



AI Discovery Under the EU AI Act: An Operating Briefing

The EU AI Act is now part of the operating backdrop for AI programmes, but AI discovery should not be reduced to paperwork. Policy assistants, contract search, case memory, and internal knowledge systems need practical controls when answers influence decisions: inventory, decision-influence tiering, provenance, human gates, evaluation, and owners. This briefing translates the Act context into an operating packet for boards and technology leaders.

David Saliba . 2026-06-30 . aimonger.com/whitepapers/ai-discovery-eu-governance/

The problem in one sentence

AI discovery systems look like search but behave like advisers. Boards need inventory, decision-influence tiers, and human gates, not a legal memo that nobody updates when the vendor ships a new retrieval mode.

AI discovery, plainly: systems that find and synthesise enterprise knowledge (contracts, policies, cases, product specs) at the moment of work, often RAG-shaped and increasingly agent-shaped. **In practice:** “Have we agreed this liability cap before?” answered in seconds from governed archives, with provenance, not a three-day email chain.

Decision influence, in short: how much a system’s output can change outcomes if wrong, from internal convenience to regulated or customer-facing commitments. **Worked case:** a wrong FAQ paraphrase is annoying; a wrong regulatory draft sent externally is a reputational and legal event.

Discovery systems are easy to underestimate

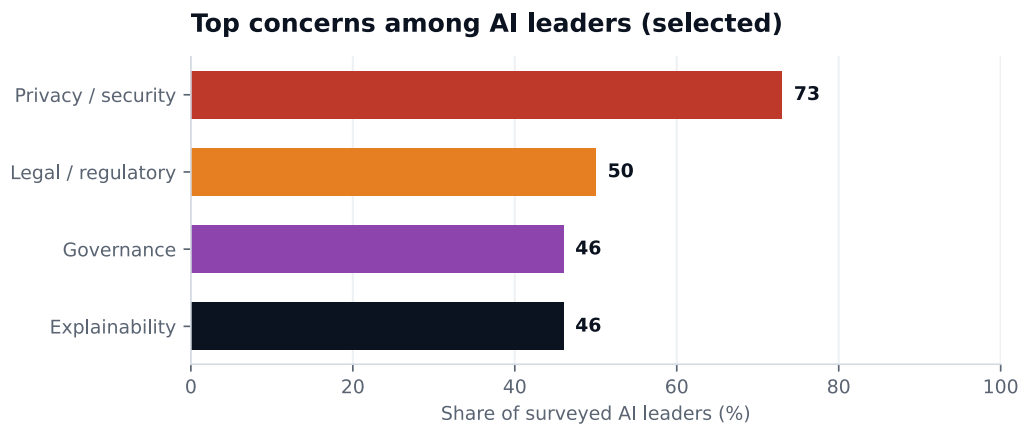
They feel like search. They behave like advisers.

In a mid-market firm, the first AI discovery systems often arrive through familiar routes: a vendor knowledge assistant, a policy search pilot, a contract Q&A tool, or a case-memory layer built by the CIO team. Because the interface looks like search, leadership can miss the operating reality: the answer may influence a customer commitment, an HR question, a regulated interpretation, or a draft sent outside the firm.

The global narrative often treats “AI governance” as paperwork: policies, forms, and committee slides. Operators need something more useful. They need discovery controls that work at the moment of use: what source was read, which version was current, who reviewed the answer, and what happens when the system is wrong.

Regulation (EU) 2024/1689 sets a staggered framework through 2027. AI literacy obligations applied from 2 February 2025. Most remaining provisions apply from 2 August 2026, with further phased duties thereafter. Boards should not wait for a court case to invent ownership. Even when an internal knowledge assistant is not Annex III high-risk, decision influence still demands an audit story.

Internal knowledge assistants are not automatically high-risk; classification depends on intended purpose, Annex I/III scope, and material influence on decisions. The GPAI Code of Practice (final published 10 July 2025) is a voluntary mechanism aimed primarily at GPAI model providers. It does not automatically satisfy an enterprise deployer's obligations for a discovery system.

