Anyway Ventures

PRIVATE EQUITY · VCS · BUYERS OF AI · 2026 EDITION

AI due diligence *question bank*.

87 questions for testing a target's AI, its exposure to AI and what AI could be worth to a new owner. Each comes with the evidence to request, what a good answer looks like and the answer that should worry you.

PDF GUIDE | WORD QUESTION LIST | EXCEL TRACKER

AUTHOR

Brannan Coady, Anyway Ventures

EDITION

Version 1.0 · October 2026

ONLINE, WITH DOWNLOADS

anywayventures.com/resources/ai-due-diligence-questions

CONTACT

brannan@anywayventures.com

Free to use, adapt and share with attribution · Not legal, tax or investment advice

CONTENTS

In this guide

1. Executive summary
2. How to use this question bank
3. The twelve areas at a glance
4. A. AI claims and how to verify them
5. B. Product and production usage
6. C. Models, providers and dependency
7. D. Data: sources, rights, consent and training use
8. E. Evaluation, quality and human review
9. F. Unit economics and inference cost
10. G. Security and AI-specific threats
11. H. Governance, policy and regulation
12. I. IP and AI-generated code
13. J. Team and capability
14. K. AI disruption exposure
15. L. Value-creation upside
16. How to test AI claims hands-on
17. Scoring approach
18. Buying an AI vendor: the operator's version
19. Worked example
20. Red flags and common mistakes
21. Glossary
22. Sources and further reading
23. About this resource

About Anyway Ventures

WHO IT IS FOR

PE deal teams, operating partners, VCs, corporate development, and operators buying AI products

ALSO AVAILABLE

PDF guide, Word question list, Excel tracker at anywayventures.com/resources/ai-due-diligence-questions

KEY TAKEAWAYS

The short version

1. Anyway Ventures' AI due diligence question bank has 87 questions in twelve areas, testing whether a target's AI is real, how AI changes its market and what AI could be worth to a new owner.

2. A Red on any of seven AI due diligence gate questions is a deal issue on its own, covering function, manual work, data rights, margin at scale, cross-customer leakage and EU AI Act status.

3. In AI due diligence, ask for evidence rather than answers, because twelve months of AI provider invoices and the customer contract templates usually tell you more than the deck.

4. When testing an AI target, a two-hour live session on production with inputs you chose, plus a side-by-side test against a general assistant, settles more than a week of management presentations.

5. Re-base gross margin before valuing an AI business by moving inference and human review into cost of sales, which commonly takes several points off reported margin.

This is a question bank for AI due diligence: 87 questions in twelve areas, each with why it matters, the evidence to request, what a good answer looks like, the answer that should worry you, and the deal stage it belongs to. It is written for private equity deal teams, operating partners, VCs and corporate development teams assessing a company whose value depends on AI, or whose market AI is changing. A shorter vendor version covers operators buying an AI product rather than a company.

1. Executive summary

1. **AI due diligence has three jobs, and most questionnaires only do one.** They test whether the target's AI is real. They rarely test how AI changes the target's market, or what a new owner could make from AI, costed. This bank covers all three: claims (areas A to J), exposure (K) and upside (L).
2. **Ask for evidence, not answers.** Every question names the document, log, invoice or test that settles it. Twelve months of AI provider invoices and the customer contract templates usually tell you more than the deck.
3. **Run the product yourself.** A live session on production with inputs you chose, plus a side-by-side against a general assistant with a decent prompt, settles more in two hours than a week of management presentations. The test protocol is in its own section below.
4. **Seven gate questions decide whether there is a deal issue.** Evidence of function, undisclosed manual work, training data rights, customer data rights, margin at scale, cross-customer data leakage and EU AI Act high-risk status. A Red on any of them is a finding by itself, whatever else scores well.
5. **Re-base gross margin before you value anything.** Inference and human review belong in cost of sales. Moving them there commonly takes several points off reported margin, and margin then has to be modelled at the usage the business plan assumes.
6. **Security questions for AI are different.** Prompt injection, excessive agency and cross-tenant leakage through retrieval indexes are not covered by a standard penetration test. Ask whether anyone has tested the AI features specifically.
7. **The regulatory calendar moved but didn't go away.** The EU's Digital Omnibus pushed high-risk obligations to December 2027 and August 2028. Both dates sit inside a normal hold. UK duties arrive through UK GDPR, as amended by the Data (Use and Access) Act 2025, and sector regulators.
8. **Findings should land in a specific place.** Exposure findings move price. Claims findings move structure: warranties, specific indemnities, retentions, earn-outs. Upside findings go into the model as a base case with a realistic start date, and the stretch case stays out of the price.

2. How to use this question bank

Who it is for. Deal teams and their technical advisers on a control or significant minority deal; VCs at Series A and later (at seed, the pre-LOI questions and the hands-on test are usually enough); corporate development teams buying AI capability; and operators procuring AI products, who should start with the vendor version.

When to use it. Each question is tagged with a stage:

STAGE	WHEN	WHAT IT NEEDS
Pre-LOI	First management meeting, IM review, before exclusivity	Management's time and the information memorandum. 17 questions.
Confirmatory	During exclusivity, usually two to four weeks	Data room access, documents, sessions with engineers and finance. 65 questions.
Post-deal	Conditions, 100-day plan and first board packs	Owners and dates agreed at completion. 5 questions, plus follow-ups from earlier findings.

How long it takes. On a lower-mid-market deal, a focused red, amber and green view on claims and exposure takes about a week. Hands-on testing and a proper cost-to-serve analysis add a second week. A costed value-creation case adds a third. That matches the one to three week scoping in AI due diligence for private equity.

What you need from the target. The core data room request: product roadmap and architecture diagram; a model and AI vendor inventory; twelve months of AI provider and cloud invoices; every customer contract template ever used, with the number of customers on each; pricing history and revenue split by pricing model; headcount by role; any AI policy, DPIAs and incident logs; and the sales materials that make AI claims. Then four or five sessions: the CTO, a hands-on engineer without the CTO, the product lead, the CFO on AI costs, and two or three operations leads.

How to cut it down. Not every question applies. A target with no AI product still needs areas K and L, and probably G5, H1 and J6 on its own use of AI tools. A seed-stage company needs areas A, B, C and F at the pre-LOI depth. Mark the rest N/A in the tracker rather than deleting them, so the gaps stay visible.

The companion files. The Excel tracker has every question as a row, with stage and gate filters, owner, date asked, answer summary, evidence received and a Red, Amber, Green rating. A summary tab rolls ratings up by area and stage, flags any gate question rated Red and applies the scoring rules below. It also has a 20-test log for the hands-on session, the vendor questions, and the worked example filled in. The Word question list has the questions and evidence requests only, formatted to paste into your own information request list.

3. The twelve areas at a glance

AREA	WHAT IT ANSWERS	FEEDS THE IC SUMMARY LINE
A. AI claims	Is what the materials say true?	AI claims
B. Production usage	Do paying customers use it, every week?	AI claims
C. Models and dependency	Whose AI is it, and what happens when the model changes?	AI claims
D. Data rights	Does the company have the right to use the data the AI depends on?	AI claims, Regulation
E. Evaluation and review	How does the team know it works, and who checks?	AI claims
F. Unit economics	What does it cost to run, and does margin hold at scale?	Cost to run
G. Security	Can the AI be turned against the business or its customers?	AI claims (own line if serious)

AREA	WHAT IT ANSWERS	FEEDS THE IC SUMMARY LINE
H. Governance and regulation	Who owns AI risk, and which rules apply?	Regulation
I. IP and AI-generated code	Does the company own what it built, and is it clean?	AI claims (own line if serious)
J. Team	Who can run and change it after completion?	AI claims
K. Disruption exposure	How does AI change demand, pricing and the moat?	Market exposure
L. Value-creation upside	What can a new owner make from AI, costed?	Opportunity

The IC summary lines are the five used in AI due diligence for private equity: market exposure, AI claims, cost to run, regulation and opportunity. The bank is deliberately longer than that summary. The IC reads five lines; the workings behind them run to 87 questions.

Gate questions

Seven questions are marked as gates. A Red on any of them is a deal issue on its own: it goes to the IC as a named finding with a proposed remedy, and it can't be offset by good answers elsewhere.

GATE	WHY IT IS A GATE
A2 Live on our inputs	If it can't run on your data, nothing else in the claims area means much
A4 Human touch rate	Undisclosed manual work changes the business model, the margin and the multiple
D1 Training data rights	A model built on data the company had no right to use is a liability on the balance sheet
D2 Customer data rights	Practice that has run ahead of the contracts is a known issue W&I won't cover
F4 Margin at scale	If margin inverts at the usage the plan assumes, the plan is wrong
G3 Cross-tenant leakage	One customer's data in another's output is a breach, a lost customer and a claim
H3 EU AI Act classification	High-risk status with no plan is a compliance cost with a 2027 or 2028 deadline

4. A. AI claims and how to verify them

Start with the claims register: every sentence about AI in the IM, the deck, the website and the sales materials, each mapped to what the software actually does. Most AI washing is optimism plus loose language, not fraud, but the valuation consequence is the same either way. The protocol behind these questions is set out in AI washing in due diligence.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
A1 List every AI claim in the IM, deck, website and sales material. For each one, what does the software do and what do people do? ***Pre-LOI***	The equity story is built on these sentences. Regulators treat false AI claims as misstatements, and so will your warranty claim.	A claims register: each claim mapped to a named feature, model, owner and supporting evidence.	**Good:** Every claim maps to a feature an engineer recognises. Management has already toned down anything it can't prove. **Worrying:** "The AI handles it." Claims the engineering team has never heard of, or can't locate in the product.
A2 Can we run the product live, on production, with inputs we choose? ***Pre-LOI · gate***	A demo on prepared data shows the sales team is good. Your own messy inputs show what the product does.	A live session on production, then a sandbox or trial login you can use without management present.	**Good:** Agrees within days. Runs your inputs end to end and fails visibly when it fails: flags low confidence or routes to a person. **Worrying:** Only a "demo environment". Your inputs are "not representative". A week's delay to "prepare".
A3 Which features are model-driven, which are rules or search, and which are manual? ***Confirmatory***	Rules engines and RPA scripts get described as AI. Each has a different cost, risk and defensibility profile.	Architecture diagram annotated by component, plus a screen-shared code walk of one feature end to end.	**Good:** A clean split, with rules used where rules are better and no embarrassment about it. **Worrying:** Keyword matching described as machine learning. Nobody can draw the boundary.
A4 What share of tasks does a person touch today, and what was it six and twelve months ago? ***Confirmatory · gate***	Undisclosed manual work is the commonest form of AI washing, and it sits in cost of sales whether booked there or not.	Intervention or exception logs, operations team rotas, offshore and contractor invoices, org chart with job descriptions.	**Good:** A measured rate, falling, disclosed to customers, with the review cost in the margin model. **Worrying:** "None." Or a team of "operations analysts" or "data reviewers" nobody can explain.
A5 Where does each accuracy, automation or time-saving number in the materials come from? ***Confirmatory***	An untested accuracy number is the most common problem in AI marketing and the one investors check least.	For each figure: the test set, sample size, date, method and who ran it.	**Good:** Reproducible on a named, dated test set drawn from real customer work. Caveats stated. **Worrying:** A single headline figure with no test set. "95%" becomes "up to 95%" becomes "in our best cohort".

