SERPIFAI FIELD GUIDE

Evidence Sufficiency Before Budget Commitment

Use this guide to answer one question: **is the current evidence sufficient for this next bounded commitment?** It does not calculate a confidence score or recommend a monetary amount.

> **Decision rule**
> The evidence threshold should rise with the consequence and difficulty of reversing the next commitment. A small reversible learning action can be supportable while meaningful uncertainty remains; a large hard-to-reverse rollout needs stronger evidence.

Version 1.0.0 - 2026-08-05. Illustrative financial amounts in the examples are not investment or financial-performance recommendations.

1. The 30-second answer

DO NOT ASK

Do we know everything?

Certainty is usually unavailable. Waiting for it can turn research into delay without improving the decision.

ASK INSTEAD

What does this evidence safely support now?

Judge sufficiency against the next bounded action, its downside, reversibility and the material uncertainty that could reverse the decision.

Hard stop: Do not authorize a large or difficult-to-reverse commitment when one material unresolved condition could reverse the decision. Do not average that blocker away with several favorable signals.

Small-test rule: Do not delay a genuinely bounded, reversible learning action merely because non-material uncertainty remains - but only when the test actually addresses the uncertainty controlling the next decision.

2. Same evidence, different commitment

A fictional company is considering a search and content expansion. The evidence set includes search demand, competitor evidence, technical feasibility and commercial evidence, but one buyer assumption is unresolved and one signal is contradictory.

Next commitment	What the evidence may support	Why
EUR 5k research / pilot	A smaller reversible test may be supportable.	The test can be designed to resolve the buyer assumption before scale.
EUR 30k production rollout	Conditional; may require more evidence or a narrower tranche.	The commitment is larger, so independence, first-party validation and contradiction treatment matter more.
EUR 150k broader expansion	Do not infer sufficiency from the pilot evidence alone.	A material buyer assumption or unresolved contradiction could reverse a hard-to-undo scale decision.

The amounts are fictional and illustrative. The method determines a permitted action class, not a recommended budget.

3. The nine-step operating method

1. **Define the next commitment.** Name the next bounded action, not the whole strategy.
2. **Test consequence and reversibility.** Ask what happens if the decision is wrong and how difficult or expensive it is to undo.
3. **Inventory relevant evidence.** Use evidence that actually bears on this decision boundary.
4. **Check independence, scope and freshness.** Several records can still represent one signal if they share provenance.
5. **Identify material contradiction and material unknowns.** Separate missing evidence from evidence that actively conflicts.
6. **Ask whether more information could change the action.** If it cannot, generic research has little decision value for this gate.
7. **Choose the permitted action.** Commit, run a smaller test, gather specific evidence, resolve scope/contradiction, hold or reopen.
8. **Define review and reversal conditions.** Every funded tranche should state when the decision is reviewed and what would cause hold, rollback or stop.
9. **Reopen when the decision changes.** A new scope, commitment size, contradiction or exposure class creates a fresh evidence-sufficiency gate.

4. Evidence conditions that matter

Condition	Beginner question	What changes
Relevance	Does this evidence actually answer this decision?	Irrelevant proof cannot support the commitment.
Independence	Are these really separate signals?	Duplicate provenance should not be counted as corroboration.
Scope	Does the buyer, offer, market and geography match?	Wrong-scope evidence can create false comfort.
Freshness	Is the evidence still current enough for this decision?	Evidence can decay when markets, products or conditions change.
First-party validation	Do we need our own customer or sales evidence here?	High commercial exposure may require first-party validation.
Contradiction	Do important signals disagree?	A material unresolved conflict is not the same as missing evidence.
Material unknown	Could what we still do not know reverse the decision?	One binding unknown can block a high-exposure commitment.

5. When to keep researching - and when to stop

Continue research when...

- a material unknown could reverse the commitment;
- a decision-relevant contradiction remains unresolved;
- the scope is wrong or incomplete;
- commercially necessary first-party validation is absent;
- the commitment is difficult to reverse and the evidence base is still weak or dependent.

Stop generic research when...