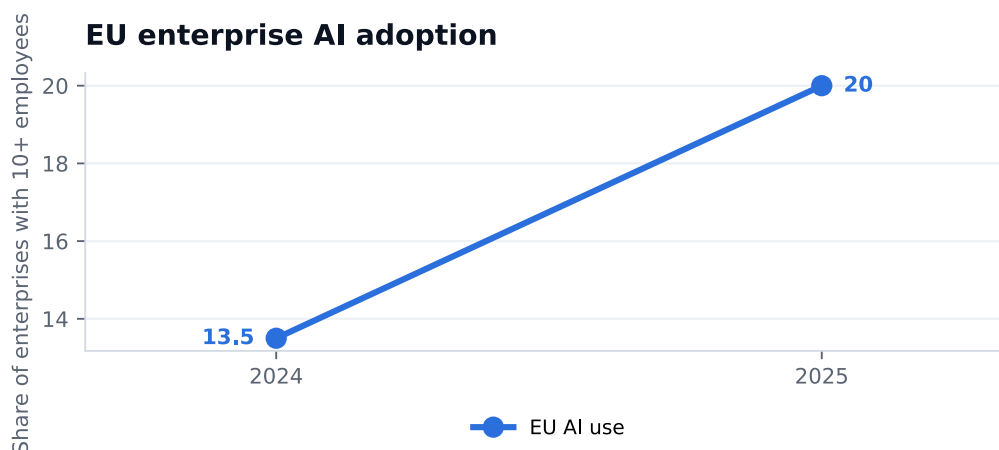


Deloitte State of AI 2026 (survey). AIMonger redraw.

Figure 1. Top concerns among surveyed AI leaders (selected). Plain read: discovery systems touch the worries boards already name, privacy, legal exposure, governance. **Operating case:** your contract Q&A tool is a data-processing and oversight question, not only a search upgrade. Source: Deloitte State of AI in the Enterprise 2026 (survey). AIMonger redraw.

Deloitte reports data privacy and security cited by 73 per cent of surveyed AI leaders as a top concern, legal/IP/regulatory by 50 per cent, governance oversight by 46 per cent, and model quality/explainability by 46 per cent. Discovery programmes sit at the intersection of all four.

Eurostat shows EU enterprise AI use rose to 20.0 per cent in 2025 from 13.5 per cent in 2024. Governance debt accumulates as assistants spread faster than inventories.



Eurostat, 11 Dec 2025 (official statistic). AIMonger redraw.

Figure 2. Share of EU enterprises (10+ employees) using AI. In short: more firms deploy AI each year; fewer maintain a discovery inventory. In practice: ask for the inventory slide before the adoption celebration slide. Source: Eurostat, 11 December 2025 (official statistic). AIMonger redraw.

Inventory fields

Field	Why
System name / vendor	Know what exists
Workflows served	Tie to absorption portfolio
Data sources	Residency and permissions
Decision influence tier (D0-D3)	Sets control intensity
Human gate	Who overrides
Logging	Incident readiness
Eval status	Trust readiness
Owner	Accountability

Provenance means: a record of which documents, versions, and retrieval steps produced an answer.

On Monday: the UI shows “Handbook v2024-03, sections 4.2-4.5”, not a confident paragraph with no source.

Maintain one inventory row per system, including shadow SaaS discovered in audits. Gartner forecasts heavy agentic cancellation when risk controls are weak; discovery systems that grow tool use inherit that failure mode.

Decision-influence tiers for discovery

Tier	Example	Controls
D0	Public FAQ paraphrase	Basic logging
D1	Internal policy guidance	Provenance + training
D2	Customer-facing draft	Human gate before send
D3	Regulated interpretation support	Gate + enhanced docs + eval

Most under-governance happens when D2/D3 systems are treated as D0 because the UI looks like search.

Human gate, put simply: a named role or rule that must approve before an AI draft becomes an official action or external communication. **Operating case:** no customer pricing email sends without a finance owner click, even if the model drafted it perfectly yesterday.

Operating controls that match discovery risk

- 1. User transparency where required.** Staff see clear disclosure when discovery output is AI-assisted, so reliance, override, and escalation behaviour matches regulatory and internal policy expectations at the moment of use. Transparency is not a label buried in a tooltip; it is a visible signal that the answer came from a model and may be wrong. Without it, staff treat synthesis like static search results and skip the scepticism a draft deserves.
- 2. Provenance on consequential answers.** High-influence answers show document version, section, and retrieval path, so a wrong paraphrase can be traced before it becomes an external commitment or HR decision. Provenance is what turns “the AI said” into “the 2024-03 handbook section 4.2 said”, which is fixable. Without it, a wrong answer is an orphan nobody can trace to a source.
- 3. Human review before external send.** Customer, regulator, and partner-facing drafts pass a named gate before send, so D2 and D3 systems cannot behave like casual search boxes on high-stakes text. The gate is where a human owns the outcome, not a formality. A D2 draft that skips the gate is a regulator incident waiting for its date.

