Research Methodology Handout

How to test research logic, evidence, engineering claims, and AI-assisted thesis reviews

PROBLEM → EVIDENCE → DECISION → CLAIM

Includes five sharp thesis-review stress tests, engineering and computer-engineering indicators, traceability matrices, and evidence-bound AI review guidance.

WRITTEN BY

Punpiti Piamsa-nga

July 2026

Contents

> *Student reading material. English only. This handout explains the main parts of a research report, the logic connecting them, and the evidence expected in each part. The hunger-example data are hypothetical and are not nutrition evidence.*

Version: 2026.07 monthly public edition.
Source page: https://punpiti.github.io/ai-assisted-thesis-review-handout/

Preface

This handout was prepared for students who need to review theses, manuscripts, or research reports before asking AI for help. Its central message is simple: review the research logic before polishing the language. A complete-looking document can still fail if the title is unfindable, the abstract is unclear, the research question is not really a question, the method cannot answer it, the evidence does not support the claim, or the references are unverifiable.

The handout uses a small hypothetical hunger example to keep the logic visible. The example is intentionally simple; it is not medical or nutrition evidence. The same review discipline should be applied to computer engineering work: define the problem, justify the method, specify metrics and mandatory gates, trace results to claims, and keep contribution inside the evidence.

The companion Thesis Review Prompt Pack is meant to generate review notes and review questions, not missing truth. Students remain responsible for every claim, citation, calculation, table, figure, ethical statement, and final decision.

Essential Terms Before Reading

- **Artificial intelligence (AI):** A computational system that can generate or analyze text; its output still requires verification.
- **Claim:** A statement the author asks the reader to accept.
- **Evidence:** Data, observations, calculations, sources, or results used to support a claim.
- **Research argument chain:** The connected reasoning from problem to conclusion. A chain is broken when one part does not support the next.
- **Research question (RQ):** The specific question the study is designed to answer.
- **Hypothesis:** One declarative, testable, and falsifiable statement for which predefined evidence permits a `Yes/No` decision: supported or not supported.
- **Objective:** The intended change or resolution of the problem stated in the problem statement; it is not merely an activity.
- **Output:** The direct, observable result showing that the stated problem has been solved according to the objective.
- **Outcome:** What happens after the problem has been solved. An outcome may be beneficial, harmful, intended, or unintended, and must be measured before it is claimed as a finding.
- **Research methodology:** The reasoning that connects the research question, design, data, methods, analysis, and interpretation.
- **Literature review:** An evidence-based reconstruction of how prior knowledge and attempts lead to an unresolved need and justify why the present study is necessary.
- **Method:** The selected way to address the stated problem and produce evidence. It should be chosen from plausible alternatives and justified with literature, selection criteria, and constraints.
- **Solution or intervention method:** The selected action intended to change the problem condition.
- **Research or evaluation method:** The design, measurement, and analysis used to determine whether the selected solution actually solves the problem.
- **Research gap:** A bounded absence, limitation, inconsistency, or unresolved issue demonstrated through prior knowledge.
- **Construct:** An idea that cannot always be observed directly, such as hunger, trust, or satisfaction.
- **Intervention:** An action or condition deliberately introduced, such as eating a specified meal.
- **Measured response variable:** The variable observed to provide evidence. Some fields call this an outcome variable; in this handout, **Outcome** is reserved for what happens after the problem has been solved.
- **Baseline:** The reference measurement taken before an intervention or comparison.
- **Confounder:** A factor that may influence both the condition studied and the measured response, making interpretation difficult.

- **Unit of analysis:** The entity about which the study makes a conclusion, such as a person, organization, event, or document.
- **Operational definition:** The exact rule used to observe, measure, or classify a concept in this study.
- **Scope or delimitation:** What the researcher deliberately chooses the study will and will not do.
- **Assumption:** A condition accepted as true or sufficiently stable for the design or analysis.
- **Limitation:** Something the study intended, needed, or would reasonably benefit from doing, but could not do because of a constraint; it reduces the strength of evidence, interpretation, or claim.
- **Validity:** The extent to which an interpretation or measurement is appropriate for its intended claim.
- **Reliability:** The consistency or repeatability of a measurement or procedure.
- **Generalizability:** The extent to which findings may apply beyond the studied sample and conditions.
- **Traceability:** The ability to follow a claim backward through results, evidence, method, and research question.
- **Sample size `n`:** The number of independent units studied; `n=1` means one unit, not one measurement.
- **Citation laundering:** Repeating a claim from a secondary source without checking the original source that supposedly supports it.
- **Systematic review:** A literature review that follows a predefined, transparent, and reproducible search and selection method.
- **Ground truth:** The reference label or value treated as the standard against which observations or predictions are evaluated.
- **Data leakage:** Information improperly passes between training, testing, or analysis stages and makes performance appear better than it should.
- **Counterfactual:** What would have happened under a relevant alternative condition; a comparison group or baseline may help estimate it.
- **Sensitivity or robustness analysis:** Testing whether a conclusion remains similar when reasonable assumptions, data choices, or analysis settings change.
- **Anonymization:** Altering data so individuals cannot reasonably be identified; simply removing names may not be sufficient.

Core Reading Questions

For every section of a research report, consider:

1. What function should this section perform?
2. What evidence should be present?
3. What other sections must agree with it?
4. Is information missing, weak, or internally inconsistent?
5. Does the claim stay within the design, data, and limitations?

Five Sharp Thesis-Review Tests

Run these tests before language editing. Failure may reveal a blocking research-logic problem that clearer writing cannot repair. If students cannot answer these five tests, the chance of the work failing is very high.

Sharp Test 1 — Kill the Problem Before Judging the Method

Do not begin by admiring the technique. First ask whether the research problem survives challenge.

- Does the stated undesirable or unresolved condition actually exist, and is there evidence that it matters?
- Is the condition still unresolved, or has prior work already solved it adequately within the stated context?
- Is the proposed work addressing the problem, or merely applying a technique, building an artifact, or completing a technical exercise?
- Could a simpler, safer, cheaper, or already established solution resolve the same problem sufficiently?
- Would the proposed Output demonstrate that the problem is gone according to the stated definition?

Decision: If the problem is not defensible, reframe or stop the study before debating Method. Technical sophistication cannot rescue an unnecessary study.

Sharp Test 2 — A Study Must Be Able to Fail

A study designed so that every possible result appears successful cannot provide convincing evidence of success.

- Pre-specify what evidence would support each hypothesis and what evidence would not support it.
- Define study-level failure: the result, threshold violation, contradiction, failure case, or unacceptable uncertainty that would show the selected method did not solve the stated problem.
- Include realistic unfavorable conditions and failure cases rather than testing only convenient examples.
- Treat a negative result as evidence, not as permission to rewrite the objective or decision rule after seeing the data.
- Separate `the system ran` from `the method solved the problem under the required conditions`.

