

FIELD GUIDE / VOLUME 01

THE EVIDENCE OF MAKING

How to make decisions inspectable.

A practical workbook for creators, researchers, builders and teams.



Inside: three candid worked examples, an evidence ladder, test plans, decision records, printable worksheets, retrospective prompts, and reusable checklists.

The governing idea

A convincing story is not the same as proof. Record the difference.

SAHIL TAYAL

sahiltayal.com • sahil@sahiltayal.com

What counts as evidence?

This guide offers portable methods, not a certificate of success.

Great portfolios reveal polished outcomes. Useful casefiles reveal the **decision path**: what was uncertain, what was tried, which signals mattered, and what remains unknown.

FIVE LABELS WORTH KEEPING SEPARATE

- **Observed:** A trace someone could independently check: a test result, screenshot, timestamp, accessible reproduction or user quote.
- **Inferred:** A reasoned interpretation of observations. State what else could explain them.
- **Proposed:** A plan or design choice that has not yet been tested.
- **Illustrative:** A fictional or composite example for learning. Helpful, but not evidence of a delivered result.
- **Unknown:** A meaningful gap. Record it instead of hiding it.

DISCLOSURE

The three project examples throughout this guide are **illustrative design exercises** adapted from the Casefiles concept. They do not claim independently verified speeds, accessibility conformance, client outcomes, or experiments.

PRACTICE

Find a project claim you have made. Label it **observed, inferred, proposed, illustrative, or unknown**. What artifact would allow a stranger to check it?

02 / METHOD

The eight-line decision record

Capture before, during, and after the choice. One page can prevent weeks of confusion.

01 THE DECISION

What precise fork in the road are you deciding?

PROMPT

Name the choice, not the whole project.

02 THE USER AND GOAL

Whose life improves, and what outcome do they actually need?

PROMPT

A user story is more useful than a slogan.

03 CONSTRAINTS

Separate non-negotiables from preferences. Include accessibility, time, maintenance and privacy.

PROMPT

Make trade-offs explicit.

04 OPTIONS

Record at least two plausible alternatives, including doing nothing.

PROMPT

Do not compare against a straw man.

02 / METHOD CONTINUED

Complete the decision record

Give evidence, accepted costs and revision criteria equal attention.

05 ASSUMPTIONS

What must be true for your favoured option to work?

PROMPT

Pick falsifiable predictions.

06 EVIDENCE AND GAPS

What do you know; how do you know; what remains untested?

PROMPT

Label observations and uncertainty.

07 TRADE-OFF AND OWNER

What becomes worse; for whom; who accepts responsibility?

PROMPT

A decision without a cost is incomplete.

08 REVISIT TRIGGER

Specify what would change your mind and when to inspect it again.

PROMPT

Allow revision without shame.

03 / EVIDENCE LADDER

How strong is this claim?

Claims need evidence proportionate to their consequences.

LEVEL 0 – CONCEPT

An idea is sketched. Appropriate wording: "We propose" or "This prototype explores."

LEVEL 1 – INSPECTABLE ARTIFACT

The interface or source is available. This demonstrates existence, not usability.

LEVEL 2 – REPRODUCIBLE TEST

The test method, device, environment, and result are recorded. One run is not a universal guarantee.

LEVEL 3 – REPEATED EVIDENCE

Results survive more than one device, task, visitor, or time point; variability is acknowledged.

LEVEL 4 – FIELD OUTCOME

A measured improvement in the real setting, with plausible alternatives and limitations considered.

LANGUAGE SWAP

Replace **"100% accessible"** with **"Keyboard paths were checked against documented cases; screen-reader evaluation remains pending"** unless stronger evidence exists.

SMALL EXERCISE

Claim: "Our site is instant." What does instant mean? To whom? On which device and network? Which measurement method would falsify the claim?

04 / CASE STUDY A

A website with multiple modes

Illustrative web architecture casefile — no historical performance metrics are asserted.

THE DECISION

Should one content collection support several visual modes, or should each experience ship as its own site? The visitor wants variety; the maintainer wants one dependable source of truth.

OPTION A / SEPARATE PAGES

Each mode has its own markup and content. Easy to experiment independently; duplicated updates can drift.

OPTION B / SHARED DATA

Modes reuse semantic content and change presentation. Consistency improves; architecture and testing become more demanding.