4. **Log retention and access controls.** Query and gate logs follow retention and access rules, so incident response can reconstruct events without exposing sensitive content to unauthorised administrators. Logs are the artefact the regulator asks for first, and they must be readable without handing the whole corpus to every admin. Retention caps also keep old discovery queries from living forever in a SaaS bucket outside firm control.
5. **Vendor subprocessor review.** Subprocessors, residency, and training-use terms are reviewed on renewal, so SaaS discovery tools do not silently expand data processing beyond the inventory row. A subprocessor change is a data-processing event under GDPR, and the review is how the firm catches it before auto-renewal locks it in. The inventory row is the artefact the review updates.
6. **Incident notification path.** A documented severity ladder and owner list exists before promotion, so the first harmful answer triggers response instead of a cross-functional email chain started from zero. The path is rehearsed, not written and forgotten, because an untested path fails on the first real incident. A discovery system without an incident path is a system that will fail opaquely.
7. **Knowledge trust scorecard alignment.** Holdouts, temporal marking, and refusal rules merge with governance review, so paperwork and evaluation argue from one artefact instead of diverging checklists. Two artefacts that disagree are how a system passes legal review and fails in production. One combined review means legal and engineering argue from the same evidence.

The control set should be light for low-consequence search and stricter when discovery shapes external commitments, employee outcomes, or regulated work.

NIST AI 600-1 recommends inventorying GenAI systems with provenance, model/version/access mode, human oversight, known issues, and supplier context. This is practical for discovery assistants even outside US jurisdiction.

Documentation pack (practical)

Keep a living folder per system:

- **Purpose and out-of-scope uses.** Document what the system is for and what staff must not use it for, so tier disputes resolve against written scope instead of UI assumptions about “harmless search.” The scope line is what stops a D0 FAQ tool from quietly absorbing a D2 client-reporting workflow. Without it, the tier is whatever the user assumes it is.
- **Data sources and refresh rules.** List corpora, sync cadence, and permission dependencies, so stale or over-broad sources do not invalidate answers between quarterly reviews. A source that synced six months ago is a source that stopped being current five months ago, and stale sources produce confident wrong answers. The refresh rule is how the firm knows whether the answer is still grounded.
- **Roles and human gates.** Name owners, reviewers, and override rules, so consequential drafts never leave the building without an accountable human when tier D2 or D3 applies. A gate without a named owner is a rule nobody enforces, and a D2 draft without a gate is an external commitment on autopilot. The names are what make the gate operational.
- **Evaluation summary and promotion date.** Record holdout results and last promotion decision, so trust claims rest on measured behaviour, not vendor marketing slides alone. A trust claim without a holdout result is marketing, and a promotion without a date is a decision nobody can audit. The summary is what lets the board ask “how do you know it works.”
- **Known limitations.** Maintain an honest limitations section covering temporal drift, language gaps, and refusal gaps, so operators know when to escalate instead of trusting confident wrong answers. A system that does not document its gaps is a system that will be trusted past its competence. The limitations section is how the firm teaches operators to escalate.

- **Incident and near-miss history.** Log prior harmful outputs and responses, so repeat failure modes feed eval priorities instead of disappearing when staff turnover. A near-miss that is not logged is a lesson that is not learned, and the same failure recurs because the next operator never saw it. The history is the firm's memory of what the system actually does wrong.
- **Vendor DPAs and subprocessors.** Keep data-processing terms, subprocessors, and audit-export rights current, so renewal negotiations happen before an incident exposes a gap. A DPA that is six months stale is a subprocessor list that is six months wrong, and the gap is discovered during the incident, not before. The current terms are what let the firm notify clients and regulators within its own obligations.

This is useful even when a formal conformity assessment is not required.

