

Research reproducibility checklist

Freeze the question, inputs, decision timestamps and comparison rules before you evaluate an agentic experiment.

Before the run

- Write one question and one decision you will make from the result.
- Save the baseline, candidate configuration and repository commit.
- Freeze input files; record a SHA-256 checksum and retrieval time.
- Record when each observation occurred and when it became available to the decision.
- Define missing data, stale data, fees and retry handling in advance.
- Separate exploratory examples from a held-out evaluation.

During the run

- Record the exact command, dependency versions and random seed where relevant.
- Keep excluded observations and reasons alongside admitted observations.
- Compare both rules on the same frozen evidence and cost assumptions.
- Keep failures and uncertain outcomes visible. A timeout is not evidence that an external action failed.

Before sharing

- Re-run from a clean environment.
- Ask another person to follow the command without undocumented steps.
- Mark synthetic inputs and worked examples clearly.
- Report sample size, exclusions, costs and limitations alongside the result.
- Keep secrets, account identifiers and private customer data out of the package.

Fill in your experiment

Question: _____

Commit and configuration: _____

Input file and checksum: _____

Decision-time availability rule: _____

Baseline and candidate: _____

Held-out window: _____

Run command: _____

Observed result and limitations: _____

This checklist supports engineering reproducibility. It does not establish trading profitability or predict an investment outcome.

Subscribe for the next practical issue

Subscribe to this newsletter

Free download. Email signup is optional and requires confirmation. Teaching examples are hypothetical.