Decision: If no possible observation can make the researcher answer `No`, the evaluation is not genuinely testing the claim.

Sharp Test 3 — Mandatory Gates Come Before Optimization

Engineering evaluation must distinguish requirements that cannot be traded away from metrics used to rank acceptable alternatives. Table 1 separates the roles of these two metric groups.

Table 1: Mandatory gates and optimization metrics

Metric role	Meaning	Decision rule	Examples
Mandatory gate / acceptance criterion	A necessary condition for an acceptable solution	Every gate must pass	Correctness, safety, security, legal compliance, maximum error, minimum required capacity
Optimization metric	A quantity used to compare solutions that already pass all gates	Improve, minimize, maximize, or balance	Response time, throughput, cost, energy, convenience, maintainability

- State each gate, threshold, test condition, and evidence before evaluation.
- Do not combine all metrics into one weighted average that allows a severe gate failure to disappear.
- Report guardrail metrics even when they are not the primary optimization target.

Safety example: A food-recommendation system may produce a cheaper or more popular meal plan, but it is not acceptable if it recommends ingredients that violate an allergy, medication restriction, religious constraint, or food-safety requirement. Safety is a gate. Taste, cost, variety, and convenience are optimization criteria only after the safety gate passes.

Sharp Test 4 — Measurement Quality Sets the Ceiling for Conclusion Quality

The conclusion cannot be stronger than the weakest justified link in this chain:

```
Construct → Operational indicator → Instrument / Ground truth
→ Data and procedure → Analysis → Claim
```

- A precise measurement of the wrong proxy does not measure the intended construct.
- More decimal places, samples, or statistical significance cannot repair invalid operationalization or biased ground truth.
- Calibration and repeatability matter only after the instrument is shown to measure the relevant quantity.
- Name the inference that connects a proxy to the claim and provide evidence for that inference.

Food example: Eating a luxury meal, eating until satiated, eating an expensive meal, eating for social display or courtship, and eating safely are not judged by the same measurement. Price does not measure satiety. Taste does not measure nutrition. Atmosphere does not measure allergy safety. Nutrient value does not remove the requirement that the food be safe to eat. If the claim is “safe,” the evidence must measure safety; if the claim is “nutritious,” the evidence must measure nutrition; if the claim is “satisfying,”

the evidence must measure satisfaction or satiety. A precise measurement of the wrong thing only makes the wrong conclusion look cleaner.

Decision: If the construct–indicator–instrument link fails, reduce the claim or redesign the measurement; do not polish the conclusion.

Sharp Test 5 — Do Not Solve One Problem by Creating Another

A target improvement is not a complete solution if it creates an unacceptable new problem for the system, users, stakeholders, or environment.

- Identify intended and unintended outcomes, affected stakeholders, and plausible externalities.
- Evaluate relevant guardrails such as safety, privacy, security, accessibility, cost, energy, maintainability, fairness, and legal or ethical obligations.
- Report trade-offs explicitly; do not present one improved metric as total success.
- Define which harms or threshold violations make the solution unacceptable.
- Check whether the burden or risk has merely been transferred to another person, group, process, or stage of the lifecycle.

Decision: Claim problem resolution only when the target is achieved and every mandatory harm-related guardrail passes. Otherwise report a trade-off, partial resolution, or new unresolved problem.

Integrated Five-Test Decision Card

Test	Blocking question
Problem survival	Is there still a real, evidenced problem that requires this study?
Genuine test	What possible evidence would make the answer `No`?
Mandatory gates	Did every necessary acceptance criterion pass?
Measurement ceiling	Does the evidence measure the construct required by the claim?
No new problem	Did the solution avoid creating or transferring unacceptable harm?

The Research Argument Chain

```
Problem
→ Literature trajectory: who did what, with what evidence
→ What worked, what did not, and under which conditions
→ Unresolved need / Gap / Why another study is necessary
→ Objective / Research Question / Hypothesis
→ Known solution or method alternatives
→ Selection criteria and rationale
→ Definitions and boundaries
→ Selected solution / intervention method
→ Research design / Data / Measurement / Analysis
→ Results and evidence
→ Output: the problem is solved as defined
→ Outcome: what happens after the problem is solved
→ Discussion / Conclusion / Contribution
→ Limitations and future work
```

Every arrow requires a defensible reason. Every claim must be traceable to evidence.

Running Example: Hungry → Eat → Satiated

- **Observation:** One person reports hunger after not eating for six hours.
- **Researchable problem:** The change in hunger score after a specified meal, under defined conditions, is not yet recorded for this case.
- **Problem:** The case has a hunger score of 8, above the defined satiety threshold, under the stated condition.
- **Prior-work trajectory:** Not supplied in this mock example. A real study must explain who investigated relevant hunger problems, which interventions and measures were used, and what their evidence did or did not establish under stated conditions.
- **Unresolved need and study necessity:** Not established at field level because no real literature is reviewed. The mock is justified only as a teaching case; it must not claim an academic gap.
- **Objective:** Reduce the hunger score to the defined threshold of 2 or lower within 30 minutes.
- **Known alternatives:** Different meal types, portions, preparation methods, prices, access conditions, or other appropriate responses to hunger. Literature is needed to map and compare them.
- **Selection rationale:** Select one specified meal using explicit criteria such as expected effect, safety, cost, availability, preparation time, suitability, and study feasibility—not because it is merely convenient or assumed to be best.
- **Solution/intervention method:** Eat the selected specified meal.
- **Research/evaluation method:** Record a 0–10 hunger score before eating and at 15, 30, and 60 minutes to test whether the objective is achieved.
- **Hypothetical result/evidence:** Hunger scores at 0, 15, 30, and 60 minutes are $8 \rightarrow 4 \rightarrow 2 \rightarrow 3$.
- **Output:** The score reaches 2 at 30 minutes; the defined hunger problem is solved at that time point.
- **Possible outcome:** Improved comfort or concentration after satiety. This outcome is not measured in the mock study and must not be claimed as a finding.
- **Bounded conclusion:** The defined output is achieved for this one case and condition.
- **Not supported:** The meal is best, works for everyone, improves concentration, or should guide health policy.

Part I — Research Framing

1. Research Title

- **Must do:** Identify the phenomenon, key variables or intervention, population/context, and sometimes the design.
- **Check:** Does the title describe the study that was actually conducted?
- **Findability:** A weak title makes the work hard to find. Readers, reviewers, and search systems should be able to infer the topic, context, and main object of study from the title.
- **Common failure:** Words such as *effective*, *optimal*, or *validated* are used without supporting comparison or evidence.
- **Example:** "Change in Self-Reported Hunger During 60 Minutes After a Specified Meal: A Single-Case Observation."