For a Malta-origin or mid-market firm, the pack should be small enough to maintain. A folder nobody updates is theatre. A living owner, tier, source list, evaluation note, and gate rule is operational evidence.

Worked example: Luxembourg fund services firm

Composite: 180 staff, multi-jurisdiction clients, SharePoint plus specialist DMS, vendor policy assistant plus internal case-memory pilot.

Month 1 mistake: treat both as D0 “search” because the UI has a query box.

Classification after worksheet:

- **Vendor assistant on public FAQs (D0).** Public FAQ paraphrase stays at D0 with basic logging, because wrong answers annoy staff but do not create external commitments or regulated interpretations. The tier is low because the blast radius is low, and basic logging is the proportionate control. A wrong FAQ answer is corrected in the next staff chat, not in a regulator letter.
- **Internal policy guidance (D1).** Employee handbook guidance runs at D1 with provenance and training, because operators rely on it for HR and operational decisions even when customers never see the text. The tier rises because the answer influences an internal decision, and provenance is what lets the operator verify the section before acting. Without provenance, a wrong handbook paraphrase becomes a wrong HR decision.
- **Client reporting draft assistant (D2).** Client reporting drafts require a human gate before send, because a wrong figure or commitment reaches customers and auditors, not only internal readers. The gate is non-negotiable because the blast radius is external, and the draft is a commitment if sent. A D2 draft without a gate is a regulator incident on a timer.
- **AML pattern discovery (D3).** AML case-note discovery sits at D3 with enhanced evaluation and legal review, because pattern suggestions can influence regulated decisions and enforcement exposure. The tier is highest because the output touches a regulated decision, and the controls include legal review, not just a gate. A D3 system without enhanced eval is a system the firm cannot defend in front of a regulator.

Operating packet per system: one-page inventory row, five-page living folder, quarterly review on tier disputes. Board sees tier counts, not legal jargon.

Outcome: faster policy answers for operators; no external send without gate; counsel engaged only on D3 promotion, not on every pilot slide.

Alignment with the knowledge trust scorecard

Governance checklists that ignore holdouts and provenance create paperwork without safety. Merge the trust scorecard into review, so legal, risk, engineering, and the business argue from one artefact.

Documentation without evaluation is incomplete. Evaluation without documentation fails audit. Run one combined review artefact: inventory plus trust scorecard plus tier plus owner signatures.

Training and transparency

Staff must know:

- **When discovery is AI-assisted.** Operators recognise AI-assisted answers in the UI, so they apply the right scepticism, override habits, and escalation path instead of treating synthesis like static search results. The signal is what tells the operator to check the provenance before acting, and without it the answer is trusted by default. A system that hides its AI assistance is a system that will be over-trusted.
- **Paste rules by system and data class.** Staff know what they may paste into which approved tool, so client confidential material does not enter consumer tiers during deadline pressure. The rule is what stops a paralegal from pasting a live contract into a public chatbot at 11 p.m. before a filing. Without it, the deadline chooses the tool, and the tool chooses the data breach.
- **Mandatory human review triggers.** Consequential tiers publish when review is mandatory before send or decision, so D2 and D3 workflows cannot skip gates because the interface felt fast. The trigger is a rule, not a suggestion, and it is what stops a D3 draft from leaving the firm on a fast interface. A gate that is optional is a gate that will be skipped.
- **Harmful-answer escalation.** Everyone knows how to report a harmful or non-compliant answer without fear of blame, so near-misses feed evaluation instead of staying in private chats. The escalation path is how a near-miss becomes an eval case instead of a war story. A firm that punishes reporting is a firm that never learns what its system actually does wrong.

Transparency is an operating control, not only a statutory phrase. EU AI Act literacy duties from February 2025 imply organisations must ensure staff understand AI use at an appropriate level. Discovery training is part of that, not a separate “AI week.”