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
A6 How much revenue comes from AI features, and how do you measure it? ***Pre-LOI***	If the multiple assumes AI revenue, you need to know how much exists and how it is counted.	Revenue by product and add-on, attach rates, price list history, and how "AI revenue" is defined.	**Good:** AI is priced or tracked separately, with attach and retention rates by cohort. **Worrying:** All revenue relabelled "AI-enabled" since the last fundraise.
A7 If the AI features were switched off tomorrow, what would customers lose? ***Confirmatory***	This is the take-the-AI-away test. It separates a feature customers rely on from one that decorates the product.	Feature-level usage telemetry, support tickets about AI features, and your own reference calls.	**Good:** Named workflows break, and customers say so unprompted. **Worrying:** Very little. The AI sits in a tab a minority of users open.

> **Why it matters.** Regulators now treat AI claims as claims. The US SEC brought its first AI washing cases in March 2024, and in January 2025 settled charges against Presto Automation after finding it claimed its drive-thru AI eliminated human order-taking when most orders through one version needed human intervention. The FTC's 2025 order against Workado concerned a real product advertised as 98% accurate that, according to the FTC's complaint, scored 53% on general content in independent testing. Question A5 exists because of cases like that one.

5. B. Product and production usage

A demo proves a capability exists once. These questions establish whether it is used, every week, by customers who pay for it, and what happens when it goes wrong. More on reading demos in beyond the demo.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
B1 How many customers use AI features weekly, and how has that moved since launch? ***Pre-LOI***	Pilots convert to logos on a slide easily. Logos convert to weekly use much less often.	Weekly active usage by customer and by feature for twelve months.	**Good:** Rising usage, concentrated inside the daily workflow. **Worrying:** A launch spike then decay. Usage concentrated in two friendly accounts.
B2 Which customers pay for AI in production, and which are pilots, trials or design partners? ***Confirmatory***	Free pilots and design partners inflate traction and rarely convert at the rate the plan assumes.	Customer list with contract status, start dates and invoices for AI line items.	**Good:** A clean split, with pilot-to-paid conversion rates by quarter. **Worrying:** Pilots counted as customers. Design partners paying nothing, shown as logos.
B3 Show us a sample of anonymised production logs: real inputs, outputs and failures. ***Confirmatory***	Logs show what the product does when nobody is presenting it.	200 or so randomly sampled interactions under NDA, including failures and escalations.	**Good:** Messy real inputs, visible failures, and evidence the team reviews them. **Worrying:** Logs not kept, kept but never read, or only a curated sample offered.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
B4 What are p50 and p95 latency and availability for AI features in production? ***Confirmatory***	A three-second demo answer can be forty seconds under load, which in some workflows kills adoption.	Monitoring dashboards and SLA reports for six to twelve months.	**Good:** Known numbers, inside the SLA offered to customers, with alerting. **Worrying:** Nobody knows. Latency only measured in the demo.
B5 What has gone wrong with the AI in production, and what did you change as a result? ***Confirmatory***	A team that has run AI in production has a list of things that broke. No list means no production or no monitoring.	Incident log, post-mortems, AI-related complaints and credits issued.	**Good:** Specific incidents, root causes and fixes, each with an owner. **Worrying:** "No incidents." Complaints handled informally by the founder.
B6 Can we speak to two customers we choose, without management on the call? ***Confirmatory***	References picked by management describe a nice tool. Ones you pick describe what changed, or didn't.	Customer list with usage data, so you can choose one heavy and one light user.	**Good:** Customers describe a changed process and what they would lose. **Worrying:** Only hand-picked references, or a refusal citing confidentiality.
B7 What is on the AI roadmap for the next twelve months, and what was cut from the last one? ***Confirmatory***	Roadmap discipline shows whether the AI strategy is planned or reactive to each model release.	Current and previous roadmaps, with what shipped against what was planned.	**Good:** A realistic roadmap with an honest delivered-versus-planned record. **Worrying:** A roadmap rewritten each quarter around whatever a lab announced last.

6. C. Models, providers and dependency

Almost every AI product is built on someone else's model, and that is usually the right choice. The questions are what the company owns on top of the model, and what happens when the provider changes its price, its terms or the model itself. The full argument is in wrapper or moat.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
C1 Which models, from which providers, do each job, and why those? ***Pre-LOI***	You can't assess dependency, cost or data flows without knowing what runs where.	A model inventory: model, version, provider, hosting region, feature, owner and monthly spend.	**Good:** A complete, current inventory, with a reason for each choice and a known alternative. **Worrying:** The CTO isn't sure which model versions are live. The inventory is assembled for the data room.
C2 If you swapped your main model for the best alternative tomorrow, what would break and how long would it take? ***Confirmatory***	Portability is the single best test of dependency. Ask the engineer, not the founder.	Results of a run on a second provider against the eval set, with prompt changes needed.	**Good:** "We ran it on two providers; quality drops 3% on our evals; about a week of prompt work." **Worrying:** Never tried. Prompts tuned to one model's quirks. "We'd be fine."

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
C3 What do your provider terms say about data use, retention, rate limits, price changes and notice? ***Confirmatory***	Provider terms decide whether customer data is retained or used for training, and how fast your cost base can move.	Signed provider agreements, enterprise or API terms, data processing addenda, any committed-spend contracts.	**Good:** A senior person has read them and can summarise the risks. Business or enterprise terms with no training on inputs. **Worrying:** Default or consumer terms. Nobody has read them. Spend on a founder's credit card.
C4 What happened the last time a provider retired a model you depended on? ***Confirmatory***	Deprecation happens to everyone. Providers publish notice periods of weeks to months, and replacement models behave differently.	Migration tickets, eval runs before and after, any customer-facing regressions.	**Good:** Migrated in days on the back of the eval set, with measured quality impact. **Worrying:** A scramble. Quality drop found by customers. A retirement date coming up with no plan.
C5 What does your product do that a user with a ChatGPT, Claude or Copilot subscription can't do in an afternoon? ***Pre-LOI***	If the answer is "better prompts", the product is a thin wrapper and the next model release competes with it.	A live comparison: same task, raw model with a decent prompt versus the product.	**Good:** Integrations, owned data, workflow and accountability they can name and demonstrate. **Worrying:** Prompt quality and a nicer interface.
C6 What happens to customers during a provider outage or rate-limit event? ***Confirmatory***	Provider outages are routine. A product with no fallback inherits every one of them.	Outage history, fallback design, rate-limit tiers and any credits paid to customers.	**Good:** A fallback provider or graceful degradation, tested, with outage history measured. **Worrying:** The product stops. Nobody has measured how often.
C7 Which fine-tuned or self-hosted models do you run, on what base model, under what licence? ***Confirmatory***	Open-weight base models carry licence terms. Fine-tunes can be impossible to reproduce if the person who made them leaves.	Model cards, base model licences, training scripts, data used and where the weights are stored.	**Good:** Licence permits your commercial use. Training is reproducible from versioned code and data. **Worrying:** Weights on a departed contractor's account. A base licence nobody has read.
C8 How has net revenue retention moved since the last major model release? ***Pre-LOI***	Every time a frontier model improves, thin wrappers lose customers to it. Retention through releases is the market's own test.	Monthly NRR and logo churn for 24 months, annotated with major model releases.	**Good:** Stable or rising through each release. **Worrying:** Churn spikes after releases. Lost customers say they now "just use ChatGPT".

> **Tip.** Deprecation happens to everyone. OpenAI's deprecations policy gives at least six months' notice for generally available models and as little as two weeks for previews; Anthropic's policy commits to at least 60 days for publicly released models. Check the inventory from C1 against both pages during diligence. A retirement date that falls in the first six months of ownership belongs in the 100-day plan.

The provider market itself moves quickly. Menlo Ventures' 2025 State of Generative AI in the Enterprise estimated Anthropic's share of enterprise LLM API usage at 12% in 2023 and 40% in 2025, with OpenAI falling from 50% to 27% over the same period. A company that can switch provider in a week captures each improvement. One hard-wired to a single provider waits.

