



AI vendor selection and contracting

Choosing and contracting AI tools and partners with clear tests, terms, exit routes and accountability.

WORKS WITH THE ASSISTANTS TEAMS ALREADY USE



EDITION 39 / EMBER & VIOLET

EXECUTIVE WORKING EDITION - 08 OCT 2026

39

AT A GLANCE

Choosing and contracting AI tools and partners with clear tests, terms, exit routes and accountability.

AI products change quickly, marketing claims are hard to compare and the contract terms that matter most, such as how your data is used and what happens when you leave, are often buried or left at defaults. This playbook gives a disciplined route from need to signed agreement: define the need before looking at products, move from a long list to a short list, run scripted pilots with scoring, ask the right security and data questions, review key contract terms with counsel, spot red flags and manage the vendor after signature. It is a practical guide for business and procurement teams, not legal advice.

WHO IT IS FOR

Procurement, IT, security, legal, privacy and business owners who select, buy or renew AI tools and AI services, and the executives who sign off on them.

TIME TO USE

70 minutes to read; use as a working checklist across a 6 to 12 week selection

What you will have at the end

- A written needs definition with must-have and nice-to-have criteria
- A long-list to short-list method that screens quickly and fairly
- A scripted pilot with a scoring sheet that uses your own data and tasks
- A set of security, data and contract questions ready for counsel
- A vendor management routine that continues after signature

Contents

01	Why AI buying differs from ordinary software buying	03
02	Define the need before you meet vendors	04
03	From long list to short list	06
04	Scripted pilots and scoring	08
05	Security and data questions to put in writing	09
06	Contract terms to review with counsel	11
07	Red flags in selection and in the contract	13
08	Manage the vendor after signature	15
09	A first quarter that produces evidence	16
10	Terms used in this publication	17

EXECUTIVE BRIEF

Why AI buying differs from ordinary software buying

01

K'

The product can change under you, the outputs are probabilistic and your data may be an input, an output and a risk at the same time.

Traditional software does the same thing for the same input. AI products often do not. Outputs vary, quality depends on your data and tasks, and the vendor may update models and features frequently, sometimes changing behavior without a visible release. Many products also rely on other providers underneath, so the party you contract with may not be the party that processes your data. Demonstrations are tuned to succeed. These features make a conventional feature checklist insufficient. You need tests on your own work, clear terms on data, and a plan for change.

The same features make good contracting more valuable. Where the technology moves fast, the agreement is the one stable thing that fixes what the vendor may do with your information, what quality and availability you can expect, how you will be told of changes and how you leave. Selection and contracting should therefore run as one process, with legal, security and privacy staff involved early rather than at the end, when momentum makes it hard to raise concerns.

- Define the need and the success measures before you look at products
- Test on your own data and tasks, not on vendor demonstrations
- Treat data use, retention and exit as selection criteria, not as late legal details
- Ask who the sub-processors and underlying model providers are
- Plan for the product to change, and require notice of material changes
- Keep the ability to leave: export, portability and a tested exit plan

😊

PLAIN TEST

If this vendor changed its model, raised its price by a large amount or went out of business next quarter, what would you do on the following Monday? If you have no answer, the selection is not finished.

WORKING QUESTION

Who in legal, security and privacy has already seen this purchase, and what did they ask that you could not yet answer?

NEEDS

Define the need before you meet vendors

02

A one-page requirements brief keeps the process honest and stops vendors from defining the problem for you.

Start with the work, not the tool. Describe the process, the users, the volume, the inputs, the outputs, the current pain and how success will be measured. Separate the problem from the proposed solution: the brief should say we need to cut proposal drafting time without lowering quality, not we need a generative platform. List the systems the tool must connect to, the data it will touch and the constraints it must respect, including sector rules, regional data requirements and internal policies. Then rank criteria so the team does not rediscover its priorities during scoring.

