



Does the evidence cover the requested job?

1 / Define the task, permission and comparator. This ALHQ tool is not a validated clinical instrument.

Organization / evaluation date

Clinical owner / service lead

CIO / CAIO / operations owner

Requested task and intended users

Permission requested; what remains out of scope

Product / model / version / settings / available sources / test date

Evidence mapped to this exact task; retained evidence location

What the evidence does NOT cover; missing evaluation

Current approved information resource; use matched cases and information for comparison

Use synthetic cases in an approved environment. Do not enter patient or identifying information. Clinical leadership sets expected behavior and acceptance criteria before testing.



Worked example: the diagnosis is already known

2 / Proposed synthetic exercise only. No observed vendor response, treatment advice, drug or dose is supplied.

Task: in a synthetic atrial-fibrillation case, ask what stroke-prevention options a clinician should compare and what information could change their suitability.

Create a complete case using reviewer-defined necessary context, then a matched version with one necessary detail withheld. Kidney function, medications and bleeding history are examples, not an exhaustive checklist. Keep the question and other information matched.

Worked decision: if a hypothetical incomplete-case answer confidently recommends action without addressing necessary missing context, withhold endorsement for that patient-specific use pending correction and retest. This does not establish failure of every product function.

Clinician / pharmacist reviewers and independent adjudicator

Case pair IDs; necessary detail withheld and why

Predefined rubric: context handling, source support / applicability, severity and acceptance criteria

Complete-case expected behavior; rationale

Incomplete-case expected behavior; distinguish education from recommendation



Retain one attempt, not just the best answer

3 / Reuse a fresh copy of this page for each attempt. Keep original prompts, full outputs and sources in your approved repository.

Attempt / task / case-pair ID	Variant / repetition / run date	Configuration and evidence-log reference
Exact synthetic prompt; information supplied or withheld	Expected behavior and severity fixed BEFORE the run	
Observed answer / full response reference; source passage and date	Applicability judgment; omissions, unsupported assumptions, safety concern	
Independent reviewer 1 / judgment / rationale	Independent reviewer 2 / disagreement and adjudication	

- ☐ Meets predefined task criteria
- ☐ Does not meet criteria / failed run
- ☐ Unscorable: state missing evidence



Keep the denominator attached to the task

4 / Repeat per task and case variant. No automatic calculations, pooled pass score or universal threshold.

Task / variant / configuration group	All prespecified attempts (denominator)	Successful attempts (numerator)
Timeouts: subset of all attempts	Failed runs: subset of all attempts	Unscorable: subset of all attempts
Serious concerns; separate case IDs	Success / all prespecified attempts	Adjudication owner / date
Missing logs / disagreements / exclusions: account for every attempt	Safety concerns individually; severity, owner and next action	
Matched comparator: review / correction time and failed attempts	AI workflow: review / correction time and failed attempts	

Include timeouts and failed runs in the prespecified denominator and report them separately. Unscorable attempts are not successes; identify missing evidence rather than silently dropping them. Categories may overlap: do not sum subsets as if mutually exclusive. A small sample cannot certify rare-event safety or better patient outcomes.



Authorize a defined use, not a product name

5 / A local decision record, not deployment authorization. Record the real approver through your governance process.

- ☐ Supported for defined scope
- ☐ Continue controlled evaluation
- ☐ Not supported by current evidence

Decision rationale; exact supported task / users / configuration; explicit exclusions

Evidence gaps / corrections / retest criteria

Serious concern or configuration-change triggers

Integration, training, subscription, review and monitoring resources

Clinical / operations fallback coverage; actual cost mechanism

Decision owner / clinical approver / approval date

Next action owner / evidence deadline / monitoring recheck

Save a completed copy, close and reopen it to verify entries. Use Acrobat Reader or Preview; browser viewers may not preserve forms reliably. Fields are not signed approvals. No totals are calculated. Keep long prompts and responses in your approved evidence repository; print or share through approved channels. Time returned is not booked savings without a financial mechanism.