- the next action is sufficiently bounded;
- remaining uncertainty cannot change the permitted action;
- a real-world test can resolve the controlling uncertainty more effectively;
- additional research has lower decision value than acting within the supported boundary.

> **Practical value-of-information question:** Would learning this actually change what we do next? If not, collect it only if there is another explicit purpose - not because “more evidence” sounds safer.

6. Pilot or test? Only when it can learn the right thing

A pilot is not automatically the safe answer. It is useful only if the test is itself bounded and can materially reduce the uncertainty controlling the next commitment.

Question	Good test	Weak test
Learning objective	Names the uncertainty that blocks the next decision.	“See what happens.”
Decision link	Outcome can change expand / hold / stop / revise.	Produces activity metrics that cannot change the decision.
Exposure	Bounded cost, time, audience and operational blast radius.	Nearly the full rollout disguised as a pilot.
Review / reversal	Pre-defined review trigger and kill/reversal condition.	No point at which the team must reconsider.

7. Controlled outcomes - no score

Outcome	Meaning
Sufficient for this commitment	Proceed only with the defined commitment. This does not imply that a larger tranche is supported.
Sufficient for a smaller reversible test	The proposed larger commitment is not supported, but a bounded learning action can address the controlling uncertainty.
More evidence required	Gather the specific decision-relevant evidence that can change the permitted action.
Resolve contradiction	Disposition or resolve a material decision-relevant conflict before this commitment.
Resolve decision scope	Define the decision boundary before judging evidence sufficiency.
Do not commit yet	A material blocker could reverse a high-exposure decision and a safe decision-changing test is not yet available.
Reopen prior decision	Material new evidence or changed conditions require reassessment before further commitment.
Invalid input	Fix missing, malformed, wrong-case or unsupported controlled inputs before interpreting a result.

8. Logic result is not record completeness

A decision can be logically supportable while the handoff record is incomplete. Keep these states separate:

- **Logical decision result:** what the current evidence supports.
- **Record completeness:** whether owner, scope, evidence, review and reversal details are documented.
- **Publication / governance readiness:** whether organizational approval requirements have been satisfied.

Do not turn incomplete documentation into a favorable logic input, and do not turn missing paperwork into a false claim that the evidence itself is insufficient.

9. Review and reversal checklist

- Next commitment is written as one bounded action.
- Buyer, offer, market and geography are the intended decision scope.
- Consequence and reversibility are stated plainly.
- Evidence is relevant to the decision, not just available.
- Independent corroboration is distinguished from repeated provenance.
- Required first-party validation is present or explicitly missing.
- Material contradiction is resolved or blocks the commitment.
- Material unknowns are named; none is averaged away.
- Any proposed test directly addresses the controlling unknown.
- Review trigger is defined before funding the tranche.
- Reversal / kill condition is defined before scale.
- Reopen triggers are recorded for new evidence or changed exposure.

10. One-page commitment handoff

Decision owner: _____

Next commitment / tranche: _____

Decision scope: Buyer _____ Offer _____ Market _____ Geography _____

Consequence if wrong: Low / Moderate / High

Reversibility: Easy / Moderate / Hard

What the evidence supports: _____

What it does not support: _____

Material unknown or contradiction: _____

Next evidence or test: _____

Review trigger: _____

Reversal / kill condition: _____

Permitted next action: _____

For a reusable record with controlled inputs and the same decision logic, use the companion Evidence Sufficiency Register workbook and web-based Evidence Sufficiency Check.

11. Research basis and limits

The method synthesizes decision-under-uncertainty, evaluation, staged-governance and evidence-appraisal principles. Each source below supports a specific piece of the method; none is presented as endorsing SERPIFAI or defining a universal commercial evidence threshold.

1. **HM Treasury - The Green Book (2026)**
   https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government/the-green-book-2026
   Supports: Evidence should inform appraisal; uncertainty should be explicit; scrutiny and analytical technique should be proportionate to risk/cost; decision trees and real options can help with sequential or hard-to-reverse decisions.
   Does not establish: Does not define a universal commercial evidence threshold or endorse SERPIFAI.