7. D. Data: sources, rights, consent and training use

Data a competitor can't get is the strongest moat in AI, and the most overclaimed. "Our customers upload documents" usually means the customer's data, under the customer's contract, which may not permit training anything. Read every contract template the company has ever used, not only the current one, and count the customers on each.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
D1 What data are your models trained, fine-tuned, grounded or evaluated on, and what right do you have to use each source? ***Confirmatory · gate***	A model trained on data the company had no right to use is a liability, whatever the accuracy numbers say.	A data rights register: each dataset, source, licence or contract basis, personal data content and permitted uses.	**Good:** A register per dataset with the legal basis stated and documents behind it. **Worrying:** "Public internet." Scraped sites. "Customer data" with no clause behind it.
D2 Do your customer contracts permit using customer data to improve models, and across customers? ***Confirmatory · gate***	Cross-customer data rights are rare and valuable. Practice that has quietly run ahead of the contracts is a known issue W&I won't cover.	Every MSA template ever used, variations and side letters, DPAs, and the customer count on each template.	**Good:** An explicit clause, opt-outs honoured and tracked, practice matching contract. **Worrying:** Contracts silent. A pooled model trained on data from customers whose template says nothing.
D3 What personal data flows through AI features, and what is the UK GDPR lawful basis for each purpose, including training? ***Confirmatory***	Training is a separate purpose from delivering the service and needs its own lawful basis, assessment and transparency.	Record of processing, DPIAs for AI features, legitimate interests assessments, privacy notices.	**Good:** A DPIA per AI feature, lawful basis per purpose, privacy notices that mention AI use. **Worrying:** No DPIA. Training on personal data assumed to be covered by "performance of contract".
D4 Where does data go when it reaches a model provider, how long is it kept, and in which country? ***Confirmatory***	Data sent to a provider may be logged, retained or reviewed, and may leave the UK.	Data flow diagram, provider retention settings, sub-processor list, international transfer mechanisms.	**Good:** Limited or zero retention, known regions, transfer mechanisms in place, sub-processors disclosed to customers. **Worrying:** Unknown. Free-tier keys in production. Customers never told a model provider is a sub-processor.
D5 Has any customer, data subject or rights holder objected, opted out or complained about your use of data for AI? ***Confirmatory***	Complaints and letters before action are the early signs of a claim, and they rarely make it into the IM.	Complaints log, subject access and erasure requests that mention AI, legal correspondence.	**Good:** Few, logged and handled properly. **Worrying:** Unlogged objections, or a rights holder letter sitting with the founder.
D6 What data does the product generate itself that a competitor can't get, such as corrections, outcomes and edge cases? ***Pre-LOI***	Data generated by use, with the rights to use it, is the most durable moat an application company can build.	Schema and volumes of captured feedback, labelled corrections and outcome data, and how it is used.	**Good:** Captured, labelled, rights confirmed, and already improving the product measurably. **Worrying:** Not stored, or stored but never used.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
D7 Can you delete a person's or customer's data from training sets, retrieval indexes, caches and logs on request? ***Confirmatory***	Erasure duties and customer exit clauses reach embeddings and fine-tunes too, which most teams forget.	Deletion procedure, a recent test, and where the vector store and logs fit in.	**Good:** A documented, tested process that covers indexes and logs, with a stated position on trained weights. **Worrying:** Nobody has thought about the vector store. "Once it's trained it's trained."

> **Red flag.** The pattern to look for is a pooled model, trained across customers, sitting on contracts that say nothing about model improvement. It is common, it is fixable (consent, retraining or a narrower model), and once found it is a known issue, which warranty and indemnity insurance normally won't cover. Price it, indemnify it or hold back a retention. Leaving it in the report is not a remedy.

Training on personal data is a separate purpose from delivering the service, and under UK GDPR it needs its own lawful basis, transparency and usually a DPIA. The ICO's guidance on AI and data protection is the main reference; the ICO marks it as under review following the Data (Use and Access) Act 2025, so check the current version when you use it.

8. E. Evaluation, quality and human review

The eval set is the least glamorous asset in an AI company and one of the most valuable. It lets the team change models in days, prove quality to customers and catch regressions before they ship. Human review is often the right design; it just has to be measured and costed.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
E1 Show us the eval set: how many cases, where they came from, and how they are scored. ***Confirmatory***	The eval set is how a team knows the product works, changes models safely and proves quality to customers.	The eval set itself under NDA, scoring method, and score history by release.	**Good:** Hundreds to thousands of real cases with known good answers, versioned, reviewed by domain experts. **Worrying:** None, or twenty hand-picked examples in a spreadsheet.
E2 Is the eval run before every model or prompt change, and what was the last regression it caught? ***Confirmatory***	An eval that runs occasionally doesn't stop regressions reaching customers.	CI configuration, eval run history, and the last blocked release.	**Good:** Automated on every change, with a recent, specific example of a caught regression. **Worrying:** Run by hand before big releases. No example to give.
E3 How is quality monitored in production, and who is alerted when it drops? ***Confirmatory***	Model behaviour drifts when inputs, providers or prompts change. Someone has to notice before customers do.	Quality dashboards, alert thresholds, on-call rota and recent alerts.	**Good:** Live quality metrics with thresholds and a named owner who gets paged. **Worrying:** "Customers would tell us."

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
E4 How are confident wrong answers detected and handled? ***Confirmatory***	Fluent errors are the characteristic failure of generative AI, and the liability lands on the business, not the model.	Grounding and citation design, confidence thresholds, review queue volumes.	**Good:** Outputs grounded in sources and checked, low-confidence cases routed to people, error rates measured. **Worrying:** No mechanism. Disclaimers in the terms instead.
E5 Who reviews outputs, how many people, against what criteria, and at what cost? ***Confirmatory***	Human review is often the right design, but it is a variable cost that has to be in the margin.	Review process, staffing, QA rubric, reviewer throughput and cost per reviewed item.	**Good:** Trained reviewers, a written rubric, sampling rates and cost per item in cost of sales. **Worrying:** Unmeasured review by whoever has time. Not in the cost model.
E6 Where AI outputs affect people (hiring, credit, pricing, access), how have you tested for bias and unfair outcomes? ***Confirmatory***	Equality, data protection and the EU AI Act all bite hardest where AI makes or shapes decisions about people.	Bias testing results, protected characteristic analysis, any external audit.	**Good:** Tested where it matters, results documented, mitigations in place, owner named. **Worrying:** Decisions about people, never tested. "The model is neutral."
E7 Which quality metric will the board see each month, and what threshold triggers action? ***Post-deal***	If quality isn't on the board pack, it isn't managed after completion.	Proposed board KPI set for AI, with thresholds and owners.	**Good:** One or two agreed quality metrics, thresholds, and an owner. **Worrying:** No metric proposed. Quality reported as anecdotes.

> **Why it matters.** Liability for a wrong answer lands on the business, not the model. In Moffatt v Air Canada (2024), a Canadian tribunal held the airline liable for its chatbot's incorrect statement about bereavement fares and rejected the argument that the chatbot was responsible for its own words. The sums were small. The principle is the one a buyer inherits.

9. F. Unit economics and inference cost

AI features have a running cost that scales with usage, which most SaaS finance teams are not used to modelling. The two most common findings are inference booked in R&D or "platform" costs, which flatters gross margin, and human review sitting in customer success or operations rather than cost of sales.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
F1 What does one unit of work cost you, all in: model fees, retrieval, hosting and human review? ***Pre-LOI***	Founders who haven't worked this out haven't priced the product.	A per-task cost breakdown with twelve months of trend.	**Good:** A per-task figure, all components included, stable or falling. **Worrying:** Unknown, or excludes model fees "because they're coming down".

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
F2 Provide twelve months of AI provider and cloud invoices, mapped to features and customers. ***Confirmatory***	Invoices tell you more than the deck does. Mapping them shows who is profitable to serve.	Invoices, usage exports from provider consoles, and a mapping of spend to feature and customer.	**Good:** Spend tagged by feature and customer, reconciling to the ledger. **Worrying:** One unallocated line. Spend spread across personal cards and expense claims.
F3 Where are inference and human review booked: cost of sales or operating expenses? ***Confirmatory***	Inference booked under R&D or "platform" flatters gross margin and the multiple that goes with it.	Chart of accounts mapping, management accounts and the finance team's policy note.	**Good:** Both in cost of sales, consistently, with a note on the policy. **Worrying:** In R&D or opex. Finance unaware the AI has a variable cost.
F4 What is gross margin per customer at today's usage, at three times today's usage, and for your heaviest 10%? ***Confirmatory · gate***	AI costs scale with usage, and a flat price can turn a heavy user from profitable to loss-making.	Customer-level margin model using real usage, at current and planned volumes.	**Good:** Modelled, and margin holds at the volumes the business plan assumes. **Worrying:** Never modelled, or margin inverts for the top users the plan relies on.
F5 How is AI priced to customers, and does price move with usage? ***Confirmatory***	Pricing decides who keeps the saving as models get cheaper and who pays when usage grows.	Price list, contract terms on usage, fair-use limits and overage history.	**Good:** A usage or outcome component, or fair-use limits that are enforced. **Worrying:** Unlimited AI for a flat fee, with no limit in the contract.
F6 What cost controls are in place: caching, model routing, token limits, per-customer budgets and alerts? ***Confirmatory***	Unbounded consumption is on the OWASP LLM Top 10 for a reason: one runaway loop or abusive user can cost thousands overnight.	Cost control design, budget alerts, and the largest unexpected bill in the last year.	**Good:** Caching and routing in place, per-tenant limits, alerts, and savings measured. **Worrying:** None. A surprise bill last quarter nobody investigated.
F7 What would moving routine tasks to a smaller or cheaper model save, and what quality would it cost? ***Post-deal***	Model prices fall fast. A team with evals can bank the saving; one without can't risk it.	Results of a routing or downgrade test against the eval set.	**Good:** Tested, with the saving and quality impact quantified. **Worrying:** Never tried, so every task runs on the most expensive model.

> **Tip.** Model prices fall fast. Epoch AI's 2025 analysis found the price of reaching a fixed level of model performance fell by between 9x and 900x a year depending on the benchmark. That helps a target with evals, which can move routine work to cheaper models safely (F7). It helps a target with flat per-seat pricing only if customers don't demand the saving.