2. Abstract

- **Must include:** Problem/need, objective, design/data, method, principal result, bounded conclusion.
- **Check:** Do the numbers and claims agree with the main report?
- **Common failure:** The abstract contains motivation and method but no actual result.
- **Reader expectation:** If the abstract is unclear, readers will not expect the rest of the report to become clear. A weak abstract often ends the reading process.
- **Rule:** An abstract is a compact research argument, not an advertisement.

3. Background and Context

- **Must do:** Move from a broad context to a specific research problem.
- **Check:** Who is affected, under what conditions, and why does the situation matter?
- **Common failure:** A long background never identifies what remains unresolved.
- **Synthesis failure:** Some reports shorten and paste other people's work, then add only a tiny statement of the writer's own idea. Background is not a pile of summaries; it must synthesize evidence into the current problem.
- **Distinction:** Background explains context; it does not replace the problem statement.

4. Research Problem / Problem Statement

- **Form:** Write the problem statement as one clear **declarative sentence** expressing one central problem. If supporting details are needed, expand them in the background, rationale, scope, or operational definitions rather than packing several ideas into the problem sentence.
- **Not a question:** An interrogative sentence belongs in the Research Question section, not in the problem statement.
- **No method or solution:** Do not include a questionnaire, algorithm, experiment, model, tool, or preferred solution. Those belong in the Method section. Putting a method inside the problem statement assumes the answer before the research design is justified.
- **It must sound like a problem:** The sentence should describe an undesirable condition, unresolved uncertainty, missing capability, inconsistency, or inadequate knowledge.
- **It is not a field:** "Hunger and satiety research" names an area, not a problem.
- **It is not a topic:** "Hunger after meals" identifies a topic, not what is wrong or unknown.
- **It is not a desire or activity:** "I want to study meals" and "I want to build a system" describe the researcher's intention.
- **Must be supported:** Evidence in the background or literature should show that the stated condition or uncertainty exists.

Table 2 distinguishes statements often confused with a problem statement. A research question is not a research problem, and a research topic is still not a research problem.

Table 3: Problem-statement distinctions

Type	Example	Why it is or is not a problem statement
Research field	“Hunger and satiety research.”	Names a field only
Topic	“Hunger after eating.”	Names a subject only
Desire	“I want to study how meals affect hunger.”	Describes what the researcher wants to do
Research question	“How does hunger change after a meal?”	A question; belongs in the RQ section
Method-loaded statement	“A questionnaire should be used to measure hunger after eating.”	Prescribes a method before the problem is established
Problem statement	“Under the defined condition, this case has a hunger score of 8, which is above the stated satiety threshold.”	Declarative, one central undesirable condition, no prescribed method

- **Final check:** Is the statement specific, significant, researchable, bounded, and recognizably problematic?

5. Research Gap

- **Possible types:** Knowledge, methodological, empirical/contextual, inconsistency, practical.
- **Check:** Does the gap emerge from a comparison of prior evidence?
- **Prior-work impact check:** If strong prior work has already solved the same problem under the same conditions, the impact of the new study may be low. The report must explain what still remains unresolved, what condition differs, or why another study is still necessary.
- **Common failure:** “No one has studied this” is asserted without a transparent search.
- **Rule:** Personal interest or unavailable local data is not automatically a field-level gap.

6. Significance and Rationale

- **Must explain:** Academic/practical value, affected users or stakeholders, timing, cost, risk, and feasibility.
- **Check:** Does the expected benefit follow from the objective?
- **Common failure:** Significance is written as promotional language rather than an evidence-based rationale.
- **Reference failure:** Significance is often asserted with no citation, with weakly related citations, or with distant policy slogans such as national strategy or SDG language that barely connects to the actual project. A broad policy reference does not make a small, disconnected study important.

7. Objectives

- **Core meaning:** An objective states the change or resolution the research intends to achieve for the problem in the problem statement.
- **Direct relationship:** Each objective must solve, reduce, explain, or resolve a stated problem. If an objective cannot be traced to a problem, it is an orphan objective.
- **Not an activity:** “Eat a meal,” “collect data,” and “build a system” describe actions. The objective states what problem condition should change.
- **Check:** Can the objective be answered or verified by a method, result/evidence, and output?
- **Hunger example:** Reduce the hunger score from 8 to the defined satiety threshold of 2 or lower within 30 minutes.

8. Research Questions

- **Must be:** Specific, answerable, aligned with the objective, and free from the desired answer.
- **Not a research problem:** A research problem is an undesirable condition or unresolved uncertainty to be addressed; a research question is the specific question the study is designed to answer.
- **Form:** A research question is a question, not a statement.
- **Check:** What data and analysis would constitute an answer?
- **Method relationship:** The research/evaluation method should be able to produce evidence that answers the question. If the supposed answer is already embedded as the chosen method, the question may be method-loaded rather than evidence-seeking.
- **Example:** "How does self-reported hunger change during 60 minutes after a specified meal?"

9. Hypotheses or Propositions

- **Form:** A hypothesis is a declarative sentence, not a question. It states one expected relationship, difference, condition, or threshold.
- **Binary-decision test:** Before collecting results, the researcher must be able to complete: `Supported? Yes / No`, using specified evidence and a decision rule.
- **Atomicity:** One hypothesis should contain one claim. If the honest decision could be "partly," split the statement into H1a, H1b, and further sub-hypotheses until each can receive its own `Yes/No` decision.
- **Required content:** Variables or constructs, expected relationship/direction/threshold, population or system, conditions, time or workload when relevant, and evidence that will decide support.
- **Pre-specify:** Measurement, comparator/baseline, acceptance threshold or statistical test, and the result that would fail to support the hypothesis.
- **Interpretation:** `Yes` means the stated evidence supports the bounded hypothesis; it does not prove universal or permanent truth. `No` is a valid research result.
- **Statistical caution:** For null-hypothesis significance testing, report `reject` or `fail to reject` the null hypothesis rather than "prove" or automatically "accept" it.
- **Qualitative work:** May use propositions or guiding questions rather than forced statistical hypotheses.
- **Common failures:** A question disguised as a hypothesis; vague wording such as "the method should work"; multiple claims joined by *and*; no failure condition; or a decision rule created after seeing results.

Hunger hypotheses:

- H1a: Under the defined condition, the 30-minute hunger score is 2 or lower. `Supported? Yes/No.`
- H1b: Under the defined condition, the 30-minute hunger score is lower than the baseline score. `Supported? Yes/No.`

Computer engineering decomposition:

- Too compound: "The proposed meal-recommendation system gives useful, affordable, safe, and satisfying recommendations."
- H1a: For every test profile with a stated allergy, the system produces no recommendation containing a prohibited ingredient.
- H1b: For every test profile with a stated budget, the recommended meal cost is within that budget.
- H1c: Under the specified rating procedure, the average usefulness score is not lower than the baseline recommender.
- H1d: Under the specified explanation checklist, each recommendation states the main reason it fits the user's constraints.

Each sub-hypothesis requires its own evidence and `Yes/No` decision. High usefulness cannot hide an allergy failure, and a safe recommendation still needs separate evidence for cost, usefulness, and explanation quality.

