

09. Add Quality Harnesses And Gold Examples

Quality | 90 minutes

A harness lets the learner stop trusting vibes. It checks whether listing outputs meet a quality bar and catches repeated mistakes.

Do This Now

- Read the listing rubric.
- Create a gold set with good, bad, and borderline listing examples.
- Ask Codex to evaluate 10 listings against the rubric.
- Turn any repeated failure into a regression case.

Where To Paste This

Paste this into Codex inside the starter workspace.

Prompt

Create a simple listing quality harness.

Inspect:

- rubrics/listing-quality-rubric.md
- examples/good-listing.md
- examples/bad-listing.md
- reports/batch-10-listings.md if it exists

Create:

- reports/listing-quality-harness.md
- examples/borderline-listing.md
- reports/listing-regression-cases.md

The harness should check:

1. no invented facts,
2. target buyer is clear,
3. strongest feature is concrete,
4. generic adjectives are reduced,
5. verification warnings are present,
6. rewrite is different enough from the original,
7. human approval is required before publishing.

Evaluate 10 sample listings and name the repeated failure patterns.

What You Should See

Codex should create a harness document and identify repeated failure types. It should not claim the workflow is perfect.

If It Goes Wrong

- If it only gives scores, ask for specific evidence and failure types.

- If it approves everything, ask it to apply an adversarial reviewer lens.
- If the rubric is too vague, ask for measurable checks and examples.

Practical Output

A listing quality harness, a borderline example, and first regression cases.

Codex task card	Details
Files to inspect	rubrics, examples, reports.
Expected files	reports/listing-quality-harness.md, examples/borderline-listing.md, reports/listing-regression-cases.md.
Verify	Repeated failures become named cases, not vague advice.
Follow-up	Update the listing improver skill so it explicitly avoids the top two regression failures.