BRIEF SECTION	CONTENTS	OWNER
Problem and outcome	Process, users, volume, pain, measurable target	Business owner
Data and systems	Sources, sensitivity classes, integrations, identity	IT and data lead

BRIEF SECTION	CONTENTS	OWNER
Risk and compliance	Applicable policies, sector rules, regional requirements to check	Security, privacy, legal
Must-haves	Pass or fail criteria that rule a vendor out	Cross-functional team
Weighted criteria	Scored items with weights summed to 100	Cross-functional team
Budget and timeline	Pilot, rollout, run cost range and decision date	Sponsor and finance
Exit conditions	Data export, portability, transition support needed	Procurement and legal

Must-have candidates to consider

- Data is not used to train vendor models without your explicit written agreement
- Single sign-on and role-based access that match your identity system
- Audit logs you can export
- Region or residency options if your obligations require them
- Ability to export your data and configuration in usable formats
- Human review or approval steps for outputs that carry risk
- Clear support and incident communication commitments

COMMON MISTAKE

Letting the first impressive demo set the criteria. Write and weight criteria before any vendor presents, and share them with every vendor in the same form.

WORKING QUESTION

Which of your must-have criteria would immediately remove the vendor you currently like best, and is it written down?

LONG LIST TO SHORT LIST

From long list to short list

03

Screen widely and cheaply on documented facts, then invest effort only in the few vendors that pass.

Build a long list from several sources: peer organizations, analyst coverage you trust, your existing vendors' roadmaps, open marketplaces and a small number of specialist firms. Include your current providers, who may offer a capable option with lower integration cost, and include at least one non-AI alternative so the comparison stays honest. Resist the temptation to long-list dozens; ten to fifteen is manageable. Then screen by sending the same short questionnaire to all and reviewing public documentation, security reports and terms.

01 Collect

Gather ten to fifteen candidates across product types, including current vendors and a non-AI option.

02 Screen on must-haves

Apply the pass or fail criteria using documents and a short written questionnaire. Remove those that fail, and record why.

03 Check basics

Review the vendor's security documentation, standard terms, sub-processor list, financial stability indicators and reference customers in similar settings.

04 Hold structured calls

Give each remaining vendor the same agenda and ask them to show the product on a scenario from your brief, not their default demo.

05 Choose three to four

Rank by weighted criteria and advance a short list to scripted pilots.

SCREEN	QUESTION	EVIDENCE
Fit	Does it address our process and users?	Scenario walk-through on our example

SCREEN	QUESTION	EVIDENCE
Data	Where is data stored, processed and who can access it?	Documentation, sub-processor list, written answers
Security	What independent assurance exists?	Current third-party reports; ask for scope and date
Viability	Will they be here and supporting in three years?	References, ownership, product history, concentration risk
Commercial	Is pricing predictable at our volume?	Price model with usage scenarios
Exit	Can we leave with our data?	Export formats, termination assistance terms

When you speak to references, ask questions that reveal operations, not satisfaction: how long did the rollout take compared with the plan, what broke, how did the vendor respond to an incident, what did the first renewal look like, what would you negotiate differently. Ask for reference at a customer of similar size and sector, and for one that has used the product for more than a year.

WORKING QUESTION

On what documented evidence, not on a demonstration, did each vendor on your short list earn its place?

PILOTS

Scripted pilots and scoring

04

A scripted pilot gives every vendor the same tasks, the same data and the same scoring, so differences reflect the product and not the presenter.

Design the pilot before inviting vendors. Choose a set of real tasks, typically thirty to fifty, that cover common cases, difficult cases and cases where the right answer is to decline or escalate. Prepare reference answers or acceptance rules. Decide what data may be used, using synthetic or de-identified data where possible, and put basic data terms in place before any real data is shared. Fix the duration, the users, the support available and the criteria for passing. Run vendors in the same period where you can, and blind the scorers to which vendor produced which output.