10. Expected Contributions

- **Possible forms:** Empirical evidence, theory, explanation, method, instrument, dataset, benchmark, design principle, validated artifact, practical implication.
- **Check:** What evidence would demonstrate each contribution?
- **Common failure:** A list of activities or deliverables is presented as a contribution.
- **Disciplinary expectation:** For a master's or doctoral thesis in computer engineering, readers expect

a contribution to computer engineering, such as a method, metric, system design principle, dataset, benchmark, empirical evidence, or validated artifact. A generic activity or implementation alone is usually not enough.

11. Output and Outcome

Teaching definitions

- **Output = problem solved:** The direct and observable result showing that the objective has been achieved and the stated problem has been solved according to its definition.
- **Outcome = what happens after the problem is solved:** The subsequent effect, consequence, or change that follows the output.
- **By-product = other effect:** An additional effect that may occur beside the intended output or outcome. It can be useful, harmful, neutral, or irrelevant, and should not be sold as the main outcome unless the design measures it.
- **Evidence precedes output:** A table, measurement, observation, or analysis provides the evidence used to decide whether the output has occurred.
- **An expected outcome is not yet a finding:** If the study does not measure the downstream effect, report it as an expected or possible outcome, not a demonstrated outcome.

Required relationship

```
Problem → Literature Trajectory → Unresolved Need/Gap → Justification → Objective → Alternatives → Selection Rationale →
↪ Selected Method → Result/Evidence → Output → Outcome / By-product
```

- The **problem** describes the undesirable or unresolved condition.
- The **literature trajectory** explains how knowledge and attempted solutions developed: who studied what, what approach was used, what evidence was produced, what worked or did not work, and under which conditions.
- The **unresolved need/gap** follows from that comparison and provides a defensible reason why another study is still necessary.
- The **objective** states the intended resolution of that exact problem.
- The **literature review** may also identify solution or method alternatives and evidence relevant to selecting the current approach.
- The **selection rationale** compares alternatives using criteria relevant to the problem, scope, stakeholders, risks, resources, and constraints.
- The **selected method** is the justified choice used to address the problem; the research/evaluation method then produces evidence of whether it worked.
- The **output** demonstrates directly that the problem has been solved according to the objective.
- The **outcome** describes what follows after the problem is solved.
- The **by-product** records additional effects that may affect the value, safety, cost, or burden of the design.
- If any item refers to a different condition, population, time, or construct, the chain is broken.

Hunger example

Table 3 provides a template for checking that the literature is not merely a list of article names, but a reasoned path into the current study.

Table 4: Literature-to-current-study logic check

Element	Example	Logic check
Problem	Hunger score is 8, above the satiety threshold	Sounds like a problem; no method included
Prior attempts/evidence	Not supplied in the mock	A real review must show who did what, with what evidence, and under which conditions

Element	Example	Logic check
Unresolved need / why still necessary	No field-level necessity established; teaching case only	Do not invent a gap when literature is absent
Objective	Reduce the score to ≤2 within 30 minutes	Directly resolves the stated problem
Alternatives from literature	Compare relevant meal types, portions, preparation, price, availability, expected effects, and risks	A secondary literature-review function; shows that the selected response is one of several possibilities
Selection rationale	Choose one specified meal against explicit criteria and study constraints	Explains why this option is suitable; does not claim it is universally best
Solution/intervention method	Eat the selected specified meal	Implements the chosen way of addressing hunger
Research/evaluation method	Measure hunger over time using a defined scale	Produces evidence for judging whether the objective was achieved
Result/evidence	Hunger scores at 0, 15, 30, and 60 minutes are 8 → 4 → 2 → 3	Provides observations for judging success
Output	Score reaches 2 at 30 minutes	Direct evidence that the defined problem is solved at that point
Outcome	Improved comfort or concentration after satiety	Not measured; may be stated only as an expected outcome
By-product	Leftovers, extra cost, time spent preparing food, or discomfort after eating	May affect design value, but must be measured before being claimed

Common failures

- Objective describes an activity but not the resolution of the problem.
- Literature is a list of work that looks similar, with no account of what the evidence established or why another study is necessary.
- A gap is claimed because no paper uses the exact same topic wording, dataset, location, or method combination.
- Only one method is presented, so the reader cannot see what alternatives were considered.
- A method is selected by habit, convenience, prestige, or preference without literature and explicit criteria.
- The report describes an intervention but provides no evaluation method capable of showing whether it solved the problem.
- Output is a deliverable that does not show the problem was solved.
- Outcome is attractive but unrelated to the output.
- Outcome is claimed even though it was never measured.
- By-products are ignored even though they affect the value, risk, or feasibility of the design.
- Problem, objective, output, and outcome refer to different populations, variables, or time periods.

Part II — Knowledge Base and References

12. Literature Review

- **Core purpose:** Reconstruct the intellectual and technical path that makes the present study necessary. It is not a shelf of studies that merely look similar to the current work.
- **Must explain:** Who addressed which part of the problem; what theory, data, method, system, or intervention they used; what evidence they produced; what performed well or poorly; under which population, environment, workload, assumptions, and boundaries; and what remains unresolved.
- **Required ending:** State a reasoned bridge from prior work to the present study: `What is known → What has been attempted → What the evidence shows → What remains inadequate or uncertain → Why this study is still needed → What this study will address.`
- **Additional functions:** Define concepts; identify validated measures and baselines; reveal candidate solutions and method alternatives; compare effectiveness, cost, risk, feasibility, conditions, and limitations; justify the selected method and design.
- **Synthesis:** Organize evidence into a developing argument—by problem, approach, evidence, chronology, debate, or boundary condition—not as one isolated summary per paper.
- **Check:** If the final paragraph and the current-study name were removed, could a reader still infer why another study is necessary and what unresolved issue it should address?
- **Common failures:** Listing "studies similar to ours"; describing only what each author did without judging evidence; reporting only positive results; using criticism that is unrelated to the current objective; or declaring a gap merely because the exact combination of words has not appeared before.

Literature-to-Study Matrix

Prior work / group	Problem addressed	Approach and evidence	What worked / strength	What did not work / boundary	What remains unresolved	Consequence for this study
Study or research line	Specific part of the problem	Theory, data, method, comparison, result	Evidence-supported value	Limitation, contradiction, failure, or untested condition	Remaining need	Why the current objective or design is necessary

13. Theoretical Framework

- **Must do:** Explain the phenomenon and inform expected relationships, interpretation, and boundary conditions.
- **Check:** Does the theory influence the research question (RQ), method, or discussion?
- **Hunger example:** A simple satiety framework might connect time since previous meal, meal composition, portion size, taste, eating atmosphere, social context, and personal condition to self-reported hunger over time. The framework tells the researcher what to measure, what to control, and what must remain outside the claim.
- **Common failure:** A theory is named but never used.

