burden may fall on different teams, creating resistance even when the overall case looks positive.
What useful evidence looks like	Stakeholder value map, resource demands, risks, dependencies and distribution of accountability.
Likely access level	Shared after a serious enquiry
Reference	https://www.nist.gov/itl/ai-risk-management-framework

4. Who owns the outcome and which resources are committed?

Why it matters	Executive sponsorship is meaningful only when decision rights, resources and accountability are explicit.
What useful evidence looks like	Named sponsor, operating owner, budget, staffing and governance commitments.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01767-z

5. What evidence would justify advancing, pausing or rejecting the opportunity?

Why it matters	Predefined decision criteria reduce ambiguity and prevent momentum from replacing evaluation.
What useful evidence looks like	Decision criteria, required evidence, risk tolerance, stage gates and stop conditions.
Likely access level	Testing for the proposed use case
Reference	https://www.nature.com/articles/s41746-025-01767-z

6. What must leadership monitor after implementation?

Why it matters	Leadership still needs visibility into value, safety, adoption and material changes after go-live.
What useful evidence looks like	Executive dashboard, review cadence, thresholds, incident reporting and strategic reassessment.
Likely access level	Shared after a serious enquiry
Reference	https://www.nist.gov/itl/ai-risk-management-framework

Internal champion

1. Which stakeholders must support the decision and what does each need to understand?

Why it matters	The champion cannot carry one generic case into a decision made by several functions.
What useful evidence looks like	Stakeholder map, decision task, evidence need, concern and approval role.
Likely access level	Visible publicly
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

2. Which evidence can be shared publicly, after qualification, under NDA or through local testing?

Why it matters	A clear access path prevents the champion from repeatedly requesting or reconstructing information.
What useful evidence looks like	Evidence register with access level, owner, request route and freshness date.
Likely access level	Shared after a serious enquiry
Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

3. What do the strongest public claims establish and where are their boundaries?

Why it matters	A champion needs language that is persuasive without asking colleagues to accept more than the evidence supports.
What useful evidence looks like	Claim-to-evidence map, context, limitations, applicable setting and approved wording.
Likely access level	Visible publicly

Reference	https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

4. Which unanswered questions currently prevent the next step?

Why it matters	A useful case separates what is known from what must still be resolved.
What useful evidence looks like	Open-question register with owner, evidence required, decision affected and due stage.
Likely access level	Shared after a serious enquiry
Reference	https://www.gov.uk/data-ethics-guidance/a-buyers-guide-to-ai-in-health-and-care

5. Which cross-functional meetings are required and what decision must each produce?

Why it matters	Meetings become useful when participants, evidence inputs, questions, output and decision gate are defined.
What useful evidence looks like	Meeting designs covering participants, inputs, questions, required output and next decision.
Likely access level	Shared after a serious enquiry
Reference	https://www.nature.com/articles/s41746-025-01900-y

6. Can the case be circulated without rebuilding it manually?

Why it matters	A decision slows when the champion must translate scattered pages into a new narrative for every stakeholder.
What useful evidence looks like	A concise decision pack with stakeholder routes, evidence boundaries, open questions and source register.
Likely access level	Visible publicly

Reference	https://www.gov.uk/data-ethics-guidance/a-buyers-guide-to-ai-in-health-and-care

Source library

These sources inform the question architecture. This toolkit does not replace clinical, legal, security, procurement, regulatory or organisation-specific review.

- UK Government, A buyer's guide to AI in health and care: https://www.gov.uk/data-ethics-guidance/a-buyers-guide-to-ai-in-health-and-care
- FDA, Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles: https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles
- NIST, AI Risk Management Framework: https://www.nist.gov/itl/ai-risk-management-framework
- WHO, Ethics and governance of artificial intelligence for health: https://www.who.int/publications/i/item/9789240029200
- FAIR-AI, A practical framework for appropriate implementation and review of AI in healthcare: https://www.nature.com/articles/s41746-025-01900-y
- Early pipeline framework for assessing vendor AI solutions to support return on investment: https://www.nature.com/articles/s41746-025-01767-z
- DECIDE-AI reporting guideline: https://www.bmj.com/content/377/bmj-2022-070904
- ONC HTI-1 Final Rule: https://healthit.gov/regulations/hti-rules/hti-1-final-rule/
- HHS Healthcare and Public Health Cybersecurity Performance Goals: https://hhscyber.hhs.gov/cybersecurity-performance-goals.html
- FDA, Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cybersecurity-medical-devices-quality-system-considerations-and-content-premarket-submissions
- LinkedIn and Bain, The Principles of Buyability: https://www.linkedin.com/business/marketing/blog/research-and-insights/the-principles-of-buyability-why-strong-deals-stall-and-what-separates-the-vendors-who-get-chosen
- Forrester, 2026 State of Business Buying: https://www.forrester.com/press-newsroom/forrester-2026-the-state-of-business-buying/