

AI Tool Evaluation Scorecard

Use this worksheet to compare 2 to 5 AI tools or models for one real task.

This is a quick decision worksheet, not a full buyer kit.

Fill it out during or right after testing so your notes stay grounded in evidence instead of hype.

1) Decision context

Task being evaluated:

User type or team context:

Must-have requirement:

Main constraint (circle or mark one): budget / privacy / speed / reliability / team fit

Optional context note:

2) Candidate list

#	Tool or model name	Plan or pricing tier considered	Deployment mode, if relevant
1			
2			
3			
4			
5			

3) Evidence capture

Use this section to separate direct observations from weaker evidence.

Candidate 1: _____

- What was actually tested:
- Date/version:
- Source of evidence (mark one or more): firsthand use / trusted review / vendor claim / benchmark / peer input
- Confidence note:

Candidate 2: _____

- What was actually tested:
- Date/version:
- Source of evidence (mark one or more): firsthand use / trusted review / vendor claim / benchmark / peer input
- Confidence note:

Candidate 3: _____

- What was actually tested:
- Date/version:
- Source of evidence (mark one or more): firsthand use / trusted review / vendor claim / benchmark / peer input
- Confidence note:

Candidate 4: _____

- What was actually tested:
- Date/version:
- Source of evidence (mark one or more): firsthand use / trusted review / vendor claim / benchmark / peer input
- Confidence note:

Candidate 5: _____

- What was actually tested:
- Date/version:
- Source of evidence (mark one or more): firsthand use / trusted review / vendor claim / benchmark / peer input
- Confidence note:

4) Tradeoff scoring

Score each tool from 1 to 5.

- 1 = poor fit
- 3 = usable with tradeoffs
- 5 = strong fit for this specific task

Tool or model	Task fit	Output quality	Speed / responsiveness	Reliability	Cost	P
---	---	---	---	---	---	---

5) Notes that prevent bad decisions

Candidate 1: _____
- Best use case:
- Biggest weakness:
- Deal-breaker risk:
- Open question before buying:

Candidate 2: _____
- Best use case:
- Biggest weakness:
- Deal-breaker risk:
- Open question before buying:

Candidate 3: _____
- Best use case:
- Biggest weakness:
- Deal-breaker risk:
- Open question before buying:

Candidate 4: _____
- Best use case:
- Biggest weakness:
- Deal-breaker risk:
- Open question before buying:

Candidate 5: _____
- Best use case:
- Biggest weakness:
- Deal-breaker risk:
- Open question before buying:

6) Decision output

****Current front-runner:****

****Why it leads right now:****

****Next action**** (choose one): test deeper / buy / reject / keep comparing

****Re-evaluate later if this changes:****

Quick use rule

Do not treat this as an objective leaderboard.

Use it to compare a short list against one real task, based on actual evidence and clear tradeoffs.

If you need more than this

If you need weighted scoring, category-specific versions, deeper test planning, or a stakeholder-ready summary, the ****AI Tool Evaluation Kit**** is the next step. It extends this scorecard with a weighted scoring worksheet, test-run template, and recommendation memo.