PILOT DIMENSION	HOW TO TEST	EVIDENCE
Output quality	Score outputs on the test set against a rubric by two reviewers	Scores, disagreement notes, failure examples
Reliability	Repeat the same tasks across days and at peak times	Variation in results, errors, outages
Fit to workflow	Real users complete tasks end to end	Task time, steps, user feedback
Safety behavior	Include tricky inputs, sensitive data and attempts to extract restricted content	Refusals, leakage, handling of edge cases
Admin and control	Configure roles, logging, retention and export	Admin time, missing controls
Support	Raise realistic questions and issues	Response time and quality
Cost	Meter usage during the pilot	Actual cost per task at pilot volume

Scoring sheet essentials

- Criteria and weights fixed before the pilot starts

- Each criterion has a definition and a 1 to 5 scale with anchors
- Two scorers per item with disagreements discussed
- Must-have failures recorded separately and not averaged away
- Failure examples captured verbatim for the vendor to respond to
- A separate score for risk findings that the weighted total could hide

COMMON MISTAKE

Averaging away a serious flaw. A product that scores high on features but leaks restricted data in one test should not win on its total. Keep pass or fail items outside the weighted score.

WORKING QUESTION

Could an outsider look at your scoring sheet and explain exactly why one vendor beat another?

SECURITY AND DATA

Security and data questions to put in writing

05

Ask specific questions, get written answers, and check that the answers match the contract. Marketing pages are not commitments.

Data questions matter most in AI because your information may be used as a prompt, stored for logs, used for tuning or sent to another provider. Ask where each data type goes, who can see it, how long it stays and what can be deleted. Ask whether the vendor uses customer content to train or improve models, whether that is off by default and whether it can be turned on only with your written consent. Ask for the list of sub-processors and underlying model providers and how you are told when it changes. Align the questions with your own classification of data.

TOPIC	QUESTION TO ASK	WHAT A GOOD ANSWER LOOKS LIKE
Training use	Is our content used to train or improve any model?	No by default, stated in the contract, with scope covering sub-processors
Retention	How long are prompts, outputs and logs kept, and can we set it?	Configurable retention, deletion on request, deletion on exit with confirmation
Location	Where is data stored and processed, including support access?	Named regions with commitments, and disclosure of support access from elsewhere
Access	Who at the vendor can see our data and under what controls?	Role-limited, logged, justified access with customer visibility
Sub-processors	Who else touches our data?	Current list, advance notice of changes, flow-down of obligations
Assurance	What independent assessments exist?	Current reports with scope and date; willingness to share under confidentiality
Incidents	How and how fast will we be told a breach?	Defined notice time and content, named contact, cooperation duty
Customer controls	Can we restrict features, set roles and export logs?	Admin controls and exportable audit logs

Also test technical safeguards that matter for AI specifically. Ask how the product handles prompt injection, where hidden instructions in documents or web pages try to steer the model. Ask what limits exist on actions the system can take on your behalf, such as sending messages or changing records, and whether those actions require approval. Ask how access permissions in your source systems are respected in retrieval. For regulated data or high-risk uses, check current legal and sector requirements, such as data protection law and any applicable AI regulation, with your advisers before the pilot uses real data.

- Ask for written answers signed by someone accountable, and keep them in the vendor file
- Ask for the questionnaire answers to be incorporated by reference into the contract where possible
- Compare the product's default settings with the settings you need and who controls them
- Check whether features marked beta or preview carry weaker commitments
- Re-ask the questions at renewal, because architectures and sub-processors change

WORKING QUESTION

Can you state, for each data type the tool will touch, where it goes, how long it stays and who at the vendor can see it?



CONTRACT TERMS

Contract terms to review with counsel

06



These are the points that most often decide who bears the cost when something goes wrong. Have qualified counsel review the actual language for your jurisdiction.

This section is a map for discussion with your lawyers, not legal advice. Standard vendor terms are written to protect the vendor, and many terms are negotiable, particularly for larger or strategic purchases. Where the vendor cannot move, decide whether you can accept the risk, mitigate it by how you use the product, or must walk away. Keep a log of requested changes and outcomes. Check that order forms, data processing terms, security addendums and online terms do not conflict, and that the order of precedence is clear.