10. G. Security and AI-specific threats

A standard web application penetration test does not cover the ways an AI feature fails. The OWASP Top 10 for LLM Applications 2025 is the best shared vocabulary: prompt injection is first, sensitive information disclosure second, and excessive agency, system prompt leakage, vector and embedding weaknesses and unbounded consumption all make the list. The questions below map to it.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
G1 Has the AI product been tested for prompt injection, direct and indirect, and by whom? ***Confirmatory***	Prompt injection is first on the OWASP Top 10 for LLM applications. Indirect injection arrives in documents, emails and web pages the AI reads.	Penetration test scoped to the AI features, findings and remediation, mapping to the OWASP LLM Top 10.	**Good:** An independent test of the AI features, findings fixed, tests repeated after model changes. **Worrying:** Only a standard web application pen test. "Our system prompt tells it not to."
G2 What can the AI do: which tools, systems and data can it read or change, and with whose permissions? ***Confirmatory***	Excessive agency turns a prompt injection from an embarrassing answer into a payment, a deletion or a data export.	List of tools and integrations the AI can call, permission model, approval steps for consequential actions.	**Good:** Least privilege, actions run with the user's own permissions, human approval for anything consequential. **Worrying:** A shared service account with admin rights. Agents that can send email or move money unattended.
G3 Can one customer's data appear in another customer's output, through retrieval, caching, fine-tuning or logs? ***Confirmatory · gate***	Cross-tenant leakage is a breach notification, a lost customer and a claim at once.	Tenant isolation design for indexes and caches, test results, any shared fine-tunes.	**Good:** Tenant-scoped indexes and caches, isolation tested, no shared training without contractual rights. **Worrying:** One shared index filtered by a metadata field. Never tested.
G4 How are system prompts, API keys and model endpoints protected? ***Confirmatory***	System prompts leak. Anything secret placed in one should be assumed public.	Secrets management, key rotation records, and what the system prompts contain.	**Good:** Secrets in a vault, rotated, nothing sensitive in prompts, keys scoped per environment. **Worrying:** Keys in code or prompts. One key shared across production and staff laptops.
G5 Which public AI tools may staff use, under what terms, and how is it controlled? ***Confirmatory***	Staff pasting code or customer data into consumer AI tools is the most common AI data leak in practice.	AI acceptable use policy, list of approved tools and licence tiers, SSO and data loss prevention settings.	**Good:** Approved tools on business terms, single sign-on, a short policy people have read. **Worrying:** No policy. Source code and customer files pasted into free accounts.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
G6 What AI supply chain risks exist: open-weight models, plugins, agent tool servers, AI libraries? ***Confirmatory***	Models, plugins and tool servers are code from third parties, with the same supply chain risk as any package.	Inventory of third-party models, plugins and tool connectors with source, version and scanning.	**Good:** Known sources, pinned versions, scanned, with an owner for updates. **Worrying:** Models downloaded from anywhere. Community tool servers given production credentials.
G7 How are model outputs checked before they reach code, browsers, databases or customers? ***Confirmatory***	Model output is untrusted input. Passing it straight into a query, a page or a shell is an old vulnerability in a new place.	Output validation and encoding design, and code paths where output is executed or rendered.	**Good:** Validated against schemas, encoded, never executed directly. **Worrying:** Generated SQL or code run as-is. Raw output rendered as HTML.
G8 Does the incident response plan cover AI incidents: leakage, harmful output, provider breach, runaway cost? ***Post-deal***	AI incidents need different first steps: disabling a feature, rotating keys, contacting a provider.	Incident response plan, AI-specific playbooks, and any tabletop exercise.	**Good:** AI scenarios written down and rehearsed once. **Worrying:** Generic plan with nothing on AI.

> **Red flag.** Indirect prompt injection is the threat most targets haven't thought about: instructions hidden in a document, email or web page the AI reads on a user's behalf. In 2025 researchers disclosed EchoLeak, a zero-click prompt injection in Microsoft 365 Copilot (CVE-2025-32711, rated critical by Microsoft) in which a crafted email could cause data to be exfiltrated without the user doing anything. Microsoft fixed it server-side. If a product that large could be caught out, ask what a ten-person engineering team has done about it. Any product that ingests invoices, emails, CVs or web pages needs G1 and G2 answered properly.

11. H. Governance, policy and regulation

The UK has no AI Act. Duties reach UK companies through UK GDPR, consumer and equality law, sector regulators and, for anyone selling into or affecting people in the EU, the EU AI Act. Governance questions also tell you whether anyone in the business is in charge of AI risk at all.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
H1 Who owns AI risk at executive level, and when did the board last discuss it? ***Pre-LOI***	Without an owner, nothing below gets done and nobody notices when it slips.	Board minutes, risk register extract, responsibilities chart.	**Good:** A named executive, minuted discussion in the last two quarters. **Worrying:** Nobody, or "the CTO" by default with no time set aside.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
H2 Is there an inventory and risk register covering AI the business builds and AI it buys? ***Confirmatory***	You can't govern, classify or insure what you haven't listed.	AI inventory and risk register with owners, purposes and risk ratings.	**Good:** Current, covers built and bought tools, reviewed on a schedule. **Worrying:** None, or built for the data room last week.
H3 Where does each AI system sit under the EU AI Act, and is the company the provider or the deployer? ***Confirmatory · gate***	UK companies are in scope if they sell into or affect people in the EU. High-risk duties now land on 2 December 2027 and 2 August 2028: inside a normal hold.	Classification memo per system: prohibited, high-risk (Annex III or Annex I), transparency (Article 50) or minimal, with reasoning.	**Good:** Documented reasoning per system. If high-risk, a costed compliance plan with an owner. **Worrying:** "We're UK, so it doesn't apply" while selling into the EU. High-risk use with no plan.
H4 How do you meet the Article 50 transparency duties: telling people they are dealing with AI, and marking generated content? ***Confirmatory***	Article 50 has applied since 2 August 2026. Marking duties for systems already on the market apply from 2 December 2026.	Disclosure wording in the product, and the technical approach to marking generated content.	**Good:** Disclosure in place where users interact with AI; marking plan on track if the product generates content. **Worrying:** Unaware the duties exist.
H5 Which UK rules and regulators apply to your AI (ICO, FCA, CMA, Equality Act, sector rules), and how have you addressed them? ***Confirmatory***	The UK has no AI Act, so the duties come through data protection, consumer, equality and sector regulation.	Regulatory map, ICO correspondence, sector regulator guidance followed, any regulatory engagement.	**Good:** A short, accurate map with the gaps known and owned. **Worrying:** "There's no UK AI law, so nothing applies."
H6 Is there an AI policy, are there AI literacy measures for staff, and can you show people follow them? ***Confirmatory***	A policy nobody follows is evidence against you. Article 4 of the EU AI Act asks for literacy measures where it applies.	AI policy, training records or literacy log, sample of approvals or exceptions.	**Good:** A dated policy, a training record, and evidence of it being used. **Worrying:** A template downloaded and never read.
H7 Are you certified to ISO/IEC 42001, working towards it, or asked for it by customers? ***Confirmatory***	ISO/IEC 42001 is voluntary, but enterprise and public sector buyers increasingly ask for it, and it shows governance is real.	Certificate and scope statement, or gap assessment, and customer requests for it.	**Good:** Certification or a plan where customers demand it, with a meaningful scope. **Worrying:** Certification claimed in sales material, with a scope that excludes the product.
H8 Does the product make solely automated decisions with legal or similarly significant effects on people, and what safeguards apply? ***Confirmatory***	The Data (Use and Access) Act 2025 rewrote the UK rules on automated decision-making: more is permitted, but safeguards are mandatory.	Description of automated decisions, the safeguards offered, and how people can contest a decision.	**Good:** Decisions identified, information given, human review on request, a working route to challenge. **Worrying:** Nobody has asked whether any decision is solely automated.

The EU AI Act timetable after the Digital Omnibus

The Digital Omnibus on AI, Regulation (EU) 2026/1744, entered into force on 27 July 2026. The dates a deal team needs, as set out by the Commission and in summaries from Gibson Dunn and the Cloud Security Alliance:

OBLIGATION	APPLIES FROM
Prohibited practices (Article 5)	2 February 2025
AI literacy (Article 4), now softened to "measures to support" literacy	2 February 2025, amended July 2026
General-purpose AI model obligations	2 August 2025
Transparency (Article 50): disclosing AI interaction, deepfake labelling	2 August 2026, not deferred
Machine-readable marking of generated content, for systems already on the market before 2 August 2026	2 December 2026
High-risk systems listed in Annex III (employment, credit, education, essential services and others)	2 December 2027
High-risk AI in products regulated under Annex I (machinery, medical devices and others)	2 August 2028

For a hold starting now, both high-risk dates fall inside the first two years. A target whose product sits in an Annex III category needs a compliance cost and an owner in the plan now, not at exit. The literacy change is covered in more detail in EU AI Act AI literacy for UK businesses.

ISO/IEC 42001, the international standard for AI management systems, is voluntary. It is worth having when enterprise or public sector customers ask for it, and worth checking when claimed: in November 2025 BSI said it had become the first certification body accredited by UKAS to certify against it, so ask who issued the certificate and what its scope covers.

