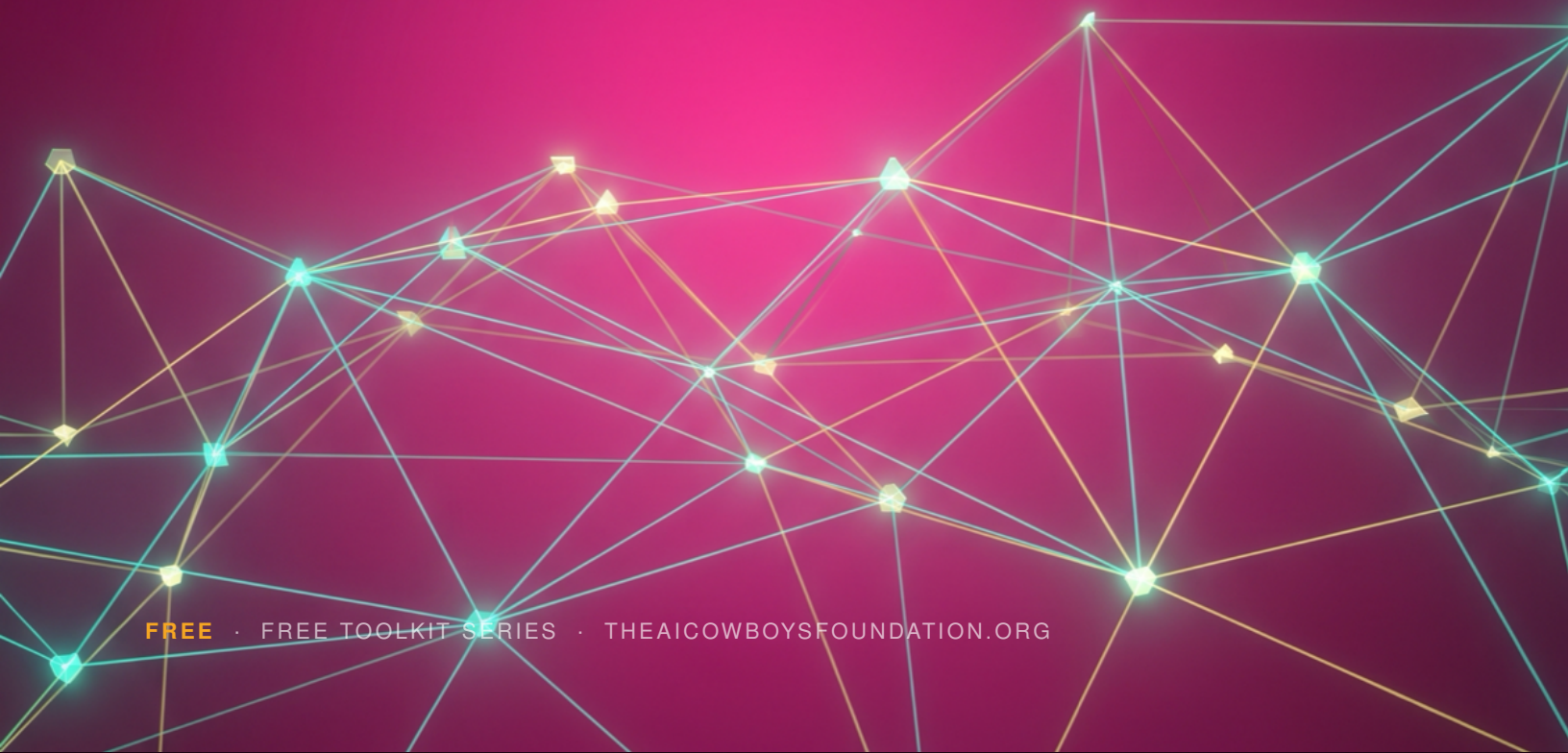
THE AI COWBOYS FOUNDATION

TOOLKIT 03 · BUYING

AI Vendor Evaluation Sheet

Compare AI tools on the same weighted criteria and choose with evidence instead of hype.



AI Vendor Evaluation Sheet



The demo rewards polish. A scorecard rewards what you will actually live with.

Compare AI tools on the same criteria so you can choose with evidence instead of hype. Score each option from 1 to 5 on every row, add the weighted total, and keep the sheet for your records.

Why a scoring sheet beats a demo

Every AI vendor demo looks impressive. That is what demos are for. The problem is that demos showcase the best case while your team will live with the average case, the support queue, and the invoice. A weighted scoring sheet forces every option through the same questions, and it leaves you a written record of why you chose what you chose. When leadership or an auditor asks later, you have an answer.

WHAT THE DATA SAYS ABOUT AI PURCHASES

Stanford University's 2025 AI Index found that while 78 percent of organizations now use AI, the financial returns are usually modest: most companies reporting cost savings from AI put those savings under 10 percent, and the most common revenue lift is under 5 percent.

The lesson: the tool alone rarely transforms the business. Value comes from fit, adoption, and honest measurement, which is exactly what this sheet forces you to evaluate before you sign.

8

weighted criteria score every tool the same way

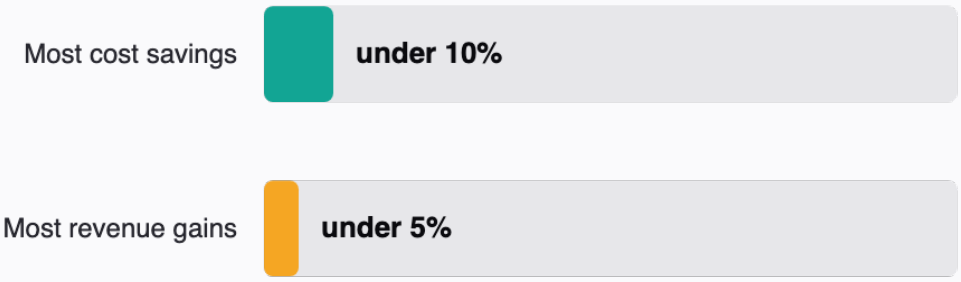
60

is the minimum total score before you should buy

<10%

typical cost savings, so fit matters more than flash

Typical AI impact per function



Source: Stanford HAI, 2025 AI Index

Returns are usually modest, so fit and adoption decide the winner. Most organizations report cost savings under 10 percent and revenue gains under 5 percent per function. Source: Stanford HAI, 2025 AI Index.