TERM	WHAT TO LOOK FOR	RISK IF MISSED
Data use and ownership	Your content and outputs remain yours; no training on them without consent; limits on use of derived data	Your information improves a product used by competitors
Retention and deletion	Set periods, deletion on request on exit, certification of deletion	Data lingers in logs or backups with no end date
Intellectual property	Who owns outputs; rights to use them; handling of third-party material in outputs	Unclear right to use or defend what the tool produces



TERM	WHAT TO LOOK FOR	RISK IF MISSED
Indemnity	Vendor defends and compensates for certain third-party claims, such as IP claims about outputs, subject to conditions	You carry claims that arise from the vendor's product
Liability limits	Caps, exclusions and carve-outs, particularly for data breach and confidentiality	A cap far below the realistic loss
Service levels	Availability, support response, credits, and how measured	Outages with no remedy
Change notice	Advance notice of material change to models, features, sub-processors and terms; right to terminate if the change harms you	Product behavior shifts and you find out from users
Exit and portability	Export in usable formats, assistance period, deletion, continued access during transition	Locked in, or losing history and configuration

Points that deserve a second look

- Order of precedence between master terms, order forms and linked online terms that the vendor can update
- Unilateral change rights: terms that may change by posting a notice on a website
- Auto-renewal and price escalators, and notice windows for non-renewal
- Usage-based pricing: overage rates, minimum commitments and what counts as a unit
- Audit rights, regulatory cooperation and support for your own compliance obligations
- Subcontracting and assignment, including change of control of the vendor
- Acceptable use restrictions that could limit your intended use

COMMON MISTAKE

Negotiating price hard and accepting default data and exit terms. The financial effect of a data incident or a forced migration can exceed several years of fees.

WORKING QUESTION

Which two contract terms would hurt you most if the vendor were acquired by a company you would not choose to work with?

RED FLAGS

Red flags in selection and in the contract

07

Some signals should slow you down or end the process. Learn them, and write them into your review routine.

RED FLAG	WHY IT MATTERS	RESPONSE
Cannot or will not state how customer data is used	Unknown training and retention exposure	Require written terms; walk away if refused
Claims of accuracy or savings with no method or sample	Unverifiable promises	Ask for method; rely on your own pilot
Pilot only on vendor-chosen data or scenarios	Results will not reflect your work	Insist on your tasks and blind scoring
Refuses to disclose sub-processors or model providers	You cannot assess where data goes	Require a list and change notice
No export route or export only in proprietary form	Lock-in	Require documented export and test it
Pressure to sign by a date to keep a discount	Compresses legal and security review	Hold the review timeline; negotiate extension
Security answers are generic or evasive	Controls may be missing	Ask for specifics, evidence and the scope of reports

RED FLAG	WHY IT MATTERS	RESPONSE
Rights to change terms at will	Commitments can disappear	Require notice and a termination right
Reference customers cannot be reached or are not comparable	Little proof of operation	Ask for another reference or reduce scope

Also watch internal red flags. A business sponsor who has already announced the choice, a security review scheduled after commitment, a pilot with no pass mark and a budget that covers licenses but not integration or evaluation are all signs that the process is serving the decision instead of informing it. Raising these early is a service to the sponsor, and a good procurement lead says so plainly.

ILLUSTRATIVE SCENARIO (HYPOTHETICAL)

A team prefers a vendor whose demo was strongest. In the scripted pilot, outputs on difficult cases fall below the pass mark, and the vendor cannot state which sub-processor handles logs. The runner-up passes and accepts a no-training clause and a change-notice term. The team switches, with the written pilot record to justify the decision.

WORKING QUESTION

Which red flag on this list have you already seen in your current process and decided to overlook?

ONGOING MANAGEMENT

Manage the vendor after signature

08



The contract is the start of the relationship. Run it with an owner, a calendar and evidence.