12. I. IP and AI-generated code

Two new IP questions sit on top of ordinary IP diligence: whether AI coding tools have introduced licence or security problems into the codebase, and whether the company owns, and can protect, what its AI produces.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
I1 How much of the codebase was written with AI coding tools, which tools, and on what terms? ***Confirmatory***	Licence tier decides whether code is used for training and whether the vendor offers any IP protection.	List of AI coding tools and licence tiers, coding policy, admin settings.	**Good:** Business or enterprise tiers with no training on code, a policy, and review standards for generated code. **Worrying:** Personal accounts on free tiers. No idea how much code is generated.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
I2 Has AI-generated code been scanned for open-source licence contamination and security flaws? ***Confirmatory***	AI tools can reproduce copyleft code without the licence notice. Licence conflicts are now found in most audited codebases.	Software composition analysis including snippet matching, static analysis results, remediation records.	**Good:** Snippet-level scanning in the pipeline, conflicts resolved, results available. **Worrying:** Never scanned. Only package-level scanning, which misses copied snippets.
I3 Who owns the model weights, prompts, eval sets and training data, and are IP assignments in place for everyone who built them? ***Confirmatory***	Contractors and agencies often build the AI, and without an assignment they may own it.	Employment and contractor agreements, IP assignments, agency contracts.	**Good:** Signed assignments for every contributor, including agencies and freelancers. **Worrying:** The fine-tune was built by a freelancer with no written agreement.
I4 Could AI outputs you deliver infringe third-party rights, and who carries that risk in your contracts? ***Confirmatory***	Targets sometimes give customers uncapped IP indemnities for output they can't control, with no matching cover from their provider.	Customer terms on outputs, provider indemnity terms and their conditions, insurance cover.	**Good:** Risk allocated deliberately, provider protections understood and conditions met, caps in place. **Worrying:** Uncapped indemnity on generated output, no provider cover, insurer not told.
I5 Does the value of the business rely on copyright in AI-generated content or code, and could you show human authorship? ***Confirmatory***	Protection for computer-generated works is unsettled and under review in the UK, and doubtful elsewhere.	Description of how core assets were created and the human contribution, with records.	**Good:** Core assets have clear human authorship, or value doesn't depend on copyright in outputs. **Worrying:** The main asset is unedited AI output, and the plan assumes it can be protected.
I6 Are there patents or applications covering the AI, and who are the named inventors? ***Confirmatory***	UK law requires a human inventor, and AI patents are often weaker than the deck implies.	Patent schedule, filings, inventor declarations, freedom-to-operate advice if any.	**Good:** Human inventors named, claims that cover something competitors would need. **Worrying:** Patents "pending" that turn out to be unfiled, or claims nobody would infringe.
I7 Are there any actual or threatened disputes, claims or regulator contacts about AI, data or IP? ***Confirmatory***	These should be disclosed against warranties, but it's better to hear about them before the disclosure letter.	Litigation schedule, correspondence, regulator contact log.	**Good:** None, or disclosed early with an assessment. **Worrying:** Surfaced in the disclosure letter the week before signing.

68%

of 947 audited codebases contained open source licence conflicts, up from 56% a year earlier. Black Duck OSSRA, 2026

54%

of organisations were evaluating AI-generated code for IP and licensing risk. Black Duck OSSRA, 2026

Black Duck's 2026 report links part of the rise in licence conflicts to AI tools reproducing copyleft code without its licence notice. Package-level scanning misses this; snippet matching is what finds it (I2).

UK law on AI and copyright is unsettled, which is a reason to ask, not a reason to panic. In Getty Images v Stability AI (November 2025), the High Court rejected Getty's secondary infringement claim, finding the model did not store reproductions of Getty's works; training-stage infringement wasn't decided because Getty dropped that claim, and an appeal has been allowed. The government's March 2026 report on copyright and AI dropped its earlier preference for a text and data mining exception with an opt-out and made no change to the law. On patents, the UK Supreme Court held in Thaler v Comptroller-General (2023) that an inventor must be a person.

13. J. Team and capability

AI systems tend to sit with one or two people. If they leave at completion, the prompts, the evals and the knowledge of what breaks leave with them. These questions overlap with key person risk, with an AI emphasis.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
J1 Who built the AI, are they still here, and who else could maintain it? ***Pre-LOI***	AI systems often sit with one person. If they leave, the prompts, evals and fine-tunes leave with them.	Org chart, commit history on the AI components, documentation.	**Good:** At least two people can run and change it, and the documentation is current. **Worrying:** One contractor, or the founder, holds it all in their head.
J2 Can we interview the engineer who maintains the prompts and evals, without the CTO present? ***Confirmatory***	The CTO gives you the strategy. The engineer tells you what breaks on a Tuesday.	A 45-minute session with the hands-on engineer, with screen share.	**Good:** Yes, and the story matches management's. **Worrying:** Refused, or the engineer contradicts the CTO on basic facts.
J3 What AI-specific skills does the team have in evaluation, retrieval, ML operations and AI security? ***Confirmatory***	Claims of proprietary models need people who can build and maintain them.	CVs or role descriptions for AI staff, hiring plan, external partners used.	**Good:** Skills that match the claims. Gaps known and in the hiring plan. **Worrying:** Claims of proprietary models with no machine learning engineers on staff.
J4 How will key AI people be retained and incentivised through the deal? ***Confirmatory***	AI engineers are in demand. Completion is when vesting accelerates and recruiters call.	Contracts, notice periods, equity and vesting schedules, proposed retention terms.	**Good:** Retention arrangements agreed for the people who matter, with sensible notice periods. **Worrying:** Key people fully vesting at completion, on one month's notice.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
J5 What AI projects have you started and stopped, and why did they stop? ***Pre-LOI***	An honest kill list shows judgement. No stopped projects usually means nothing was measured.	Project list with outcomes and reasons for stopping.	**Good:** A few stopped projects with clear reasons and lessons applied. **Worrying:** None stopped, or everything is still a pilot.
J6 How does the wider business use AI tools day to day, and does it show in productivity? ***Confirmatory***	Licences bought and nothing changed is common. Real use is also a leading indicator of the value case.	Tool licences and usage reports, examples of changed processes.	**Good:** Measured use in named processes with before and after numbers. **Worrying:** Licences bought, usage unknown, nothing changed.

14. K. AI disruption exposure

This area applies to every target, including the ones with no AI story at all. A target can have no AI and still carry more AI risk than one that sells it. The usual exposures in the lower mid-market are seat-based software priced on headcount that AI shrinks, and services businesses that bill hours AI compresses. The AI disruption risk scorecard goes further on this area alone.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
K1 What share of revenue is priced per seat, per hour or per head? ***Pre-LOI***	If AI shrinks the headcount your pricing is tied to, revenue falls even if you win every deal.	Revenue split by pricing model and by customer segment.	**Good:** Most revenue priced on outcomes, volume or value, or a plan to reprice already under way. **Worrying:** Most revenue tied to the headcount of roles AI is shrinking, with no plan.
K2 If your customers needed 25% fewer people in the roles you sell to, what would happen to your revenue? ***Pre-LOI***	This turns a vague worry into arithmetic the model can carry.	Management's own estimate, then your recalculation from the seat data.	**Good:** A worked answer with mitigations, such as repricing or expanding into adjacent work. **Worrying:** "That won't happen in our market."
K3 Could a well-funded team rebuild your core product with current AI tools in a year, and what would stop them? ***Pre-LOI***	AI coding tools have cut the cost of rebuilding software. What protects you now is data, workflow, distribution and trust.	Management's view, your own assessment of what a rebuild would need, competitor launches.	**Good:** Specific barriers: proprietary data, regulatory approvals, integrations, distribution, switching costs. **Worrying:** "Our code is complex." Code complexity is no longer much of a barrier.
K4 Which competitors or new entrants have launched AI-native alternatives or repriced, and what has happened to your win rates? ***Pre-LOI***	Win rates and pricing pressure show disruption before revenue does.	Win and loss data by competitor for eight quarters, discounting trends.	**Good:** Win rates steady or rising, discounting stable, a clear response to new entrants. **Worrying:** Win rates falling and nobody can say to whom.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
K5 Which customer tasks could a general-purpose assistant or agent now do without your product? ***Confirmatory***	General assistants keep absorbing tasks that used to need specialist software.	Task-level map of the product with your own hands-on test of a general assistant on the top tasks.	**Good:** Core tasks need data, integration or accountability a general assistant lacks. **Worrying:** The top three tasks can be done in a chat window today.
K6 For services: when AI cuts the hours a job takes, who keeps the saving, you or the client? ***Confirmatory***	Businesses that bill by the hour either give the saving away in fees or keep it as margin. Contracts and competition decide which.	Fee basis by service line, contract terms, recent client fee negotiations.	**Good:** Fixed or value-based fees on the work AI speeds up, with evidence clients accept them. **Worrying:** Hourly billing on the most automatable work, and clients already asking for discounts.
K7 Does any revenue depend on traffic, content or reference use that AI answer engines are replacing? ***Confirmatory***	Businesses that rely on search traffic, directories or reference content are exposed to AI answers that remove the click.	Traffic sources over 24 months, revenue by acquisition channel, content licensing terms.	**Good:** Low dependency, or a plan to license or protect content. **Worrying:** Organic search traffic down sharply with no explanation in the plan.

> **Tip.** Do the arithmetic in the meeting. £12m of ARR, 60% billed per seat to customer service teams, is £7.2m tied to the headcount of the target's customers. If those customers need 25% fewer people over three years and seats follow people, £1.8m is under pressure: 15% of the revenue base before any competitor does anything. The numbers are illustrative; the calculation isn't optional.

15. L. Value-creation upside

The opportunity case is the one that usually gets waved through with a slide. These questions turn it into lines a CFO will sign: named processes, hours from the target's own data, a measured baseline, capture rates with reasons, run costs and a start date that reflects the preconditions. The shape of the IC deliverable is in AI due diligence as a value-creation plan.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
L1 Where do the back-office and delivery teams spend their hours, process by process? ***Confirmatory***	The value case starts from hours and pounds in named processes, not from a benchmark.	Headcount by role, timesheets or time allocation, ticket exports and process volumes.	**Good:** A process-level view of hours from the target's own data. **Worrying:** "Everyone is busy." No data below headcount by department.