14. Conceptual Framework

- **Must show:** Constructs/variables, proposed relationships, and what lies inside or outside the model.
- **Hunger example:** Previous-meal interval + meal characteristics + taste + eating atmosphere + group/-social context + person factors + data-collection burden/cost → hunger over time.
- **Check:** Are all important model elements measured or explicitly treated as outside scope?

15. References, Citations, and Bibliography

- **Citation:** Points to a source inside the text.
- **Reference entry:** Provides enough metadata to locate the source.
- **Reference list:** Contains the works cited; a bibliography may include additional works depending on institutional style.
- **Check:** Exact claim support, text/list consistency, metadata, Digital Object Identifier (DOI) or web address, primary source, and attribution for quotations and reused figures/tables.
- **Common failures:** Fabricated citation, citation laundering, irrelevant source, outdated source for a current claim.
- **AI warning:** If AI fabricates a reference, the damage is severe. A hallucinated reference is not a minor formatting error; it is a research-integrity failure.
- **Rule:** If a source is unavailable, write `reference required` or `missing evidence`; never invent it.

16. Source Search and Selection

- **Record where rigor matters:** Databases, search terms/date, inclusion/exclusion, date/language range, duplicate handling, quality appraisal.
- **Check:** Could another reviewer understand how the knowledge base was assembled?
- **Rule:** Match search rigor to the claim; not every literature review is a systematic review.

Part III — Boundaries and Research Design

17. Conceptual and Operational Definitions

- **Conceptual definition:** Meaning within a theory or discipline.
- **Operational definition:** How the construct is observed, measured, or classified in this study.
- **Check:** Is the same definition used across method, results, and conclusion?
- **Example:** Hunger is a 0–10 self-report; satiety is defined as a score ≤2 for the teaching example only.

18. Scope and Delimitations

- **May define:** Population, setting, period, data, variables, method/intervention, measured results, downstream outcomes, inclusion/exclusion.
- **Plain-language test:** "We choose to study only this much."
- **Delimitation:** A planned boundary selected before or during design to keep the study focused on its stated objective.
- **Example:** The study deliberately covers one person, one selected meal, one 60-minute observation, and self-reported hunger only. Other populations, meals, periods, and outcomes are outside the chosen scope.
- **Check:** Does the conclusion remain inside these boundaries?
- **Do not write:** A chosen exclusion as though it were an unexpected failure.

19. Limitations

- **Plain-language test:** "We wanted, needed, or would reasonably benefit from doing this, but a constraint prevented it."
- **Definition:** A restriction that prevents part of the intended or desirable study from being achieved and therefore weakens evidence, interpretation, or generalization.
- **Categories:** Design, sampling, measurement, data quality, analysis, external validity, resources.
- **Example:** A suitable validated hunger scale was unavailable, and relevant contextual factors could not be fully controlled or measured; therefore measurement and interpretation remain uncertain.
- **Check:** Which claim is affected by each limitation?
- **Critical distinction:** The same fact must not be listed as both scope and limitation without explaining intent. "We chose one case" is scope. "We planned to recruit 30 participants but could recruit only one because access failed" is a limitation and a deviation from the intended scope.

20. Assumptions

- **Definition:** Conditions accepted so the design or analysis can operate.
- **Examples:** A rating scale is understood and used consistently; measurement conditions remain stable.
- **Check:** Can the assumption be examined, justified, or tested?
- **Common failure:** A critical assumption remains hidden.

21. Research Paradigm and Approach

- **Families:** Quantitative, qualitative, mixed methods, design science/engineering, case study, experimental/quasi-experimental/observational, simulation/modeling.
- **Check:** Is the approach appropriate to the research question (RQ) and required evidence?
- **Selection rationale:** State the alternatives considered, comparison criteria, supporting literature, trade-offs, constraints, and why the selected approach is the best fit for this problem and scope.
- **Important distinction:** The solution/intervention method addresses the problem; the research/evaluation method tests whether the problem was actually solved. A defensible study explains both.
- **Common failure:** A prestigious, familiar, convenient, or available method is chosen before alternatives are reviewed and compared.

22. Population, Sample, Case, and Unit of Analysis

- **Define:** Target population, accessible population, sampling frame/method, sample-size rationale, inclusion/exclusion, unit of observation, unit of analysis.
- **Hunger rule:** Four time points from one person do not equal four participants.
- **Check:** What entity does the conclusion describe?

23. Variables, Constructs, Factors, and Confounders

- **Identify:** Intervention/exposure, measured response variable, controls, mediators/moderators when relevant, confounders.
- **Hunger example:** Meal → hunger score; confounders include prior meal, sleep, stress, activity, and drink.
- **Check:** Are variables in the hypothesis present in the data and analysis?

24. Data and Dataset

- **Describe:** Source/provenance, collection period, schema/units, sampling/coverage, missingness, cleaning, labels/ground truth, bias/leakage, license/access, version/retention.
- **Check:** Can these data answer the research question (RQ)?
- **Common failure:** Cleaning removes difficult cases without justification or dataset splits leak information.

25. Instrument and Measurement

- **Include:** Construct, instrument/scale/sensor, scoring/units, calibration, validity evidence, reliability/precision, administration protocol.
- **Check:** Does the instrument measure what the claim says it measures?
- **Hunger example:** Hunger score, satisfaction with eating, taste, nutrition amount, allergy/safety status, and cost are different constructs. A satisfaction survey does not measure nutrient intake; a nutrition table does not measure enjoyment; a cheap meal does not prove safety.
- **Common failure:** A short or convenient instrument is assumed to be valid.

Engineering and Computer Engineering Indicators

Engineering research must translate words such as *better*, *faster*, *safer*, *more accurate*, or *more efficient* into operational indicators. Select indicators from the problem and objective; do not collect familiar metrics merely because a tool reports them.

For every primary indicator, specify:

- name and operational definition
- unit and direction of improvement
- baseline, comparator, or reference value
- target or acceptance threshold linked to the objective
- workload, environment, input distribution, hardware/software version, and test conditions
- aggregation, such as mean, median, percentile, worst case, rate, or distribution
- repetitions, variability, uncertainty, and failure handling

Table 4 gives a review checklist for common engineering indicators. Table 5 extends the examples to computer engineering concerns, where system, software, data, and security conditions often determine whether a metric is meaningful.

Table 6: General engineering indicator checks

General engineering concern	Possible indicators	Review question
Effectiveness	Goal-achievement rate; target error or gap	Did the system achieve the defined objective?