A PLAUSIBLE CHOICE

Start with shared structured content and progressively enhance it with distinct visual modes. Keep a stable text-first baseline before exploring complex 3D, animation or surprise interactions.

THE REAL TRADE-OFF

More expressive modes create more combinations to test: screen readers, mobile layouts, reduced motion, keyboards, and browser support. Visual ambition is not free.

WHAT WOULD PROVE VALUE?

- Run a content-consistency audit after adding or editing one entry; check every mode.
- Perform keyboard and screen-reader tasks in the baseline and each enhanced mode.
- Measure a stated performance metric on named devices; report distributions, not a decorative number.

05 / CASE STUDY A

Testing the mode engine

Turn strong language into modest, repeatable tests.

QUESTION

Can visitors discover and use key material in each mode without losing their place?

PROTOCOL

On narrow, medium and wide viewports, ask users to open an entry, change modes, return to the entry, use search, and contact the author. Record failures and confusion.

ACCESSIBILITY CHECK

Verify focus order, visible focus, interactive names, keyboard escape paths, contrast, reflow and reduced-motion alternatives. Formal conformance requires a broader audit.

PERFORMANCE CHECK

Record browser version, connection, device, trace method and a representative run set. Do not confuse cumulative layout shift with milliseconds.

FAILURE TRIGGER

If a mode hides required material or traps focus, degrade gracefully to the baseline experience.

REUSABLE PRINCIPLE

A mode is a different **way of seeing** the material, not a different standard of basic usability.

06 / CASE STUDY B

A browser-based learning interaction

Illustrative interaction-design casefile.

THE DECISION

A tiny practice widget teaches keyboard navigation. Use a heavy state-management framework or a small native DOM model? Neither is intrinsically moral or superior; the project size matters.

NATIVE FIRST

Buttons, fieldsets, details and predictable document order do much of the work. Simpler deployment and no framework runtime.

FRAMEWORK WHERE JUSTIFIED

Can help manage substantial shared state and multiple collaborators. Requires dependency, build and maintenance decisions.

A CHOSEN EXPERIMENT

Prototype the lesson with semantic controls first. Add JavaScript only for state transitions that native controls do not express. Write down every custom keyboard interaction so it can be tested.

IMPORTANT CORRECTION

Avoid claims like **"100% ARIA compliant"** or **"100-year longevity"** without a meaningful test definition. Smaller dependency graphs may aid maintenance but do not guarantee either outcome.

WHAT TO INSPECT

- Could participants complete the task using keyboard only?
- Does focus return to a sensible position when a panel updates?
- Can someone explain the learning goal without reading implementation details?

07 / CASE STUDY B

Testing a learning widget

Use learning evidence, not visual polish, as the primary signal.

A SMALL, ETHICAL USABILITY SESSION

- **1. Before:** Ask the learner to demonstrate the task without instruction. Observe, do not judge.
- **2. During:** Let them use the widget. Note confusing labels, missed announcements, dead ends and focus behaviour.
- **3. After:** Give an equivalent task. Ask what transferred and what still feels uncertain.
- **4. Follow-up:** Check whether the learner can repeat the task later, outside the widget.

EVIDENCE LOG EXAMPLE

OBSERVATION

"Learner returned to the main control after opening the example."

INTERPRETATION

"The focus behaviour may have been predictable." This is a hypothesis until the learner confirms it.

DISCONFIRMING EVIDENCE

A person can complete the demo by guessing, or the control can feel clear on desktop but confusing with a screen reader. A successful click is not automatically learning.

BUILD THE CORRECT THING

The objective is not a zero-dependency trophy. The objective is an understandable and transferable learning experience.

08 / CASE STUDY C

Translating fragrance into a screen

Illustrative physical / brand narrative casefile.

THE DESIGN PROBLEM

How can an interface help someone imagine top, heart and base notes when a screen cannot transmit smell? A well-chosen analogy can invite curiosity, but it must not masquerade as sensory equivalence.

LITERAL EXPLANATION

Name ingredients, note families, concentration caveats and an honest description of what is known.

EVOCATIVE TRANSLATION

Use tempo, typography, colour and narrative metaphors to describe a scent journey; label metaphors as interpretations.

A HUMANE COMPROMISE

Offer both modes: a clear fragrance description and an optional evocative experience. Text must carry essential information without colour, animation or audio.