QUESTION	WHY IT MATTERS	EVIDENCE TO REQUEST	GOOD ANSWER AND WORRYING ANSWER
L2 Which candidate processes have the data ready for AI today, and which need fixing first? ***Confirmatory***	Data preconditions move a line from base case to conservative case, and set its start date.	Sample extracts from the systems each process depends on, and data quality checks.	**Good:** Preconditions tested in the data room, with a fix plan and a realistic start month. **Worrying:** Assumed data is fine because the system "has everything".
L3 What baseline exists for each candidate process: volume, time per item, error rate and cost? ***Confirmatory***	Without a baseline measured before anything is built, no one can prove the saving later.	Baseline measurements, or a plan to take them in the first 30 days.	**Good:** A measured baseline the CFO agrees with. **Worrying:** No baseline, and a plan to estimate one after the build.
L4 What capture rate can you defend for each line, as conservative, base and stretch cases, and why? ***Confirmatory***	Capture rates decide the number. Each needs a reason a sceptical IC member would accept.	A line-by-line model: hours in scope, capture rates, loaded cost, build and run costs, start month.	**Good:** Three cases, reasons stated, run costs included, a realistic start month. **Worrying:** A single percentage saving from a vendor or analyst report.
L5 If we gave you £100k for AI next year, where would it go first, and how would we know it worked? ***Pre-LOI***	The answer shows whether management has thought about AI as operations or as a slogan.	Management's answer in the meeting, then a one-page note.	**Good:** A named process, a baseline, a metric and a sensible first step. **Worrying:** "An AI strategy." A hire with no defined job.
L6 Which revenue opportunities does AI open: new products, new pricing, faster delivery or new segments? ***Confirmatory***	Revenue upside is often bigger than cost savings, and harder to evidence.	Customer research, pricing tests, pipeline for AI products.	**Good:** Customer-validated and costed, with early evidence of willingness to pay. **Worrying:** Roadmap slides with no customer evidence.
L7 Who in management owns each value-creation line, and has each owner agreed the number? ***Post-deal***	Lines with an owner and a metric get delivered. Lines owned by "the business" don't.	Named owners per line, signed-off base case, link to objectives or incentive plan.	**Good:** Each line owned by a named manager who has seen and agreed the base case. **Worrying:** Owned by the operating partner, or by nobody.
L8 What happens in the first 100 days, and what will be measured at day 30, 60 and 100? ***Post-deal***	The AI case should lift straight into the 100-day plan. If it can't, it wasn't sequenced.	Sequenced plan with dates, owners, dependencies and the metric for each milestone.	**Good:** A plan that starts with the line whose data is ready, measured in hours returned per month. **Worrying:** A list of initiatives with no dates or owners.

Price the base case, sometimes only the conservative one. The stretch case belongs in the equity story and the management incentive plan. If the fund pays for the stretch case at entry, the vendor has taken the upside and the new owner has bought the work.

16. How to test AI claims hands-on

Two hours with the product, on production, with inputs you chose, is the best value in AI diligence. It is also the step most often skipped, because it needs someone who knows what to try and how to read the result. This protocol fits in one session plus a day of preparation. For venture investors at seed and Series A, the startup demo scorecard turns the same live tests into weighted scores by stage.

Before the session

- ☐ Tell management in advance that the session will be live, on production or a production-equivalent sandbox, with inputs you supply. A real product doesn't need a week to prepare.
- ☐ Build a test pack of 20 to 30 inputs from the target's domain. Get them from a reference customer, a friendly operator in the sector, or public material. Don't use anything from the target's own demo set.
- ☐ Agree any adversarial tests in writing first. Prompt injection probes should be benign (change a field, reveal an instruction) and run only in a test tenant.
- ☐ Have a general-purpose assistant open with a decent prompt written for the same task. This is the baseline.
- ☐ Ask for the hands-on engineer to attend, not only the founder.

The test pack

The Excel tracker has a 20-row test log with these categories already laid out.

CATEGORY	WHAT TO TRY	WHAT YOU ARE LEARNING
Typical	Three normal inputs from different sources	Baseline task success on real work
Messy	A poor scan, missing fields, mixed language or jargon	How far quality falls on the inputs customers actually have
Edge	Twice the usual size, just outside scope, genuinely ambiguous	Whether the product knows its limits
Repeat	The same input twice	Consistency. Material differences between runs need an explanation
Failure	Something it should refuse or escalate, something it can't know	Whether it fails visibly or produces fluent nonsense
Grounding	Check three cited sources or extracted values against the original	Whether citations and values are real
Injection	A document with hidden instructions; a user asking it to ignore its rules	Whether untrusted content can change behaviour (G1, G2)
Leakage	Ask about another customer or data outside the test tenant	Tenant isolation in practice (G3)
Baseline	Run five typical inputs through the general assistant	What the product adds over the raw model (C5)
Load	Five tasks at once	Whether latency holds (B4)

CATEGORY	WHAT TO TRY	WHAT YOU ARE LEARNING
Human	Time a flagged item from flag to resolution	Who does the work behind the automation (A4, E5)

What to record

For each test: pass, partial or fail against an expected result written before the run; seconds to complete; whether a failure was flagged; and anything said in the room. Then four numbers for the report: task success rate on typical and messy inputs separately, consistency across repeats, the margin over the baseline assistant, and how many failures were visible rather than silent. Silent failures matter most. A product that gets 80% right and flags the rest is more valuable than one that gets 90% right and hides the other 10%.

Estimating cost from the session

Ask the engineer to show the token counts or provider console entries for your test runs. Multiply by the provider's published prices and by the volume a typical customer runs a month. If the number is far from the cost per task management gave you in F1, find out why before you look at the invoices.

Watch the people as well as the product

The most useful signal is how the team reacts when the product gets something wrong in front of you. "That's a known weakness, here's how often it happens and what we're doing" is a good answer. Talking over the failure, or switching back to a prepared example, is not.

17. Scoring approach

Rating each answer

RATING	DEFINITION
Green	The answer matches "what good looks like" and the evidence has been received and checked
Amber	Real but incomplete: a gap that is known, fixable within the first year, and can be priced, conditioned or put in the 100-day plan
Red	The answer matches "worrying answer", the evidence contradicts management, or the evidence is refused
N/A	The question doesn't apply to this target, with the reason recorded

Rate the evidence, not the answer. An excellent verbal answer with no evidence is Amber at best. Evidence promised but not delivered by the end of confirmatory diligence becomes Red.

Rolling up by area

The tracker applies these rules automatically:

- **Red** if any gate question in the area is Red, or two or more questions are Red.

- **Amber** if one question is Red, or Ambers make up 40% or more of the rated questions (and there are at least two).
- **Green** once at least half the area's questions are rated and neither rule applies.
- **Overall**: Red if any gate question is Red; Amber if any area is Red; otherwise Green.

Don't average across areas, and don't net strengths against weaknesses. A target can be excellent on evals and team and still have a data rights problem that needs an indemnity. Report each area separately, with the reason in a sentence and what it means for the deal.

Scoring the AI claim itself

For the narrower question of how much the AI is worth as an asset, use the six dimensions from AI washing in due diligence, each scored 0 to 3. They draw on specific questions in the bank:

DIMENSION	DRAWS ON	0 LOOKS LIKE	3 LOOKS LIKE
Evidence of function	A2, A4, A5, B3	Demo only works on prepared data	Performs on your data, live, end to end
Proprietary element	C5, D6, K3	Prompt and interface over a public model	Data, workflow or models a rival can't copy within a year
Unit economics	F1 to F4	Inference cost unknown or hidden in opex	Cost per customer known, in cost of sales, margin stable with usage
Data rights	D1 to D3	Training data of unclear origin	Documented rights for every dataset used
Dependency	C2 to C4, C6	One vendor, no fallback, no tested migration	Tested ability to switch model with measured quality impact
Operational maturity	E1 to E3, J1	No evaluation set, no monitoring	Versioned evals, live monitoring, a named owner

Again, don't add them up. A 0 on evidence of function or data rights is a deal issue by itself. A wrapper can score 0 on proprietary element and still be a good investment if unit economics and dependency score well: price it as the good software business it is, not as an AI company.

Turning findings into deal terms

FINDING TYPE	USUAL REMEDY	EXAMPLES FROM THE BANK
Exposure	Price, or the growth assumption behind it; a funded repricing plan	K1, K2, K4, K6
Margin re-basing	Valuation on corrected gross margin; comparable set reconsidered	F3, F4, A4, E5
Overstated claims	Corrected disclosure; specific warranties on AI statements; earn-out on AI revenue	A1, A5, A6, B2
Known legal issue	Specific indemnity, retention or price adjustment (W&I won't normally cover known issues)	D2, D3, I3, I4
Fixable operational gap	Condition precedent, or 100-day plan item with an owner and date	C2, C4, E2, G1, I2

FINDING TYPE	USUAL REMEDY	EXAMPLES FROM THE BANK
People	Retention terms in the deal	J1, J4
Upside	Base case in the model with a delayed start; stretch case in the equity story only	L1 to L8, F7

The warranty set for an AI-dependent target should normally include: accuracy of AI descriptions in the disclosed materials; a complete list of third-party models and AI vendors relied on; compliance with those vendors' terms; documented rights to all training data, including personal data; no material manual processing presented to customers as automated; and, if it sells into the EU, its classification under the EU AI Act.

18. Buying an AI vendor: the operator's version

Operators buying an AI product need a shorter, different set. They aren't buying the vendor's margins or team, so most of areas F, J and L fall away. They are taking on the vendor's data handling, liability and continuity, so the data, security and contract questions get sharper. The 22 questions below map to the full bank, so an operator can go deeper on any of them.

QUESTION FOR THE VENDOR	WHY IT MATTERS TO YOU	GET IN WRITING
V1 What exactly does the AI do in our use case, and what is the evidence it works on data like ours? ***Maps to A1, A5***	Vendor accuracy figures are measured on someone else's data.	A pilot on your data with success criteria agreed before it starts.
V2 What share of tasks is reviewed or completed by your staff, and will our data be seen by them? ***Maps to A4, E5***	Human review changes the data protection picture and the vendor's cost base.	Disclosure of human review, who does it and where.
V3 Which model providers and sub-processors will handle our data, in which countries? ***Maps to C1, C3***	Each one is a sub-processor under UK GDPR.	A sub-processor list, with notice and objection rights for changes.
V4 Will our inputs or outputs be used to train or improve any model, yours or a provider's? ***Maps to C3, D4***	Default terms vary and change.	A clear contractual no, or a narrow, opt-in yes.
V5 How long are prompts, files and outputs retained, and can we have them deleted, including from indexes and logs? ***Maps to D4, D7***	Your data may outlive the contract in caches and embeddings.	Retention periods and deletion on request and at exit, with confirmation.
V6 How will you tell us when the underlying model changes, and how will you show quality hasn't dropped? ***Maps to C4***	A model swap can change behaviour overnight.	Notice of material model changes and access to eval results.
V7 What service levels apply to the AI features, and what happens during a provider outage? ***Maps to C6, B4***	Many AI features are excluded from standard SLAs.	SLAs that cover the AI features, with credits.