General engineering concern	Possible indicators	Review question
Efficiency	Useful output per unit of input, time, energy, or cost	Are output and input defined and comparable?
Performance and speed	Work per time; completion time	Is the workload and test condition fixed?
Safety and risk	Failure probability, severity, exposure, recovery	Are both likelihood and consequence considered?
Cost and resource use	Total cost, cost per unit, energy, material, labor	What lifecycle and cost boundary are included?
Loss or waste	Input minus useful output; defect or loss rate	What counts as useful output?
Accuracy and error	Absolute/relative error or task-specific correctness	What is the ground truth or reference?
Precision or repeatability	Variation across repeated measurements	Does “precision” mean repeatability or a classification metric?
Durability and reliability	Cycles/time before failure; failure probability	What mission time and operating condition apply?
Maintainability	Repair time, change effort, defect-removal time	What maintenance event starts and ends the clock?

Table 7: Computer engineering indicator checks

Computer engineering concern	Example indicators	Necessary context
Response time	Average, median, slowest case, or percentiles when needed	User task, request type, workload, network, and hardware
Throughput	Requests, transactions, packets, instructions, or data per second	Concurrency, payload, correctness and sustained duration
Resource utilization	CPU, accelerator, memory, storage, bandwidth	Capacity, sampling interval and bottleneck condition
Energy efficiency	Operations, transactions, or useful work per joule/watt	Power-measurement boundary and workload
Scalability	Performance as load/resources increase; speedup and efficiency	Strong/weak scaling rule, resource count and baseline
Reliability and availability	Failure rate, mission reliability, uptime, mean time to failure/repair	Observation period and whether the system is repairable
Fault tolerance and recovery	Successful recovery rate, data loss, recovery time	Injected fault model, severity and recovery criterion
Correctness and error	Error rate, test pass rate, numerical error	Oracle/ground truth, input coverage and tolerance
Machine-learning task quality	Accuracy, precision, recall, specificity, F1, calibration, task loss	Class balance, threshold, split, leakage and subgroup results
Security	Detection/false-alarm rate, attack success, breach rate, time to patch	Threat model, attacker capability and test coverage
Cost effectiveness	Cost per request/run/user or performance per cost	Pricing period, infrastructure boundary and utilization
Software maintainability	Change time, defect density, complexity, test coverage	Repository scope, change type and measurement rule

Formula and Terminology Checks

- **Effectiveness is not efficiency:** achieving a target differs from the resources consumed to achieve it.
- A generic `output/input × 100%` is not automatically meaningful; define compatible quantities and explain whether a percentage is appropriate.
- `|measured − true| / |true|` is relative error, not a universal definition of accuracy.
- **Precision is context-dependent:** measurement precision concerns repeatability, while classification precision is `TP / (TP + FP)`.
- `MTTF / (MTTF + MTTR)` is commonly used as an availability-style ratio; reliability concerns failure-free operation over a stated mission time. Do not use the terms interchangeably.
- Average response time alone can hide slow cases; report the slowest case, ranges, or percentiles when tail behavior matters.
- Classification accuracy alone may mislead with imbalanced classes; select metrics that reflect the error costs and research objective.
- A metric improvement is not sufficient if another required indicator becomes unacceptable; report trade-offs such as response time–throughput, accuracy–cost, security–usability, or performance–energy.

Computer Engineering Traceability Example

```
Problem: existing meal suggestions do not screen for stated allergy and budget constraints
→ Objective: design a machine-learning meal recommender that avoids prohibited ingredients and keeps cost within the stated
↪  budget
→ Literature/alternatives: rule-based filters, nutrition databases, collaborative filtering, constraint-based
↪  recommendation, or human dietitian review
→ Selection rationale: expected safety, usefulness, data availability, cost, explainability, maintenance effort, and
↪  legal/ethical constraints
→ Evaluation: fixed user profiles, ingredient database version, allergy rules, budget threshold, repeated tests, and human
↪  review of unsafe cases
→ Evidence: unsafe-recommendation count, budget violations, user-rated usefulness, preparation cost, and explanation quality
→ Output: recommendations satisfy the allergy and budget gates in the test set
→ Outcome: easier meal planning or improved eating satisfaction only if measured separately
```

Source inventory adapted and critically checked from Engineering Research Indicators. The source is useful for identifying candidate indicators, but every formula and term must be verified for the specific discipline and context before use.

26. Validity, Reliability, and Trustworthiness

- **Quantitative:** Construct, internal, external, statistical conclusion validity; reliability/repeatability.
- **Qualitative:** Credibility, transferability, dependability, confirmability.
- **Rule:** Use criteria appropriate to the paradigm; do not force one quality framework onto every study.

27. Procedure and Reproducibility

- **Record:** Ordered steps, materials/software/versions, parameters, roles/timing, decision rules, deviations, verification artifacts.
- **Check:** Could a competent person reproduce or audit the procedure?
- **Hunger example:** Record fasting condition, meal, schedule, context, and deviations.

28. Experimental Design, Controls, and Baselines

- **Ask:** What counterfactual does the comparison represent? Is the baseline simple, fair, and field-relevant? Are randomization, blinding, repetition, or split controls needed?
- **Check:** Does the comparison isolate the explanation relevant to the research question (RQ)?
- **Rule:** Add complexity only when the question requires it.

29. Analysis Plan

- **Pre-specify where possible:** Data preparation, descriptive summaries, primary/secondary outcomes, analysis method, missing/outlier handling, uncertainty, multiple comparisons, sensitivity/error analysis.
- **Check:** Does the analysis answer the research question (RQ) and match the data/design?
- **Common failure:** Analysis is chosen after seeing the desired result.

30. Ethics, Privacy, Legal, and Responsible Research

- **Check:** Consent, risk/benefit, vulnerable groups, an ethics committee or Institutional Review Board (IRB), sensitive data, anonymization, security/retention, copyright/license/terms of service, conflicts of interest, and responsible-AI disclosure.
- **Rule:** A simple topic does not automatically remove ethical obligations.
- **Research safety:** Protect participants and researchers during data collection, intervention, travel, food handling, device use, lab work, and fieldwork. A harmless-looking study can still create allergy, injury, privacy, embarrassment, coercion, or security risks.
- **Legal caution:** Check applicable law, institutional rules, IRB/ethics requirements, food and health claims, data-protection rules, copyright/license terms, platform terms of service, procurement rules, and safety regulations before collecting data or deploying a system.
- **AI caution:** Do not upload confidential thesis or participant data without authorization.

31. Research Plan, Feasibility, and Resources

- **Include:** Milestones, dependencies, data/access readiness, expertise, equipment/software, budget, risks/contingency, deliverables.
- **Check:** Can the study be completed with the stated resources and timeline?
- **Rule:** Feasibility does not prove scientific value, but an impossible design cannot answer the research question (RQ).

Part IV — Evidence and Reporting

32. Results

- **Must do:** Report evidence generated by the stated method and organize it by objective or research question (RQ).
- **May include:** Data/sample description, primary results, uncertainty, comparisons, null results, errors/-failures, sensitivity.
- **Rule:** Separate observation from interpretation.