Vendor addendum requirements

For SaaS discovery tools, contract for:

- **Subprocessors and residency.** Subprocessor lists and data residency commitments must match inventory tier expectations, so a D2 client-reporting tool cannot process EU client data in an unapproved region. The match is what stops a vendor from quietly moving data to a cheaper region between reviews. A residency commitment that names the region is a control; “best efforts” is not.
- **Training-use prohibition or opt-out.** Contracts prohibit vendor training on firm prompts or provide an enforceable opt-out, so confidential discovery queries do not become model improvement fodder. A training-use clause buried in a broad licence grant lets the vendor improve its model on your data and sell the improvement back to you. The prohibition or opt-out is what keeps the data boundary a conscious decision.
- **Audit log export.** Vendors must provide CSV or API audit export for D2 and D3 systems, so gate and query review does not depend solely on a vendor UI screenshot during incident response. The export is the artefact the regulator asks for first, and a screenshot is not evidence. A vendor that refuses export is a vendor that owns your incident response.
- **Retention limits.** Log and prompt retention caps align with policy, so old discovery queries do not live forever in a SaaS bucket outside firm control. A retention cap is what stops a 2024 prompt from surfacing in a 2027 data subject access request. Without it, the SaaS bucket is a permanent record the firm cannot purge.
- **Incident notice periods.** Contractual notice windows for breaches and subprocessor changes are short enough for the firm to notify clients and regulators within its own obligations. The notice

window is what lets the firm meet its own GDPR timeline, not the vendor's convenience. A 90-day notice window is a firm that cannot meet a 72-hour breach duty.

- **External evaluation access.** Where feasible, firms retain rights to run holdout evaluation outside the vendor UI, so promotion decisions rest on independent measurement, not dashboard defaults alone. Independent measurement is how a firm knows the vendor dashboard is honest. A vendor that blocks external eval is a vendor asking you to trust a number you cannot reproduce.

On Monday: renewal blocked until vendor provides CSV audit export for D2 systems.

Evidence base: inventories matter even outside Annex III

Source	Finding	Governance implication
EU AI Act	Risk-based duties; transparency; phased timeline	Classify by influence, not UI label
NIST AI 600-1	Inventory, oversight, provenance	Discovery pack maps cleanly
Deloitte 2026	Privacy 73%; governance 46%	Boards already worry about discovery risks
Eurostat 2025	20.0% EU enterprise AI use	More systems to inventory each year
Gartner 2025	>40% agentic cancellation on weak controls	Tool-enabled discovery needs gates
McKinsey 2025	High performers >5% EBIT with scaling discipline	Trust controls enable scale

Plain read: legal classification matters, but boards can act now on tiers, gates, and inventory without waiting for final counsel memos on every SKU.

Decision-influence classification worksheet

For each discovery system, score:

- **External draft capability.** Score higher if the system can draft customer, regulator, or partner text that could bind the firm if sent without review, because external drafts demand D2 gates at minimum. A system that can draft a customer email is a system that can bind the firm, and the score reflects that blast radius. The score is what sets the tier, and the tier sets the controls.
- **HR, credit, or access influence.** Score higher when outputs can influence hiring, credit, access, or benefits decisions, because those pathways attract scrutiny even when Annex III labels are debated. A system that shapes an HR decision is a system that shapes a person, and the scrutiny follows the influence, not the UI label. The score is what stops a D0-looking tool from quietly making D3 decisions.
- **Special-category or client-confidential data.** Score higher when corpora include special-category personal data or client-confidential matter files, because residency and minimisation rules tighten regardless of UI simplicity. A system that reads special-category data is a system under GDPR Article 9, and the rules do not care that the UI looks like search. The score is what forces the residency conversation before the tool goes live.
- **Use without specialist review.** Score higher when business users routinely act on answers without domain review, because unchecked synthesis at scale is where discovery systems behave like unmanaged advisers. A system that business users trust without review is a system that has no human gate in practice, regardless of what the policy says. The score is what surfaces the gap between the written gate and the actual workflow.

Higher scores raise documentation, human-gate, and evaluation obligations. Keep the worksheet with the inventory.

Counter-position: governance paperwork kills speed

Another board might argue tiering and folders slow pilots. Unmanaged speed creates incident and regulatory exposure. Use tiered controls: D0 FAQ assist differs from D3 regulated interpretation support. Speed belongs in low-consequence tiers; documentation intensifies with decision influence.

Deloitte finds many organisations still struggle to move experiments to production. Governance theatre (slides without owners) is slower than tiered operating packs, because incidents force rework.