QUESTION FOR THE VENDOR	WHY IT MATTERS TO YOU	GET IN WRITING
V8 How has the product been tested for prompt injection and data leakage, and can we see a summary? ***Maps to G1, G2***	Your documents and emails become the attack surface.	Summary of an independent test and remediation.
V9 What actions can the AI take in our systems, and with whose permissions? ***Maps to G2***	Agents that act need limits and approvals.	Permission scopes, approval steps and audit logs.
V10 How is our data separated from other customers' data in indexes, caches and fine-tunes? ***Maps to G3***	Shared retrieval indexes are a common leakage route.	A description of tenant isolation and any shared training.
V11 How are wrong answers detected, and what does the product do when it isn't confident? ***Maps to E3, E4***	Liability for a wrong answer to your customer lands on you.	Confidence handling, citations and escalation design.
V12 Is the vendor the provider and are we the deployer under the EU AI Act, and what is the system's risk classification? ***Maps to H3***	Deployers of high-risk systems carry their own duties.	The classification, and the documentation a deployer needs.
V13 Will people be told they are dealing with AI, and is generated content marked? ***Maps to H4***	Article 50 duties can fall on you as deployer.	Product features that support disclosure and marking.
V14 Which certifications cover the AI service: ISO 27001, SOC 2, ISO/IEC 42001, Cyber Essentials Plus? ***Maps to H7***	Certificates are only useful if the scope includes the product you buy.	Certificates with scope statements.
V15 Who is liable if output infringes someone's rights, and is your indemnity capped or conditional? ***Maps to I4***	Indemnities often have conditions that ordinary use breaks.	An IP indemnity covering outputs, with conditions you can meet.
V16 Does the system make decisions about people, and how can a person challenge one? ***Maps to D3, H8***	UK GDPR safeguards on automated decisions apply to you as controller.	Explanation, human review and challenge features.
V17 Has the system been tested for bias in uses like ours? ***Maps to E6***	Equality Act exposure is yours, not the vendor's.	Bias testing results relevant to your use case.
V18 How is pricing calculated as our usage grows, and what stops a runaway bill? ***Maps to F5, F6***	Usage-based AI pricing can surprise finance teams.	Price caps, usage alerts and a rate card for overage.
V19 If we leave, what do we get back, in what format, and how long do we have? ***Maps to C2***	Configuration, prompts and history are often the real switching cost.	Exit assistance and export of your data and configuration.
V20 How many people at the vendor can support and change the AI, and what happens if a key person leaves? ***Maps to J1***	Small AI vendors carry key-person risk you inherit.	Support commitments and, for critical systems, escrow or step-in terms.

QUESTION FOR THE VENDOR	WHY IT MATTERS TO YOU	GET IN WRITING
V21 What does the product do that our existing Microsoft, Google or ChatGPT licences can't? ***Maps to K5***	You may already own most of the capability.	A side-by-side test on your top tasks.
V22 What baseline and metric will we use to decide whether the pilot worked? ***Maps to L3, L4***	Pilots without success criteria drift into renewals.	Success criteria, baseline and a decision date in the pilot agreement.

Contract terms to get before signing

- ☐ No training on your inputs or outputs by the vendor or its model providers, unless you opt in
- ☐ Retention periods for prompts, files, outputs, logs and embeddings, with deletion at exit and on request
- ☐ A sub-processor list including model providers, with notice and a right to object to changes
- ☐ Notice of material model changes, and access to evidence that quality hasn't fallen
- ☐ Service levels that explicitly cover the AI features, with credits
- ☐ An IP indemnity covering outputs, with conditions you can meet in normal use
- ☐ Usage price caps, alerts and a published overage rate
- ☐ Your role and the vendor's under the EU AI Act, and the documentation a deployer needs, if it applies
- ☐ Data and configuration export in a usable format, with an exit assistance period
- ☐ A pilot with success criteria, a baseline and a decision date written into the agreement

For public sector buyers, the government's Guidelines for AI procurement (2020) and the AI Playbook for the UK Government (2025) set the expectations. In legal services, Litig's AI due diligence questionnaire is becoming a common RFP attachment. Answer it if you are a vendor; use this bank alongside it if you are a buyer, because it asks for evidence rather than self-declaration.

19. Worked example

> **Illustrative example.** Ledgerline Ltd is a fictional target. All names, figures and findings are invented to show how the question bank works. The Example tab in the Excel tracker has every question rated for Ledgerline.

The target

Ledgerline Ltd, a fictional target, sells accounts payable automation software to UK and Irish mid-market finance teams. It has 320 customers paying an average of £28,000 a year: £8.96m of ARR. 70% of revenue is priced per AP user seat and 30% by invoice volume tier. Management accounts show 81% gross margin. The IM describes "proprietary AI" with "95% touchless invoice processing". A fund is in exclusivity with three weeks to IC.

Week one: pre-LOI questions and the hands-on session

The pre-LOI questions had been asked before exclusivity. Two answers stood out. On K1, management confirmed the seat split without hesitation but had never modelled what happens if customers' AP teams shrink (K2). On A1, the "proprietary AI" turned out to be a third-party LLM for extraction, Ledgerline's own fine-tuned classifier for coding invoices to the ledger, and rules for approval routing. A sensible design, described with the wrong word.

The hands-on session used 30 real invoices from two friendly finance teams. Ledgerline processed 24 correctly end to end and flagged the other 6 for review; none failed silently. The same 30 invoices through a general assistant with a good prompt produced correct field extraction on most, but only 9 were fully right, because it couldn't match suppliers to the master data or code to the customer's ledger (C5). Ten crafted invoices with hidden text tested indirect injection: in 2 of the 10, the hidden text changed the extracted bank details without a flag (G1). The ERP's two-person approval on bank detail changes stopped that reaching a payment (G2), so it was rated Amber rather than Red.

Weeks two and three: confirmatory findings

FINDING	QUESTIONS	RATING	EFFECT
"95% touchless" was measured on the top 50 customers with clean templates. Platform-wide it is 82%. An exceptions team of 8 touches the rest	A4, A5	A5 Red, A4 Amber	Restated in the IM; specific warranty on AI statements
Inference (£346k a year) booked in platform R&D; the exceptions team (£304k) in customer success	F3, E5	Red	Gross margin re-based from 81% to just under 74%
Heaviest 10% of customers process 40,000 invoices a month but pay higher volume tiers, so margin holds at scale	F4	Green	Gate cleared
The 2019 contract template, still used by 140 customers, says nothing about model improvement. Their invoices are in the pooled training set for the classifier	D1, D2	D2 Red (gate)	Specific indemnity and retention; consent or retraining before completion

FINDING	QUESTIONS	RATING	EFFECT
One LLM provider, never tested on a second. The model snapshot in use is scheduled for retirement in the first quarter of 2027	C2, C4	Amber	Migration plan and eval run as a condition; portability test by day 60
Classifier originally built by a contractor in 2021; the IP assignment was never countersigned	I3	Amber	Signed assignment as a condition precedent
70% of ARR (£6.27m) priced per AP user seat, and the product itself reduces AP headcount	K1, K2	Red	Growth assumption cut; repricing plan funded
22 Irish customers. AP automation isn't an Annex III use; the supplier query assistant is an Article 50 chatbot that doesn't yet disclose it is AI	H3, H4	H3 Green, H4 Amber	Fix before wider launch

The margin arithmetic: £346k of inference plus £304k of review is £650k, or 7.25 points of £8.96m revenue. 81% becomes just under 74%. Nothing here is fraud, and the AI works. But a business with 74% gross margin and a variable cost of goods is valued differently from an 81% one.

The exposure arithmetic: if customers' AP teams need 30% fewer people over the hold and seats follow people, £1.88m of the £6.27m seat revenue is at risk, about 21% of ARR. The irony is that Ledgerline's own product is one of the reasons those teams will shrink. The fix is to move seat customers to per-invoice pricing at renewal, which six customers have already accepted in a test (L6). That goes in the plan, with the cost of doing it.

The value-creation case

Three lines, each traced to the target's own data:

LINE	EVIDENCE	CONSERVATIVE	BASE	STRETCH
Route routine extraction calls to a smaller model (F7)	Engineer estimates 40% of calls are routine; the smaller model costs about a fifth as much. Full saving: 40% × 80% × £346k = £111k	£55k	£83k	£111k
Support triage (L1)	1,200 support hours a month, half routine queries: 600 hours in scope. Capture 20%, 35%, 50% at £32 an hour loaded, less £12k a year to run	£34k	£69k	£103k
Fewer review touches through better confidence thresholds (A4, E4)	Touch rate from 18% to 16%, 14% or 12%; review cost scales with touches, captured by not backfilling leavers	£34k	£68k	£101k
Total run rate	Build cost about £95k across the three lines	**£123k**	**£220k**	**£315k**

The base case goes in the model with realistic starts: routing from month 4, review thresholds from month 6, support triage from month 7, once ticket categories are cleaned up (L2). Year one comes to about £136k of savings against £95k of build: roughly £41k net. Year two is the full £220k. The per-invoice repricing stays out of the price and in the equity story until more customers have accepted it.

What went to the IC

IC LINE	RATING	ONE SENTENCE
Market exposure	Red	70% of ARR is priced per seat on a headcount the product itself reduces; about £1.9m at risk over the hold without repricing
AI claims	Amber	Real, used and better than a general assistant on the full task, but overstated in the IM and resting on a data rights gap for 140 customers
Cost to run	Amber	Margin re-based to about 74% with inference and review in cost of sales; holds at heavy usage
Regulation	Green	Not high-risk under the EU AI Act; one Article 50 disclosure to fix
Opportunity	Amber	£220k base case run rate, costed and sequenced; owners proposed but not yet signed off

The deal went ahead at a lower price, reflecting the re-based margin and the seat exposure, with a specific indemnity and a retention covering the 2019 contract data, the IP assignment as a condition, and the model migration, portability test and AI-scoped penetration test in the first 100 days. On the six claim dimensions Ledgerline scored 2 on evidence of function, 2 on proprietary element, 2 on unit economics, 1 on data rights, 1 on dependency and 2 on operational maturity.