33. Tables

- **Must have:** A specific purpose, descriptive title, defined columns/units/abbreviations/sample/conditions, consistent comparisons, uncertainty where relevant, attribution, and a reference in the text.
- **Placement rule:** Introduce and cite the table in the text before the reader reaches it. Do not place a table before the text that explains why it is there.
- **Check:** Do values agree with the narrative and calculations?
- **Common failure:** A table becomes a data dump, floats without explanation, or uses unexplained precision.

34. Figures and Visual Evidence

- **Must have:** A specific purpose, a useful caption, labels/axes/units/legend where relevant, attribution if adapted, and a reference in the text.
- **Text rule:** Explain in the narrative what the figure shows, what the reader should notice, and which claim the visual evidence supports. Do not leave interpretation entirely to the reader.
- **Check:** Appropriate visual form, axes, units, legends, uncertainty, sample size, accessible colors, honest scale, useful caption, no misleading processing.
- **Common failure:** A truncated or inconsistent scale exaggerates an effect.

35. Discussion

- **Must do:** Interpret results against the research question (RQ) or hypothesis, explain plausible mechanisms, compare prior work, consider alternatives, state meaning, and integrate limitations.
- **Common failures:** Repeating results, treating every outcome as success, ignoring contradictory evidence, introducing new unsupported results.

36. Conclusion

- **Must do:** Answer the research question (RQ) directly with strength and scope proportional to evidence.
- **Check:** Does it distinguish findings from uncertainty and avoid new unsupported claims?
- **Example:** “The score decreased for this case and condition,” not “This meal works for everyone.”

37. Contributions

- **Traceability:** `Problem → Prior Attempts/Evidence → Unresolved Need/Gap → Justification → Objective → Alternatives/Selection → Method → Result → Demonstrated contribution.`
- **Check:** Which evidence demonstrates each contribution?
- **Distinction:** A proposal describes expected contributions; a final report must demonstrate them.

38. Limitations in Final Interpretation

- For each limitation, state what result/interpretation is affected, which claim must be reduced, what evidence is missing, and what future work addresses it.
- **Rule:** Limitations are not a defensive paragraph placed at the end; they shape the conclusion.

39. Recommendations and Future Work

- **Must follow from:** A limitation, unresolved research question (RQ), or new evidence.
- **Must state:** What should be tested and why.
- **Common failure:** A disconnected wish list or policy recommendation unsupported by the design.

40. References and Appendices in the Final Document

- **Appendices may contain:** Instruments, procedures, extra tables, ethics documents, code/data dictionaries, supplementary analyses.
- **Rule:** Evidence essential to the main argument must remain in the main document.
- **Check:** Are references complete, verifiable, and consistent with citations?

41. Academic Writing and Terminology

- **Check:** Consistent terms, calibrated claim language, claim–evidence–reasoning paragraphs, disciplinary tense/voice, ethical citation/quotation.
- **Rule:** Language editing must not hide a conceptual weakness.

Part V — Cross-Section Consistency

42. Minimum Traceability Matrix

Table 6 is for checking cross-section consistency. Do not use it as a decorative summary table; trace each row back to actual evidence.

Table 8: Minimum traceability matrix for cross-section consistency

Trace element	What to check	Evidence to locate
Problem	What must be solved, in one central condition	Problem statement and supporting background
Prior attempts/evidence	Who tried what, what worked, what failed, and under which conditions	Literature review and source evidence
Unresolved need	What remains inadequate or uncertain, and why another study is still necessary	Gap/rationale paragraph
Objective/RQ	Whether the intended resolution answers the same problem	Objectives and research questions
Alternatives and selection	What could address the problem and why this option fits	Literature-derived alternatives and criteria
Method/evaluation	How the option is implemented and tested	Method, measurement, data, and analysis plan
Result/evidence	What was observed and whether it came from the planned method	Results, tables, figures, logs, or analysis
Output	Direct evidence that the defined problem is solved	Output statement tied to the objective
Outcome/by-product	What follows after resolution, including intended and unintended effects	Measured outcomes, side effects, cost, or burden
Limitation	What could not be achieved because of constraints	Limitations and bounded claims

Use one row per objective or research question (RQ). Look for:

- orphan problem — no objective addresses it
- literature list with no evidence-based path to an unresolved need
- claimed gap with no reason another study is necessary
- orphan objective — no problem/gap motivates it
- selected method with no literature-derived alternatives or comparison criteria
- intervention with no evaluation method
- result without planned evaluation
- conclusion without evidence
- limitation that invalidates an unqualified claim

43. Full Hunger-Example Traceability

Table 7 gives the complete “hungry → eat → satiated” teaching example so the problem, evidence, decision, and claim can be traced together.

Table 9: Full “hungry → eat → satiated” traceability example

Element	Teaching example	Review caution
Observation	Hungry after six hours	Anecdote alone is not a general problem
Problem	Hunger score is 8, above the satiety threshold	One declarative undesirable condition; no method
Gap	Not claimed before literature review	Never invent a gap
Objective	Reduce the score to ≤2 within 30 minutes	Directly resolves the stated problem
Alternatives	Different meals, portions, preparation, prices, and access conditions	Must be mapped and compared through literature
Selection rationale	Expected effect, safety, cost, access, time, suitability, and feasibility	Explains fit; does not prove the choice is universally best
Scope	One person, one selected meal, 60 minutes, self-reported hunger only	Deliberately chosen boundary
Solution/intervention method	Eat the selected specified meal	The chosen way to address the problem
Research/evaluation method	Repeated self-rating over 60 minutes	Tests whether the objective is achieved within a deliberately non-comparative scope; no validated scale stated
Result/evidence	Hunger scores at 0, 15, 30, and 60 minutes are $8 \rightarrow 4 \rightarrow 2 \rightarrow 3$	Hypothetical observations
Output	Score reaches 2 at 30 minutes	Defined problem solved for this case/time point
Outcome	Improved comfort or concentration	Not measured; expected outcome only
By-product	Leftovers, added cost, preparation time, or discomfort after eating	May matter for design value but is not the same as the main output
Conclusion	Defined output achieved for this case	No population or downstream-outcome claim
Limitation	No validated scale and incomplete control/measurement of contextual factors	Desired methodological strength could not be achieved; directly restricts the claim
Future work	Use a validated scale, improve contextual control, and measure the outcome if relevant	Must address actual limitations or a new research question (RQ)

Part VI — Using AI as a Reviewer

Review-Question Anatomy

```
Review target
→ Methodology criterion
→ Evidence and location
→ Cross-section comparison
→ Problem–Objective–Output–Outcome alignment
→ Missing vs weak/inconsistent
→ Blocking / Major / Minor
→ Human-verifiable action
```

Severity

- **Blocking:** The central research question (RQ), validity, evidence, or claim fails.
- **Major:** Evidence or alignment is inadequate and must be corrected.
- **Minor:** Clarity or presentation issue that does not change core validity.