How to score

1. Rate each criterion from 1 (poor) to 5 (excellent).
2. Multiply each score by its weight.
3. Add the weighted scores for a total out of 100.
4. Anything under 60 needs a second look before you buy.

Scoring grid

Criterion	Weight	What to look for	Score 1 to 5	Weighted
	5			

Criterion	Weight	What to look for	Score 1 to 5	Weighted
Security and data handling		Where data is stored, who can see it, encryption, a signed data agreement, no training on your data by default		
Accuracy and reliability	4	Quality on your real tasks, how it handles errors, how often it is wrong		
Privacy and compliance	4	Fit with your privacy duties, data retention controls, audit options		
Cost and value	3	Total cost including seats and usage, and the time it actually saves		
Ease of use	3	How quickly your team can learn it, quality of onboarding		
Support and reliability	2	Response times, uptime record, help when something breaks		
Vendor stability	2	Company track record, funding, and likelihood of being around next year		
Control and exit	2	Your ability to export data and leave without penalty		

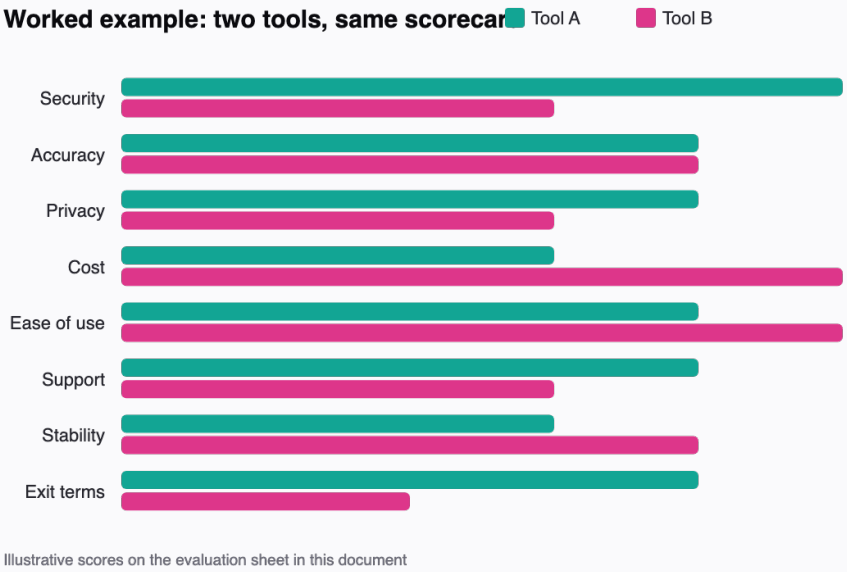
Total weighted score: _____ out of 100

Questions to ask every vendor

- Where is our data stored, and who can access it?
- Do you train your models on our data? Can we turn that off?
- What happens to our data if we cancel?
- Can you provide a signed data processing agreement?
- What is your uptime record and support response time?
- How do you test for and reduce bias and errors?
- Which recognized framework does your risk process follow, such as the NIST AI Risk Management Framework?

A worked comparison

Two tools can demo equally well and score very differently once you weight what matters. Here is the same scorecard applied to two real options.



Illustrative scores on the evaluation sheet in this document

Same criteria, different winners. Tool B scores higher on cost and ease, but Tool A wins on security and privacy, which carry the most weight. The scorecard makes the tradeoff visible instead of leaving it to the demo.

WORKED EXAMPLE: READING THE SCORES

Tool B looked better in the demo: cheaper, easier, slicker. But on the weighted sheet, Tool A wins. Security (weight 5) and privacy (weight 4) are where Tool A scores 5 and 4, while Tool B scores 3 and 3. Because those criteria carry the most weight, Tool A's total comes out ahead even though Tool B felt nicer to click through.

The point: the demo rewards polish. The scorecard rewards what you will actually live with. Trust the sheet.

Red flags worth walking away from

- The vendor cannot say plainly where your data lives or who can see it.
- Training on your data is on by default and cannot be turned off.

- Pricing only makes sense at usage levels far beyond your real workload.
- There is no export path. If you cannot leave, you are not a customer, you are collateral.

Decision record

Chosen tool: _____ Score: _____ Reviewed by: _____ Date: _____

Runner up: _____ Score: _____ Reason not chosen:

_____

Sources and further reading

- NIST AI Risk Management Framework, National Institute of Standards and Technology
- 2025 AI Index Report, Economy chapter, Stanford Institute for Human Centered AI
- Texas DIR AI templates and resources, Texas Department of Information Resources

PUT THIS TO WORK

Want help applying this in your organization?

The AI Cowboys Foundation delivers free briefings, workshops, and readiness assessments for business, government, classrooms, and veteran programs. Tell us what you are working on and we will point you to the right next step.

Contact us

contact_us@theaicowboys.com · theaicowboysfoundation.org/contact

The AI Cowboys Foundation is a Texas nonprofit organization based in San Antonio, Texas. This document is provided free for education. Adapt it to your organization and have counsel review policies before you rely on them. Veteran founded, Texas built.