20. Red flags and common mistakes

Red flags in the answers

- Management can't or won't run the product live on your inputs.
- The human touch rate is "none", or nobody can say what the operations team does all day.
- Accuracy is a single headline number with no test set, sample size or date.
- Nobody senior has read the model provider's terms.
- Inference costs are unknown, or booked somewhere other than cost of sales.
- Only the current contract template is in the data room.
- Claims retreat under questioning without anyone noticing: 95% becomes "up to 95%" becomes "in our best cohort".
- The hands-on engineer is unavailable, or every technical question is routed back to the CEO.
- Pilots, trials and design partners are shown as customers.
- "We're a UK company, so the EU AI Act doesn't apply" from a business with EU customers.

Mistakes deal teams make

- **Doing only the claims job.** Confirming the AI is real, and never asking how AI changes the target's pricing, or what it could be worth to the new owner.
- **Treating "wrapper" as a verdict.** Most good AI products call a third-party model. The questions are what the company owns around it, whether margins hold, and what a model change would break.
- **Accepting a verbal answer as evidence.** Rate the evidence. A confident CTO is not a data rights register.

- **Asking only the CTO.** The engineer who maintains the prompts tells you what breaks. Ask them C2 and E2.
- **Relying on a standard penetration test.** It won't have tested prompt injection, excessive agency or cross-tenant retrieval.
- **Averaging scores.** A data rights Red doesn't become acceptable because evals are excellent.
- **Pricing the stretch case.** The vendor then takes the upside and the fund buys the work.
- **Leaving findings in the report.** Each Red and most Ambers should land in the price, the SPA, a condition or the 100-day plan, with an owner.
- **Running the bank as a questionnaire.** Sending 87 questions to management and accepting written replies produces 87 self-assessments. Use the bank to plan sessions and evidence requests, then rate what you see.

21. Glossary

TERM	MEANING
AI washing	Overstating how much a product or business relies on AI, or what its AI does
Agent	An AI system that takes actions (calls tools, changes records, sends messages), not just answers
DPIA	Data protection impact assessment, required under UK GDPR for high-risk processing
Eval set	A library of real cases with known good answers, run against every model or prompt change
Excessive agency	An AI with more permissions or autonomy than its task needs (OWASP LLM06)
Fine-tuning	Further training a base model on specific data to improve a narrow task
Gate question	A question where a Red is a deal issue on its own
Human in the loop	People reviewing or completing some of the AI's work
Inference	Running a model to produce an output, and the cost of doing so
Indirect prompt injection	Instructions hidden in content the AI reads, such as a document or email, that change its behaviour
Open-weight model	A model whose weights are published and can be self-hosted, under a licence
Retrieval (RAG)	Searching the customer's documents and passing results to the model with the question
Thin wrapper	A product adding little beyond a prompt and an interface over someone else's model
Touch rate	The share of tasks a person handles or reviews
Vector store	The index of embeddings used for retrieval; it holds customer data and needs isolation and deletion like any database
W&I	Warranty and indemnity insurance, which normally excludes issues known before signing

22. Sources and further reading

Sources cited

- SEC, first AI washing enforcement actions, 2024, and Presto Automation order, 2025
- FTC, final order against Workado, 2025
- Menlo Ventures, 2025 State of Generative AI in the Enterprise, 2025
- Epoch AI, LLM inference price trends, 2025
- OpenAI, deprecations, and Anthropic, model deprecations
- OWASP, Top 10 for LLM Applications 2025
- EchoLeak: the first real-world zero-click prompt injection exploit in a production LLM system, arXiv, 2025
- McCarthy Tétrault, Moffatt v Air Canada, 2024
- European Commission, AI Omnibus enters into force, 2026; Gibson Dunn, EU AI Act Omnibus agreement, 2026; Cloud Security Alliance, high-risk deadline research note, 2026
- ICO, guidance on AI and data protection
- ISO, ISO/IEC 42001; BSI, UKAS accreditation for ISO/IEC 42001, 2025
- Black Duck, 2026 Open Source Security and Risk Analysis, 2026
- Bird & Bird, Getty Images v Stability AI, 2025; DLA Piper, UK government report on copyright and AI, 2026; UK Supreme Court, Thaler v Comptroller-General, 2023

Other AI due diligence questionnaires worth knowing

Each of these is good at something. None combines evidence requests, good and worrying answers, stages, a scoring method, a hands-on test and the exposure and upside questions in one place, which is the gap this bank is built to fill.

RESOURCE	WHAT IT DOES WELL	WHERE IT STOPS
Litig AI due diligence questionnaire (2024)	Detailed supplier questions in six sections, including EU AI Act risk; widely used in legal procurement	Self-declared answers, no evidence standard or scoring; written for buyers of AI, not acquirers
FS-ISAC generative AI vendor evaluation (2024)	Risk-tiering workbook that sets the depth of vendor diligence	US financial services framing; vendor risk only
CSA AI Controls Matrix and AI-CAIQ (2025)	243 control objectives across 18 domains, mapped to ISO/IEC 42001 and NIST AI RMF	A security and governance controls framework; nothing on commercial claims, margin or exposure
UK Guidelines for AI procurement (2020) and AI Playbook (2025)	Sound principles on data assessment, explainability and lifecycle management	Public sector procurement principles rather than a question list
Lexis+ AI due diligence questionnaire, private M&A, with Simmons & Simmons	A legal precedent for gathering AI documents into the data room	Subscription only; legal scope
ITechLaw AI and IP due diligence checklist (2024)	Specialist IP questions for AI transactions	IP only; gated

RESOURCE	WHAT IT DOES WELL	WHERE IT STOPS
MME AI due diligence questionnaire (2025)	Broad legal coverage across 13 areas, including employment and liability	Questions not published; legal scope
NCSC guidelines for secure AI system development and NIST AI RMF	The reference frameworks for secure design and AI risk management	Frameworks for builders, not diligence question lists

Related Anyway Ventures guides

- AI due diligence for private equity: the three jobs, scoping and the IC summary
- AI washing in due diligence: the verification protocol and the six-dimension score
- Wrapper or moat?: defensibility and dependency, with 12 questions
- AI due diligence as a value-creation plan: what the AI section of the IC paper should hold
- Beyond the demo: how investors judge an AI product on evidence
- Technical due diligence checklist: the wider engineering, security and IP review
- Technology exit readiness checklist: what a seller should fix before a buyer asks these questions

23. About this resource

Written by Anyway Ventures, the York technology and AI practice led by Brannan Coady, who has been on both the buying and selling side of technical due diligence. Version 1.0, October 2026; the regulatory dates reflect the EU AI Act as amended by the Digital Omnibus in July 2026. Free to use, adapt and share with attribution. Not legal, tax or investment advice. The question bank sits behind our AI due diligence work.

ABOUT ANYWAY VENTURES

Senior hands for technology and AI decisions

Technology and AI work for private equity funds, venture investors and UK mid-market management teams: before a deal, during the hold and at exit. Every engagement is led by Brannan Coady, an exited founder and former CEO. Fixed price, built to be handed over.

FOR INVESTORS

Technical due diligence

Code, architecture, security, IP and one-to-one team interviews, every material finding costed. Red-flag review in five to seven days.

ANYWAYVENTURES.COM/TECHNICAL-DUE-DILIGENCE

AI due diligence

Whether a target's AI claims are real, defensible and theirs, and a costed AI value-creation plan for the IC.

ANYWAYVENTURES.COM/AI-DUE-DILIGENCE

AI disruption risk review

Is AI a threat to the business or its moat? One deal, or every holding in a portfolio ranked by exposure.

ANYWAYVENTURES.COM/AI-DISRUPTION-RISK

Technical due diligence for VCs

A two to three day check before the term sheet: the product tested on evidence, not the demo.

ANYWAYVENTURES.COM/VC-TECHNICAL-DUE-DILIGENCE

Retained AI operating partner

Senior AI capacity held by the fund and deployed into the company with the strongest AI value case.

ANYWAYVENTURES.COM/AI-OPERATING-PARTNER

Exit readiness and vendor technical due diligence

Find what a buyer's diligence will find, 12 to 24 months out, fix it, then publish a report buyers can rely on.

ANYWAYVENTURES.COM/EXIT-READINESS

FOR OPERATORS

AI audit and readiness assessment

Two to three weeks to a ranked, costed list of AI opportunities, the gaps to close and a 90-day plan.

ANYWAYVENTURES.COM/AI-AUDIT

Embedded AI delivery

Internal tooling, automation and AI agents in production inside your business, built with your team.

ANYWAYVENTURES.COM/EMBEDDED-AI-DELIVERY

Fractional Chief AI Officer

One senior person who owns AI on your leadership team, two to four days a month.

ANYWAYVENTURES.COM/FRACTIONAL-CHIEF-AI-OFFICER

Hands-on AI training

In-person training where every attendee builds a working tool on their own job.

ANYWAYVENTURES.COM/AI-TRAINING

Board and NED roles

Non-executive seats where technology, product and AI decide the outcome.

ANYWAYVENTURES.COM/BOARD-AND-NED

Every enquiry starts with the same free 30-minute call.

A straight answer on whether we can help, what it would cost and how long it would take.

EMAIL
brannan@anywayventures.com

BOOK A CALL
anywayventures.com/contact

MORE RESOURCES
anywayventures.com/resources

Anyway Ventures Limited · Registered in England and Wales, company number 16318119 · Registered office: Newbury House, Main Street, Little Ouseburn, York, YO26 9TD. This resource is general guidance, not legal, tax or investment advice. Free to use, adapt and share with attribution to Anyway Ventures. Version 1.0, October 2026.