Assign a vendor owner on the business side and a counterpart in procurement, with security and privacy input as needed. Hold a regular review, often quarterly, covering service levels, incidents, support quality, usage and cost against forecast, roadmap and changes, and open risks. Keep the vendor file current with contracts, questionnaire answers, assurance reports and the register of sub-processors. Repeat a lighter version of the pilot test after significant vendor changes, because behavior may shift when models or features are updated.

01 Onboard

Configure roles, logging, retention and integrations to the agreed settings, and record the configuration as the reference.

02 Monitor

Track availability, quality samples, usage and spend against the baseline from the pilot.

03 Review

Hold scheduled reviews with the vendor, using the evidence, and record actions and owners.

04 Respond to change

When the vendor gives notice of a change, assess impact, re-test key tasks and decide whether to accept, mitigate or exercise rights.

05 Prepare renewal and exit

Start six months before renewal: re-score against the market, reconsider terms and test the exit plan.



Vendor file

- Signed agreement, order forms, data terms and security addendum
- Written security and data answers with dates and signatories

- Current sub-processor list and change notices received
- Assurance reports and their scope and expiry dates
- Pilot results, scoring sheets and acceptance decisions
- Service level reports, incident log and credits claimed
- Exit plan with data export tested at least once

Keep optionality. Avoid designing workflows that depend on proprietary features with no equivalent elsewhere, hold your own prompts, templates, test sets and evaluation results in formats you control, and keep the scripted pilot ready to run on a new vendor. This is the most practical protection against lock-in, and it makes negotiations at renewal more balanced, because the vendor knows you can compare alternatives quickly.

WORKING QUESTION

If you had to move to another vendor within 90 days, which of your assets would transfer cleanly and which would not?

30 / 60 / 90

A first quarter that produces evidence

09

Define the need, test on your own work, put data and exit terms in writing, and manage the vendor like the business-critical supplier it becomes.

First 30 days: ground it

- Write the requirements brief with must-haves, weighted criteria and an exit condition
- Build the long list and screen on documents and a written questionnaire
- Involve security, privacy and legal and agree their review points
- Prepare the pilot test set, rubric and data rules

Days 31 to 60: test it

- Run scripted pilots with blind scoring on three to four vendors
- Collect written answers on data use, retention, sub-processors and incidents
- Score, record must-have failures separately and check references
- Open contract discussions on the key terms with counsel



Days 61 to 90: decide

- Complete negotiation on data, change notice, service levels and exit terms
- Select the vendor with a documented rationale and a recorded runner-up
- Configure to agreed settings and set up the vendor file and owner
- Schedule the first quarterly review and test the data export once

GLOSSARY

Terms used in this publication

10



TERM	MEANING
Sub-processor	A third party engaged by a vendor that may process customer data on the vendor's behalf.
Data processing terms	Contract terms setting how a vendor may handle personal or other customer data on the customer's behalf.
Training use	The use of customer content to train or improve a vendor's models.
Retention	How long data, including prompts, outputs and logs, is stored.
Indemnity	A contractual promise by one party to defend or compensate the other for specified claims.
Service level	A measurable commitment such as availability or response time, usually with remedies if missed.
Change notice	A duty to tell the customer in advance of material changes to the service, models or terms.
Portability	The ability to move your data and configuration to another provider in a usable form.



TERM	MEANING
Vendor lock-in	Dependence on one provider that makes leaving costly or impractical.
Scripted pilot	A test in which every vendor performs the same tasks on the same data, scored the same way.
Prompt injection	Hidden or hostile instructions in content that try to steer an AI system.
Order of precedence	The clause that decides which document governs when contract documents conflict.

A note on use

This publication is a public orientation resource, not legal, technical, financial or professional advice. Adapt it to your people, information, workflows and obligations, and check current tool terms, regulations and risks before acting.

TAKE THE NEXT STEP

Take the free Halden assessment at halden.company/assessments, or talk to us at halden.company/contact about turning this into a plan for your organization.