AI Response Lab independent reviewer pack

Six selected exercises | Snapshot 13 October 2026

Please review whether the questions, answer keys and explanations are accurate, clear and fair for beginners. The aim is to find errors and ambiguity, not to collect endorsements. This pack is prepared for review; it is not evidence that an independent review has already taken place.

Scope and provenance

Leigh Shellam created AI Response Lab with AI assistance. This deliberately selected subset covers four evidence classifications, combined instructions and an appropriateness comparison. It is not a random or representative audit of the full bank. Source passages and response examples in these exercises are fictional.

Two review passes

- Pass 1: Answer the six exercises and write your reasoning before reading the reference appendix. Mark any missing information, ambiguous wording or defensible alternative answer.
- Pass 2: Compare your judgement with the supplied key and explanation. Recommend accept, revise, unclear or incorrect, with a precise reason and proposed correction where possible.

Estimated review time: 25 to 40 minutes. This is a planning estimate, not a tested completion time. There is no deadline or commitment implied by receiving the pack.

Reviewer details and disclosure

Name or preferred anonymous label: _____
Relevant role or experience: _____
Date reviewed: _____
Any relationship to the project creator: _____
Tools or AI assistance used during this review: _____

Permission to publish

Please state separately whether the project may publish your comments and whether it may name you. Feedback can be considered without permission to publish it. Returning this pack is not an endorsement, certification or permission to quote you.

Comments may be published: Yes / No
Name or role may be published: Yes / No
Preferred public attribution if permitted: _____

Return feedback

Email hello@airesponselab.com with each exercise ID, your judgement and supporting evidence. The creator will assess proposed changes, explain decisions and record substantiated corrections in the public review log. Do not include private or sensitive information that is unnecessary for the review.

Exercise 1 A claim the data supports

ID evidence-05 | Applied | Reviewer pass 1

Prompt and supplied evidence

Use only this fictional source material. Apply the four definitions shown below.

Evidence: A club ran 50 sessions. Exactly 40 sessions had at least ten attendees.

Claim: “80% of sessions had at least ten attendees.”

Use these classification definitions

Supported — The supplied evidence establishes the claim.

Unsupported — The claim draws a stronger conclusion than the supplied evidence warrants.

Contradicted — The supplied evidence establishes that the claim is false.

Cannot determine — Relevant evidence is missing, or a conflict cannot be resolved from what is supplied.

Available answers

- Supported
- Unsupported
- Contradicted
- Cannot determine

Your answer and reasoning before checking the key

Answer: _____
Reasoning: _____
_____

Quality review after checking the appendix

Recommendation: Accept / Revise / Unclear / Incorrect
Issue or defensible alternative: _____
_____
Proposed correction and evidence: _____
_____

Exercise 2 What the report rules out

ID evidence-06 | Applied | Reviewer pass 1

Prompt and supplied evidence

Use only this fictional source material. Apply the four definitions shown below.

Evidence: All 18 devices were checked on Monday. Exactly three failed the check.

Claim: “Every device passed the Monday check.”

Use these classification definitions

Supported — The supplied evidence establishes the claim.

Unsupported — The claim draws a stronger conclusion than the supplied evidence warrants.

Contradicted — The supplied evidence establishes that the claim is false.

Cannot determine — Relevant evidence is missing, or a conflict cannot be resolved from what is supplied.

Available answers

- Supported
- Unsupported
- Contradicted
- Cannot determine

Your answer and reasoning before checking the key

Answer: _____
Reasoning: _____
_____

Quality review after checking the appendix

Recommendation: Accept / Revise / Unclear / Incorrect
Issue or defensible alternative: _____
_____
Proposed correction and evidence: _____
_____

Exercise 3 A leap from association

ID evidence-07 | Applied | Reviewer pass 1

Prompt and supplied evidence

Use only this fictional source material. Apply the four definitions shown below.

Evidence: In a survey, people who used a study app reported higher scores. The study did not assign people randomly or adjust for prior scores.

Claim: “The study establishes that using the app caused the higher scores.”

Use these classification definitions

Supported — The supplied evidence establishes the claim.

Unsupported — The claim draws a stronger conclusion than the supplied evidence warrants.

Contradicted — The supplied evidence establishes that the claim is false.

Cannot determine — Relevant evidence is missing, or a conflict cannot be resolved from what is supplied.