BOUNDARY

"Citrus feels bright" is an evocative convention, not a universal fact of perception. People differ in associations, olfactory access and sensory preferences.

WHAT TO INSPECT

- Can a visitor distinguish literal note descriptions from artistic metaphors?
- Do choices remain usable with sound off and reduced motion enabled?
- Does the brand invite exploration without promising psychological effects?

Testing multisensory storytelling

Explore interpretation without prescribing what people should feel.

FIVE PROMPTS FOR PARTICIPANTS

- After reading the literal scent description, what do you expect? What remains hard to imagine?
- What did the visual metaphor add, if anything? Did it create a misleading impression?
- Could you find the same information without animations, colour distinction or sound?
- Did any imagery assume everyone has the same sensory memories?
- Would a different tone, cultural reference or level of specificity feel more helpful?

A TEST RECORD

Observation: A participant asked whether the imagery indicated the formula. **Possible interpretation:** The metaphor was not clearly separated from product information. **Next change:** Introduce a plain-language label and test again.

DESIGN PRINCIPLE

The goal is meaningful translation, not a claim that typography literally reproduces smell.

10 / BEFORE YOU CHOOSE

A practical trade-off matrix

Qualitative comparison often beats fake precision.

Compare options across criteria that matter to real users. Avoid summing invented 1–10 ratings into an aura of certainty.

VISITOR VALUE

Does this option help someone complete the central task?

ACCESS AND INCLUSION

Who cannot use the experience without alternatives?

MAINTAINABILITY

How hard is it to revise in three months?

FAILURE RECOVERY

What happens on old browsers, slow devices or lost connectivity?

EVIDENCE BURDEN

How difficult is it to verify the promised outcome?

REVERSIBILITY

Can we roll back if the idea fails?

DECISION STATEMENT

"We chose B because it protects access and content consistency; we accept a larger testing surface; we will reverse course if basic tasks fail in any public mode."

11 / RUN AN EXPERIMENT

A test card you can actually use

Distinguish a design preference from a claim that can fail.

CLAIM

Write one falsifiable sentence, e.g. "Visitors can locate the field guide without using the main navigation."

POPULATION AND CONTEXT

State where participants come from and how that limits what you can conclude.

TASK

Give a neutral instruction. Do not coach the answer.

SIGNAL

Record completion, errors, confusion or participant explanation as appropriate.

STOPPING RULE

Set a practical test window before seeing results.

ALTERNATIVES

Name competing causes such as prior familiarity, task wording or device differences.

INTERPRETATION

Write what results support and what they do not establish.

NEXT REVISION

Choose keep, revise, discard or gather more evidence.

12 / THINKING ERRORS

Six traps that distort project stories

Every builder meets these. The task is to notice them, not to perform infallibility.

SURVIVORSHIP BIAS

Publishing only the elegant final artifact hides failed routes. Keep a change log of meaningful abandoned ideas.

METRIC THEATRE

An impressive number without a measurement method communicates certainty but little information.

POST-HOC CERTAINTY

The path seems obvious after success. Preserve dated notes and alternative hypotheses.

TOOL LOYALTY

A preferred framework becomes an identity. Ask whether it improves the specific user task.

BEAUTY HALO

Polished visuals can make unsupported claims feel stronger. Separate aesthetics from evidence.

SCOPE DRIFT

What began as one user goal expands into novelty for its own sake. Re-state the problem before building more.

CORRECTIVE RITUAL

For any bold claim, ask: **"What would I expect to observe if this were false?"**

13 / RELIABILITY

A pre-release inspection sheet

Use the list as a trigger for testing, not proof of certification.

CONTENT

Links resolve; claims carry labels; examples are named as examples; the takeaway is genuinely complete.

INTERACTION

Every visible button acts; all panels work with keyboard; focus is visible; reset and export work.

MOBILE AND MOTION

Check narrow viewports, 200% zoom, text wrapping, touch sizes and prefers-reduced-motion.

PRIVACY

No secrets in client code; no unrequested data collection; local data is clearly described and removable.

PERFORMANCE AND RESILIENCE

Baseline readable without animation; no critical remote dependencies; offline assets linked relatively.

RELEASE

Stage before deploying; review the ZIP root; preserve a rollback copy; verify the real domain and HTTPS.

