

PARTICIPANT PACKET

Lighting Up Vendor AI Risk

The BEACON™ Framework for Executives and Risk Leaders

Most organizations don't decide to adopt vendor AI. It arrives as a feature update on a vendor they already trust. BEACON gives executives six categories of risk to ask about about any vendor AI that touches your data or your decisions, and a simple way to decide what to do about the answers.

What's inside

- BEACON at a glance
- Vendor question bank (18 questions)
- Rating scale and decision rule
- Regulatory snapshot
- Case study: we ran BEACON on ourselves
- Your 30-day action plan

The big idea: You can outsource the AI. You cannot outsource the outcome.

BEACON at a glance

Letter	The executive question	Red flag to listen for
B Bias	Could this vendor's AI treat our customers, applicants or employees unfairly, and would we know?	"We don't use race, gender or age, so the model can't be biased." Proxies such as zip code or graduation year can carry the same signal.
E Explainability	If a regulator, customer or judge asks why the AI decided something, can we answer?	"That's proprietary." A black box you can't explain becomes your liability, not the vendor's.
A Access & Data	What of our data goes into the AI, where does it go next, and does it train anyone's model?	A new AI subprocessor announced only on a web page, or a "right to use de-identified data to improve services" buried in the contract.
C Contractual Accountability	When the AI gets it wrong, who pays, and what are we entitled to?	"Beta" or click-through AI terms that override the negotiated contract.
O Oversight	Who is watching this AI, at the vendor and inside our own walls?	Consequential decisions made fully automatically, on by default, with no named owner.
N Non-determinism	Will the AI give the same answer twice, and what happens when it doesn't?	The vendor can't pin a model version and re-tests only "major" upstream updates.

Vendor question bank

Ask these of any vendor whose AI touches your data or influences decisions about people. Ask for evidence, not assurances.

B Bias

- ☐ How do you test outputs for discriminatory or skewed results affecting our customers?
- ☐ How often do you re-test, and what triggers a re-test?
- ☐ Has any independent party audited this model for disparate impact?

E Explainability

- ☐ Can you document or justify how the model reached a given decision, for audit purposes?
- ☐ What are the documented limits of this model: where should it not be used?
- ☐ Can we see the model card or equivalent documentation?

A Access & Data

- ☐ Does our data train your model? Is it segregated from other customers' data, and what's the retention and deletion policy?
- ☐ Is any of our data, including de-identified data, used to train or tune a model?
- ☐ How long are prompts and outputs kept, and who can see them?

C Contractual Accountability

- ☐ Who is liable when the AI produces a harmful or incorrect output, and what does the SLA actually say about AI-specific incidents?
- ☐ Which document governs the AI feature: our contract or separate AI terms?
- ☐ What indemnity do you offer for discrimination or IP claims caused by the AI?

O Oversight

- ☐ Is there a human in the loop, and can we override or escalate a bad output?
- ☐ What decisions happen without a human, and can we require human review?
- ☐ What monitoring reports will we receive, and how often?

N Non-determinism

- ☐ What's your measured hallucination or failure rate, and how do you monitor drift over time?
- ☐ Can we stay on a model version, and how much notice do we get before it changes?
- ☐ What guardrails stop a bad output from reaching a customer?

Rating scale and decision rule

Green	A specific, complete answer backed by evidence. No action beyond normal monitoring.
Yellow	Partly specific or partly evidenced. Gaps can usually be closed with conditions: contract language, configuration or follow-up evidence.
Red	Vague, evasive or no answer. “We haven’t thought about that” is itself a finding.

Three decisions: approve; approve with conditions; not yet / reject.

BEACON is not pass/fail. Each category is scored on how specific and complete the vendor’s answer is. Two or more Reds trigger deeper review before sign-off, however well the vendor scored on the standard security questionnaire.

Regulatory snapshot

As of September 2026. Not legal advice; confirm with counsel.

In force since Aug 2025	EU AI Act: prohibited practices, AI literacy and general-purpose AI model obligations
Aug 2026	EU AI Act: transparency duties (tell people when they interact with AI)
Jan 1, 2027	Colorado automated decision-making law (SB 26-189): notice, adverse-outcome explanations, human review
Dec 2, 2027	EU AI Act: high-risk obligations for standalone systems such as hiring and credit, deferred by the 2026 Digital Omnibus

Frameworks to ask vendors about: NIST AI Risk Management Framework (AI RMF) and ISO/IEC 42001 (AI management systems).