Evidence-Bound AI Prompt

```
Act as a strict but constructive academic reviewer.

Review this chain:
Problem → Prior Attempts/Evidence → What Worked/Did Not and Where → Unresolved Need/Gap → Why This Study Is Necessary →
↪  Objective/RQ → Alternatives/Selection Rationale → Selected Method/Evaluation → Results/Evidence → Output → Outcome →
↪  Conclusion

For every issue:
1. Identify the section and evidence/location.
2. Separate missing information from weak or inconsistent information.
3. Classify severity as Blocking, Major, or Minor.
4. Explain the impact on validity.
5. Recommend a correction without inventing data, references, or results.

Cross-check the Abstract, Method, Tables/Figures, Results, and Conclusion.
Confirm that the Objective resolves the Problem, the Output demonstrates problem resolution, and the Outcome follows from
↪  the Output and is measured before it is claimed.
Check whether the literature reconstructs who addressed what, how, with what evidence, what worked or did not work under
↪  which conditions, what remains unresolved, and why the present study is still necessary. Also check whether it
↪  identifies plausible alternative solutions or methods, whether explicit criteria justify the selected option, and
↪  whether the evaluation method can demonstrate that the selected solution worked.
Check whether each hypothesis is one declarative and falsifiable claim with a pre-specified decision rule that permits an
↪  evidence-based `Supported: Yes/No` decision; require compound hypotheses to be decomposed.
For engineering or computer engineering research, check whether every primary indicator has an operational definition, unit,
↪  direction, baseline, target, workload/test condition, aggregation, repetitions, and uncertainty; verify that formulas
↪  and technical terms match their disciplinary meaning.

Apply five sharp tests before language editing:
1. Problem survival — determine whether the problem is real, current, evidenced, unresolved, and not already solved
↪  adequately by a simpler established approach.
2. Genuine failure — identify the pre-specified evidence that would show the hypothesis, selected method, or study failed.
3. Mandatory gates — separate must-pass acceptance criteria from optimization metrics and do not allow averages to hide a
↪  gate failure.
4. Measurement ceiling — verify the construct → indicator → instrument/ground truth → data → claim chain, including proxy
↪  validity.
5. No new problem — check trade-offs, affected stakeholders, unintended outcomes, transferred burdens, and mandatory
↪  harm-related guardrails.

Output:
Criterion | Evidence/Location | Prior Evidence/Unresolved Need/Study Necessity | Alternatives/Selection Rationale |
↪  Problem–Objective–Output–Outcome Link | Issue | Severity | Action
```

How to Use the Thesis Review Prompt Pack

Use the Thesis Review Prompt Pack to produce review notes and review questions, not to let AI decide whether a thesis is correct. The student-facing source is the Thesis Review Prompt Pack: Student Edition `v1.2`:

https://punpiti.github.io/thesis-review-prompt-pack/

Minimal workflow

1. Open `index.html` in the prompt pack folder.
2. Prefer the workspace workflow in `workspace.html` when reviewing a whole thesis or manuscript.
3. Put thesis files, manuscripts, references, figures, tables, results, scripts, logs, metrics, and baselines into the review workspace.
4. If files are mixed together, use the intake-initialization command shown in the prompt pack.
5. Ask the AI to prepare review instructions from the current work before full review.
6. Then use the short evaluation command shown in the prompt pack.
7. Read the review notes, verify every important comment against the source document, and fix Blocking issues before polishing language.
8. Keep or create a review-state note so the next review pass continues from known evidence instead of starting over.

What the prompt pack should check first

- Title and abstract alignment
- Problem significance and problem survival
- Literature synthesis and reference quality
- Objective/RQ/problem alignment
- Method justification, alternatives, and baselines
- Results, expected results, evidence, and contribution
- Traceability from problem to contribution
- Ethics, IRB, legal, data, and LLM-use responsibility
- Tables, figures, writing hygiene, terminology, and reader flow

Non-negotiable rules

- Use AI to create candidate review notes, not missing truth.
- Do not edit the thesis source before review notes exist, unless explicitly doing a later revision pass.
- Do not let AI invent references, DOI, venues, rankings, datasets, metrics, results, IRB status, or legal requirements.
- Treat LLM-generated information as unverified until checked against supplied documents, official pages, PDFs, data files, or other real sources.
- Every major criticism must identify a location in the manuscript when possible: chapter, section, page, paragraph, figure, or table.
- For references, separate manuscript evidence, reference evidence, and reviewer interpretation.
- If a source is missing, write `missing input`, `reference required`, or `needs verification`; do not guess.
- Review research logic before language polishing.

Use the prompt pack as a review cycle

```
Prepare context
→ Create review instructions
→ Run the short evaluation command
→ Produce review notes
→ Verify evidence manually
→ Fix Blocking issues
→ Recheck remaining Major issues
→ Polish language only after the logic survives
```

AI Review Rules

- AI output is a proposal for human verification, not an authoritative verdict.
- Verify every comment against the source document.
- Never let AI invent a reference, gap, dataset, result, approval, or calculation.
- Do not allow rewriting to hide broken research logic.
- Human reviewers retain academic judgment and responsibility.

Final Quick Checklist

- ☐ The problem is real, specific, significant, researchable, and evidenced.
- ☐ The gap follows from prior knowledge.
- ☐ The literature review explains who attempted what, what the evidence showed, what remains unresolved, and why this study is still necessary.
- ☐ Each objective directly resolves a problem stated in the problem statement.
- ☐ Output directly demonstrates that the problem was solved according to the objective.
- ☐ Outcome follows logically from the output and is not claimed unless measured.
- ☐ Objectives and research questions (RQs) align with the problem and contributions.
- ☐ Each hypothesis is declarative, atomic, falsifiable, and decidable as `Supported: Yes/No` using pre-specified evidence.
- ☐ The problem survives challenge: it is real, evidenced, unresolved, necessary to study, and not adequately solved by a simpler established approach.
- ☐ Study-level failure criteria state what evidence would show that the selected method did not solve the problem.
- ☐ Mandatory acceptance gates are separated from optimization metrics, and every gate passes.
- ☐ The construct–indicator–instrument/ground-truth–data–claim chain supports the conclusion; no unjustified proxy substitution occurs.
- ☐ The solution does not create or transfer an unacceptable new problem, harm, risk, or burden.
- ☐ Definitions, scope, assumptions, and limitations are distinct.
- ☐ Design, data, measurement, and analysis can answer the research question (RQ).
- ☐ Engineering indicators, when applicable, specify units, baselines, targets, test conditions, aggregation, repetitions, uncertainty, and relevant trade-offs.
- ☐ References are real, relevant, and consistent.
- ☐ Tables/figures agree with the narrative.
- ☐ Results answer the objectives.
- ☐ Conclusions and recommendations do not exceed evidence.
- ☐ Ethics, privacy, legal, and responsible-AI issues are addressed.
- ☐ Every major claim is traceable through the research argument chain.