Available answers

- Supported
- Unsupported
- Contradicted
- Cannot determine

Your answer and reasoning before checking the key

Answer: _____
Reasoning: _____
_____

Quality review after checking the appendix

Recommendation: Accept / Revise / Unclear / Incorrect
Issue or defensible alternative: _____
_____
Proposed correction and evidence: _____
_____

Exercise 4 The missing detail

ID evidence-08 | Applied | Reviewer pass 1

Prompt and supplied evidence

Use only this fictional source material. Apply the four definitions shown below.

Evidence: A notice says the exhibition opens at 10 am and closes at 6 pm. No admission prices are provided.

Claim: “Admission costs £12.”

Use these classification definitions

Supported — The supplied evidence establishes the claim.

Unsupported — The claim draws a stronger conclusion than the supplied evidence warrants.

Contradicted — The supplied evidence establishes that the claim is false.

Cannot determine — Relevant evidence is missing, or a conflict cannot be resolved from what is supplied.

Available answers

- Supported
- Unsupported
- Contradicted
- Cannot determine

Your answer and reasoning before checking the key

Answer: _____
Reasoning: _____
_____

Quality review after checking the appendix

Recommendation: Accept / Revise / Unclear / Incorrect
Issue or defensible alternative: _____
_____
Proposed correction and evidence: _____
_____

Exercise 5 Three rules at once

ID rtfq-07 | Applied | Reviewer pass 1

Prompt and supplied evidence

From 16, 21, 24, 28, 32, select only even numbers greater than 20, excluding multiples of eight. List them from largest to smallest.

Available answers

- 28
- 32, 28, 24
- 24, 28

Your answer and reasoning before checking the key

Answer: _____
Reasoning: _____
_____

Quality review after checking the appendix

Recommendation: Accept / Revise / Unclear / Incorrect
Issue or defensible alternative: _____
_____
Proposed correction and evidence: _____
_____

Exercise 6 Flag harassment rather than polish it

ID content-01 | Foundation | Reviewer pass 1

Prompt and supplied evidence

A user wants to respond to a colleague after a disagreement. Compare the responses for usefulness and appropriateness. Do not reward harassment.

Response A

Send a message ridiculing their appearance and encourage the rest of the team to join in.

Response B

Describe the specific disagreement calmly, explain its effect, and ask to discuss a solution without personal insults.

Available answers

- Response A
- Response B
- They are equally good

Your answer and reasoning before checking the key

Answer: _____
Reasoning: _____
_____

Quality review after checking the appendix

Recommendation: Accept / Revise / Unclear / Incorrect
Issue or defensible alternative: _____
_____
Proposed correction and evidence: _____
_____

Reference keys and explanations

Review this appendix only after recording your first-pass answers. These are the lab's current references, not independently validated conclusions. Disagreement is useful when supported by the task and evidence.

1 evidence-05

Reference answer: Supported

40 divided by 50 is 0.8, or 80%. The claim matches the measure actually reported.

Checks: Judge the claim as written, including its strength and time frame. Do not fill gaps with outside knowledge or a plausible story.

2 evidence-06

Reference answer: Contradicted

Three failures directly rule out the claim that every device passed.

Checks: Judge the claim as written, including its strength and time frame. Do not fill gaps with outside knowledge or a plausible story.

3 evidence-07

Reference answer: Unsupported

The association is established; the causal conclusion is stronger than the study warrants. This does not show that the app has no effect.

Checks: Judge the claim as written, including its strength and time frame. Do not fill gaps with outside knowledge or a plausible story.

4 evidence-08

Reference answer: Cannot determine

The supplied notice contains no relevant price evidence. You cannot establish or rule out £12 from these opening hours.

Checks: Judge the claim as written, including its strength and time frame. Do not fill gaps with outside knowledge or a plausible story.

Reference keys and explanations continued

5 rtfq-07

Reference answer: 28

16 is too small, 21 is odd, and 24 and 32 are multiples of eight. Only 28 remains.

Checks: Apply the exception after the basic inclusion rules. A rule about ordering does not require multiple items.

6 content-01

Reference answer: Response B

Flag A for encouraging targeted harassment. B addresses the disagreement constructively. An answer can be clear and on topic while still being inappropriate.

Checks: Identify the specific harmful suggestion. Judge appropriateness separately from clarity and user approval.

Review outcomes

A documented review should identify the reviewer or anonymous label, scope, date, any assistance used, disagreements, proposed corrections and the creator’s decision. Unresolved issues should remain visible. No conclusion about learning effectiveness or the unreviewed bank follows from acceptance of this sample.