A TRACEABLE WORKFLOW

From external signals to reviewed decisions

A visible signal can justify investigation without proving a vulnerability, harmful outcome or control failure. Keep the evidence, reasoning and decision connected.

> **Observe / Structure / Correlate / Evaluate / Review / Report**

01 Observe

Define the service, assessment question and permitted scope. Record what is visible, where it was found and when. Preserve a dated reference or suitable evidence extract.

02 Structure

Describe the asset, source, timestamp, relevant excerpt and limitations consistently. Separate direct observations from interpretations and missing information.

03 Correlate

Compare independent sources where available. Check ownership, dates and context. Several copies of one source do not provide independent confirmation.

04 Evaluate

Explain what the evidence supports, the plausible consequence and the conditions needed for that consequence. Consider alternative explanations.

05 Review

A person checks the reasoning, evidence quality and remaining uncertainty. Record disagreement and information that could change the assessment.

06 Report

State the finding, practical significance, next action, responsible owner and review date. Preserve enough provenance for another reviewer to follow the reasoning.

> **Observed:** directly supported by a recorded source. **Inferred:** an interpretation with explicit reasoning. **Unknown:** information not established.
>
> An automated score is an input to review. Keep it connected to its evidence and limitations.

CHECKLIST 01

External exposure

For each item, record Yes / No / Unknown / Not applicable, a supporting source and a responsible owner. Use the checklist to guide review, not to calculate an automatic pass score.

- ☐ **01 Service inventory**
  Do we know which domains, applications and public interfaces belong to the organisation?
- ☐ **02 Ownership**
  Can we distinguish our services from suppliers, abandoned assets and unrelated lookalikes?
- ☐ **03 Public information**
  Do published pages and documents disclose only information the organisation intended to make public?
- ☐ **04 Account journeys**
  Are registration, sign-in, recovery and support routes clear and consistent?
- ☐ **05 Third parties**
  Can we identify the external services supporting the public user journey?
- ☐ **06 Privacy information**
  Can a visitor find current information about relevant data handling and contact routes?
- ☐ **07 Misuse reporting**
  Is there a working route for reporting impersonation, suspicious messages or security concerns?
- ☐ **08 Maintenance signals**
  Are public security notices, documentation and service information current?
- ☐ **09 Evidence quality**
  Have suspected issues been checked against plausible benign explanations?
- ☐ **10 Follow-through**
  Does each relevant concern have an owner, next action and review date?

> **Prioritise with context**
> What is exposed? Who could be affected? What additional conditions are necessary? How strong is the evidence? What action would reduce uncertainty or impact?

A version string, reachable page or suspicious domain name alone does not establish exploitation or malicious ownership.

CHECKLIST 02

AI in education and child-facing services

A practical conversation between service owners, educators and relevant specialists. Record an answer, supporting evidence and any unresolved question for each item.

01 Purpose. What task should the AI perform, and what decisions remain with people?

02 Users. Who is expected to use the service, and how are their needs reflected in its design?

03 Data. What enters the system? Could the task use less sensitive information or synthetic examples?

04 Lifecycle. Are storage, retention, deletion and onward sharing explained clearly?

05 Boundaries. What topics, actions or decisions should the system decline or escalate?

06 Reliability. How are unsupported answers, conflicting information and missing evidence identified?

07 Human oversight. Who can review, correct, interrupt or challenge an output?

08 Communication. Can users recognise that they are interacting with AI and understand its relevant limitations?

09 Accessibility. Can intended users understand instructions and use reporting and support routes?

10 Evaluation. Do test cases reflect the intended tasks, languages and foreseeable failure conditions?

11 Incident response. Who receives concerns, and what happens after a report?

12 Change control. Who approves material changes, and when must the service be reassessed?

> **Before a pilot**
> Record permitted tasks, test data, the responsible reviewer, evidence needed to proceed and conditions for pausing the pilot.

WORKED EXAMPLE

When evidence is not yet sufficient

Entirely fictional example. No real service, organisation or incident is represented.

SYNTHETIC CASE / A small educational organisation is considering an AI assistant for administrative questions.

Assessment question: Is there enough evidence to permit a limited pilot involving learner information?

Evidence	Recorded observation	Limitation
E-01	The supplied product description says submitted questions may be stored.	A description does not establish actual configuration.
E-02	The supplied information does not state a retention period.	Controls may exist but have not been supplied.
E-03	The proposed pilot plan includes questions containing learner details.	The pilot has not started.

Interpretation

Identifiable learner information could create an avoidable data-handling risk while retention and deletion remain unresolved.

Unknowns

Actual retention settings, deletion behaviour, supplier access and whether storage can be disabled.

Decision

Use only synthetic questions during the initial evaluation. Keep identifiable learner information outside the pilot until the service owner provides sufficient evidence for a reviewed decision.

Next evidence and review trigger

Request a current description of retention and deletion controls, applicable configuration and a documented verification result. Reassess when evidence arrives or the pilot purpose or data changes.

The conclusion concerns evidence sufficiency for this proposed use. It does not establish a breach or legal violation.

Source provenance and further reading: Adapted from SKOPION's published methodology, child-facing AI assessment scope and IGF booth description. These are instructional materials; no benchmark results are claimed.

skopion.de/en/methodology
skopion.de/en/child-and-youth-facing-ai-safety-assessment
Official IGF profile / Village Booth #20 SKOPION Intelligence

PRINTABLE WORKSHEET

Evidence and decision record

Use one record for one assessment question. Attach further evidence or notes when the available space is insufficient.

Record ID / assessment date / reviewer

Assessment question / service or asset / scope and exclusions

ID	Source / date	Observation	Limitation

Interpretation / alternative explanations / unknowns or conflicting evidence

Plausible consequence and necessary conditions

Confidence (Low / Moderate / High) and reason

Human review comments or disagreement

Decision and rationale

Action / responsible owner / target date

Evidence needed to close or reconsider / review date and change triggers

Outcome after review