DEFINITION OF DONE

A release is a starting hypothesis. The real test is whether the intended visitor can use it, take something away, and understand its limits.

14 / BETTER RETROSPECTIVES

Review the reasoning, not the ego

A review should improve judgment even when the outcome was disappointing.

A 20-MINUTE REVIEW FORMAT

MINUTES 0–4

Restate the original goal and constraints without embellishing them.

KEEP SEPARATE

What happened? / Why might it have happened?

MINUTES 4–8

Compare expected and observed signals. Show raw artifacts when possible.

KEEP SEPARATE

What happened? / Why might it have happened?

MINUTES 8–12

Identify surprising failures, unintended beneficiaries and people excluded.

KEEP SEPARATE

What happened? / Why might it have happened?

MINUTES 12–16

Distinguish bad decisions from bad luck and good decisions from lucky outcomes.

KEEP SEPARATE

What happened? / Why might it have happened?

MINUTES 16–20

Record one revised principle, one follow-up test and a revisit date.

KEEP SEPARATE

What happened? / Why might it have happened?

ONE QUESTION WORTH KEEPING

If you could replay the decision using only the information available *at that time*, what would you still choose?

Decision record — blank

Print this page or reproduce it in your notes. More room is better than false completeness.

PROJECT / DATE / DECISION OWNER

VISITOR GOAL / WHO BENEFITS

CONSTRAINTS / OPTIONS CONSIDERED

ASSUMPTIONS / EVIDENCE / UNKNOWNNS

CHOSEN OPTION AND EXPLICIT TRADE-OFF

WHAT WOULD CHANGE OUR MIND?

16 / TEAR-OUT WORKSHEET

Field test card — blank

One test, one clear question, honest limits.

CLAIM OR QUESTION**CONTEXT / PARTICIPANTS / DEVICE****TASK AND PROCEDURE****OBSERVABLE SIGNALS****RESULTS AND RAW TRACES****ALTERNATIVE EXPLANATIONS****NEXT EXPERIMENT / REVISIT DATE**

Retrospective — blank

Use after a release, experiment or abandoned prototype.

WHAT WERE WE TRYING TO ACHIEVE?

WHAT DID WE EXPECT?

WHAT ACTUALLY HAPPENED?

WHAT SURPRISED US?

WHO BENEFITED / WHO WAS LEFT OUT?

WHICH ASSUMPTION CHANGED?

WHAT WILL WE REVISE OR PRESERVE?

Twelve sharp questions to borrow

Pick three that reveal what the prettiest screenshot cannot.

- 01. Which decision would be hardest to explain to someone affected by it?
- 02. What evidence would convince us to remove our favourite feature?
- 03. Did we optimise the metric or the underlying human need?
- 04. Which alternative did we reject too early?
- 05. What does the current page imply that we have not proved?
- 06. Which visitor needs an entirely different route?
- 07. If animation disappears, does the meaning survive?
- 08. What happens if network requests fail after first load?
- 09. Can another person reproduce the result without asking us?
- 10. Is the accepted trade-off visible to the people who bear it?
- 11. Which useful question did we avoid because it might slow us down?
- 12. What should the next builder inherit from our mistakes?

REFLECTION

The archive becomes more valuable each time its owner changes their mind for a reason they can explain.

Keep the work inspectable

The project is stronger when a stranger can examine your reasoning without first accepting your self-description.

An evidence-minded practice does not mean perfection, sterile design, or working only on things already validated. Imagination creates possibilities; the record protects our ability to learn from them.

A COMPACT NEXT-PROJECT RITUAL

- Before: write a decision record and the strongest alternative.
- During: keep dated observations separate from interpretations.
- After: publish a candid outcome, an uncertainty and the next question.

CONTINUE EXPLORING

Main home: <https://sahiltayal.com/>

Interactive casefiles: <https://casefiles.sahiltayal.online/>

Write to Sahil: sahil@sahiltayal.com

OPTIONAL TECHNICAL REFERENCES

For real-world implementation, consult the current W3C Web Content Accessibility Guidelines, the WAI-ARIA Authoring Practices Guide, and MDN Web Docs. These are reading suggestions, not proof that an example in this booklet passes an audit.

A CLOSING THOUGHT

Make boldly. Document carefully. Be willing to revise. A meaningful casefile tells the truth about the journey, including the parts still unfinished.