2. **HM Treasury / Evaluation Task Force - Magenta Book: Central Government guidance on evaluation**
   https://www.gov.uk/government/publications/the-magenta-book/magenta-book-central-government-guidance-on-evaluation-html
   Supports: Evaluation should be planned before, during and after implementation; pilots and early tests can investigate risks, uncertainty, implementation and outcomes.
   Does not establish: Does not say every uncertain decision should be piloted.

3. **Evaluation Task Force - Test and Learn**
   https://www.gov.uk/government/publications/the-magenta-book/test-and-learn-html
   Supports: Test and Learn can generate robust learning before wider rollout and should be prioritized where learning can improve decisions.
   Does not establish: Does not justify a pilot whose design cannot address the controlling uncertainty.

4. **UK Government - Value for money study: mega-project governance and budgeting**
   https://www.gov.uk/government/publications/value-for-money-vfm-study-on-the-governance-and-budgeting-arrangements-for-mega-projects/value-for-money-study-governance-and-budgeting-arrangements-for-mega-projects
   Supports: Stage gates should be real decision points and additional funding may or may not be released; robust cost/risk evidence is required before final investment decisions.
   Does not establish: Applies to mega-project governance, not routine marketing spend.

5. **U.S. GAO - Cost Estimating and Assessment Guide**
   https://www.gao.gov/products/gao-20-195g
   Supports: Reliable estimates require defined scope, assumptions, data, sensitivity/risk analysis, documentation, cross-checking and updates with actuals.
   Does not establish: Does not define marketing evidence sufficiency.

6. **NIST - Assessment of Conformity, Decision Rules and Risk Analysis**
   https://www.nist.gov/publications/assessment-conformity-decision-rules-and-risk-analysis
   Supports: Decision rules should account for uncertainty and the asymmetric economic consequences of false acceptance/rejection.
   Does not establish: Metrology context; not a commercial investment rule.

7. **World Bank - Investment Decision Making Under Deep Uncertainty**
   https://ppp.worldbank.org/library/investment-decision-making-under-deep-uncertainty
   Supports: Under deep uncertainty, robust/flexible decision methods and real options can be preferable to pretending uncertainty can be eliminated.
   Does not establish: Climate/infrastructure focus; transfer requires caution.

8. **National Academies / NCBI - Uncertainty and the Value of Information**
   https://www.ncbi.nlm.nih.gov/books/NBK225344/
   Supports: Information has value when it improves decisions; collecting information should be judged against the value of improved decisions.
   Does not establish: Older and context-specific; conceptual transfer only.

9. **ISPOR Task Force / PMC - Value of Information Analytical Methods**
   https://pmc.ncbi.nlm.nih.gov/articles/PMC7373630/
   Supports: VOI addresses whether reducing uncertainty through new evidence can improve resource decisions.
   Does not establish: Healthcare framework; A35 uses the concept qualitatively.

10. **AHRQ - Assessing Applicability of Studies**
    https://effectivehealthcare.ahrq.gov/products/methods-guidance-applicability/methods
    Supports: Applicability should be assessed explicitly relative to the target population/context.
    Does not establish: Healthcare framing; maps conceptually to buyer/offer/market/geography.

11. **GRADEpro - Introduction to Evidence-to-Decision frameworks**
    https://book.gradepro.org/guideline/introduction-to-the-evidence-to-decision-frameworks

Supports: Structured decisions should consider benefits/harms, resources, values, equity, acceptability and feasibility rather than one evidence metric.
Does not establish: Healthcare framework; not directly adopted.

12. **Microsoft Azure - Architecture strategies for safe deployment practices**
https://learn.microsoft.com/en-us/azure/well-architected/operational-excellence/safe-deployments
Supports: Progressive exposure/canary rollout reduces impact by exposing changes to smaller groups before broad rollout.
Does not establish: Technical deployment context; not a marketing-budget rule.

The public article and its schema citations use this same representative source set. Context-specific guidance is transferred cautiously; commercial sufficiency remains decision-relative.