Quarterly operating review agenda

- 1. Inventory changes.** Review new SaaS tools, retired systems, and shadow discoveries surfaced in audits, so the inventory matches what staff actually use this quarter. An inventory that does not match reality is a document, not a control, and the review is what keeps it alive. The audit is where shadow SaaS is found before it finds the firm.
- 2. Tiering disputes.** Resolve business-versus-risk disagreements when a workflow is labelled D1 versus D2, so promotion does not proceed on optimistic UI labels alone. The dispute is where the tier gets honest, and the resolution is what keeps a D2 workflow from running on D0 controls. A tier that is never disputed is a tier nobody questioned.
- 3. Incidents and near-misses.** Walk harmful outputs and near-misses from the quarter, so eval priorities and gate rules update before repeat failures reach customers. The walk is how a near-miss becomes an eval case, and the eval case is how the next incident is prevented. A quarter with no walk is a quarter with no learning.
- 4. D2 and D3 evaluation status.** Confirm holdout summaries and promotion dates for high-influence systems, so trust claims rest on current measurement, not last year's pilot slide. A holdout that is a year stale is a trust claim that is a year old, and the confirmation is what forces a refresh. The promotion date is what tells the board whether the system was tested recently.
- 5. Vendor term changes.** Flag subprocessor, residency, or training-use changes on renewal, so legal review happens before auto-renewal locks in unacceptable terms. A subprocessor change that auto-renews is a data-processing event the firm did not authorise, and the flag is what catches it. The renewal calendar is where the terms are decided, not after.
- 6. Training completion.** Report discovery and data-class training completion by role, so literacy duties and operating controls stay aligned with deployed tiers. A system deployed to untrained operators is a system that will be misused, and the completion rate is what tells the board whether the training is real. The rate by role is what surfaces the gap.
- 7. Board and risk-committee decisions.** Escalate tier promotions, incidents, and resource asks that require board or risk-committee decision, instead of deferring them to informal email threads. An escalation that stays in email is a decision that stays undocumented, and the committee is where the decision gets owned. The agenda item is what makes the decision visible.

CIO decision criteria

Criterion	D0/D1 pass	D2/D3 pass
Inventory row complete	Yes	Yes
Owner named	Yes	Yes
Provenance shown to user	Recommended	Required
Human gate documented	Optional	Required
Eval holdout summary	Basic	Required before promotion
Audit export tested	Optional	Required
Incident path exercised	Tabletop	Live run in last 12 months

Defer D2/D3 production promotion if any required D2/D3 row fails.

Link to agentic discovery

Discovery systems increasingly call tools (create ticket, file draft, trigger workflow). Tool use raises interoperability and OWASP agentic risks. Inventory must note tool permissions; tiers may rise when write tools attach.

See the companion interoperability paper for gateway patterns when discovery agents multiply.

Closing position

EU AI Act work for AI discovery is not a legal textbook exercise. It is how firms keep speed without losing defensibility.

Inventory. Gate. Document. Train. Then scale.

When Eurostat shows AI spreading through EU enterprises and Deloitte shows leaders already rank privacy and governance as top concerns, discovery programmes are the early test of whether your AI operating model is serious, or still treating search boxes as harmless.

References

1. Regulation (EU) 2024/1689 of the European Parliament and of the Council (EU AI Act). <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
2. European Commission, AI Act timeline and AI literacy duties. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
3. NIST, “Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile,” NIST AI 600-1, July 2024. <https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.600-1.pdf>
4. Deloitte AI Institute, “The State of AI in the Enterprise: The Untapped Edge” (2026 edition); survey of 3,235 leaders, Aug-Sep 2025. <https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html>
5. Eurostat, “20% of EU enterprises use AI technologies,” 11 December 2025. <https://ec.europa.eu/eurostat/en/web/products-eurostat-news/w/ddn-20251211-2>
6. Gartner, “Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027,” press release, 25 June 2025. <https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027>

-
7. McKinsey & Company / QuantumBlack, “The State of AI: Global Survey 2025.” <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
 8. European Commission, GPAI Code of Practice (voluntary for GPAI providers; not a deployer compliance substitute). <https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice>
 9. OWASP GenAI Security Project, “OWASP Top 10 for Large Language Model Applications.” <https://owasp.org/www-project-top-10-for-large-language-model-applications/>
 10. OWASP GenAI Security Project, “OWASP Top 10 for Agentic Applications for 2026.” <https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/>
-

Published by AIMonger . <https://aimonger.com> . Malta . Europe