Case study: we ran BEACON on ourselves

Vendor and product names are changed. Findings are TPTH's, for TPTH's use, from published vendor documentation only, dated 24–25 September 2026. Worked reasoning, not a rating of any real product.

The situation

- Third Party Threat Hunting (TPTH) is a small advisory firm. In September 2026, during a review of our own sub-processors, we ran BEACON on ourselves.
- First we listed our AI. We found five AI surfaces. Two weren't on anyone's list: AI that arrived inside an e-signature tool, and our own AI assistant.
- The one hardest to be honest about was the assistant: Orrery Assistant, on its commercial (business) tier. We use it for drafting, analysis and code, and it has access to our website code and document drive. We bought it on purpose and had never assessed it.
- The six evidence cards summarize what the vendor publishes. We assessed from published documentation only; we didn't contact the vendor.

Your team's use

Team A: you are TPTH	The assistant drafts documents, analyzes material and writes code. It makes no decisions about people. A named person at TPTH is accountable for every deliverable.
Team B: you are a regional lender	You want the same assistant, same tier, same contract, to pre-screen loan applicants and draft adverse-action letters to the people it screens out.

Score the evidence for your team's use only.

Evidence cards

What the vendor publishes about Orrery Assistant, business tier, by BEACON letter.

B Bias	The vendor publishes a political even-handedness evaluation built on paired prompts across many topics and task types, reports its own and competitors' scores, and has open-sourced the test so customers can re-run it. There is no published benchmark of outcomes by demographic or protected characteristic.
E Explainability	A compliance API records who asked what and what the assistant answered: user, IP address, device, timestamps and full session content, exportable to a security monitoring tool. Nothing records why a particular output was produced: no reason codes, no feature attribution, no decision trail.
A Access & Data	On the commercial tier, the contract prohibits training on customer content. API data is kept 30 days, and the vendor acts as a processor under a public, self-serve data processing agreement. The consumer tier has the opposite default. Zero data retention is available only by negotiation, and since June 2026 data from newer models is kept 30 days for safety work even under those agreements. Storage is U.S.-only; there is no EU residency option.
C Contractual Accountability	Liability is capped at the previous 12 months' fees. Output accuracy is expressly disclaimed: outputs are provided "as is," with no warranty that they are accurate or complete. There is no AI-specific incident clause. On the other side: an uncapped IP indemnity, and the customer owns its inputs and outputs.

O Oversight	The vendor's usage policy requires a qualified professional to review advice or subjective decisions affecting individuals before they are shared, and requires disclosure when outputs are shown to consumers. These are binding conditions of use. (For Team A only: TPTH's own policy manual separately requires a named human accountable for every deliverable.)
N Non-determinism	No published hallucination or error rate, and no frontier AI vendor publishes one. The vendor publishes model cards and safety evaluations, offers dated model versions that can be pinned so behavior doesn't shift underneath you, and gives at least 60 days' notice before retiring a model.

Your group's scoring sheet

Circle your team: Team A (TPTH) Team B (lender)

Letter	Rating	Evidence that drove it	Control if Red
B Bias	G Y R		
E Explainability	G Y R		
A Access & Data	G Y R		
C Contractual Accountability	G Y R		
O Oversight	G Y R		
N Non-determinism	G Y R		

Number of Reds: _____ Two or more means deeper review before sign-off.

Before you leave: list your own AI

The step most firms skip. For each line, name at least one example at your organization.

Where AI hides	At my organization
A feature added to a tool you already used	
AI inside a tool you bought for something else	
AI your vendors' vendors run on your data (fourth parties)	
Your own AI tools, bought on purpose and never assessed	
Anyone using a consumer-tier AI account for work	

Your 30-day action plan

Circle the action you will start first and write a date next to it.

Action	What it means	My date
1. Find it	List your AI before you assess any of it: features added to tools you already own, AI inside tools bought for something else, your vendors' vendors and your own tools. Then ask your top 20 critical vendors in writing what AI is in their service, and on which models.	Start by:
2. Contract for it	Add AI clauses to every renewal: notice of new AI features and model changes, no training on your data, AI incident notification and audit rights.	Start by:
3. Govern it	Name an owner for each vendor AI use, set a risk appetite for AI that affects people and give your board a one-page BEACON heat map each quarter.	Start by:

My first question to a vendor

Vendor:

Question:

Go deeper. The full-day BEACON workshop covers vendor tiering, the complete question bank, contract clause language and building your board report.

Gregory Rasner | Third Party Threat Hunting LLC | Third Party Threat Hunters podcast | Fourth Party newsletter