AIMONGER
WHITE PAPER

AI Discovery

Data Permissions Are the Critical Path for Enterprise AI

Production enterprise AI cannot retrieve useful organisational data unless identity, document permissions, and permission changes propagate into the retrieval path. For firms with fragmented systems, inherited archives, and cross-border teams, models are widely available but governed access is not automatic. Boards that want AI impact must treat permissions as architecture and decision rights, with dates and owners, not as a side ticket.

David Saliba . 2026-07-09 . aimonger.com/whitepapers/data-permissions-ai-critical-path/

The problem in one sentence

Your AI programme is probably waiting on a permission, not a model.

Public evidence does **not** prove that "most" AI pilots fail solely because of access. What it does prove is stronger for architecture: production retrieval must respect identity and document permissions, and permission changes must propagate or the system becomes either blocked or unsafe. Programmes that publish model bake-offs while ACL design is unfinished are not ready for consequential use.

ACL-aware retrieval, plainly: the AI system only returns documents the asking user could already open in the source system, not a copy of the whole archive. **In practice:** a contract discovery query returns matter-level results because SharePoint ACLs propagate into the retrieval index; it never surfaces files from matters the user cannot access.

In a CIO office, the practical blocker is often not model quality. It is read access to the document store, matter-level permissions in a contract vault, or a SharePoint estate where nobody wants to decide what the AI service principal may see.

What you see globally right now is every vendor promising enterprise AI on top of your knowledge. McKinsey's impact concentration among organisations that redesign workflows assumes those organisations can touch real data. Gartner's GenAI mainstreaming projection assumes applications can connect. Permissions are the hinge.

Permission backlog as a board artefact

If the backlog does not exist, the AI portfolio is fiction. Review it monthly on the steering agenda.

Workflow	System	Access type	Owner	SLA date	Status
Contract discovery	DMS	Read + provenance			
Invoice exceptions	ERP	Read PO + invoice			
Policy assistant	Intranet	Read current policies			

Governed fast path, in short: a pre-approved pattern for low-risk read-only access with logging, expiry, and a named approver, measured in days, not quarters. **Worked case:** the policy assistant goes live in two weeks because the request mapped to pattern #1, not a bespoke six-month security review.

Critical-path mathematics

If a pilot needs three access decisions and each averages 45 calendar days with serial dependencies, the programme is a quarter late before model work begins. Parallelise independent reads, use pre-approved patterns, and escalate at SLA breach; do not discover the path length in week twelve.

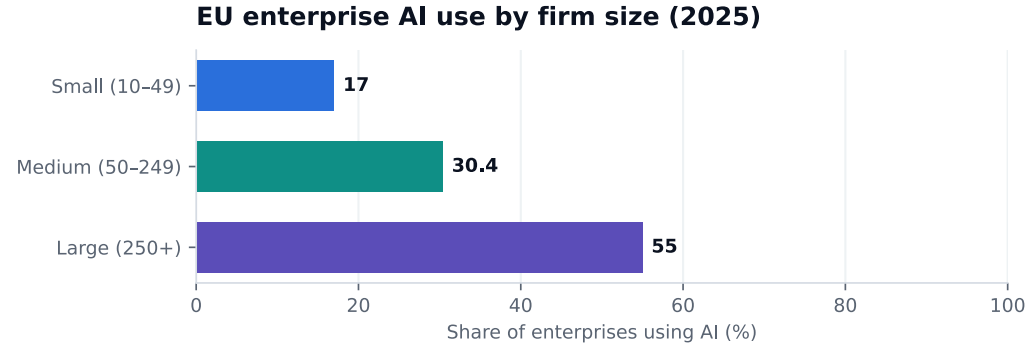

Figure 1. EU enterprise AI adoption by firm size, 2025. Plain read: larger firms adopt faster partly because data estates and access processes are more industrialised. **On Monday:** mid-market firms cannot assume large-enterprise access maturity. Source: Eurostat enterprise survey 2025 (official statistic). AIMonger redraw.

Eurostat reports 17.0 percent of small, 30.4 percent of medium, and 55.0 percent of large EU enterprises used AI in 2025. Scale advantages include data estates and access processes. Mid-market firms must industrialise permission patterns or lose tempo to peers who already have them.

Design patterns that unblock safely

- **Read-only replicas for AI workloads.** Retrieval runs against a read-only copy of the source corpus so builders cannot accidentally write back, mutate records, or broaden scope beyond the approved pattern during pilot and production phases. A read-only replica is the cheapest structural control against the most common pilot incident: a builder runs an exploratory query that updates a production record because the connection string pointed at the wrong database. Removing write access at the connection level, rather than trusting the builder to be careful, is how you keep a Monday-morning mistake from becoming a Friday-afternoon incident.
- **Row and document level security at retrieval.** Identity and ACL metadata filter every query at retrieval time so users see only documents they could already open in the system of record, not a flattened copy of the whole archive. Without this layer, a retrieval index becomes a privilege-escalation surface where a junior analyst can ask a question that returns partner-only material. The control belongs in the retrieval path, not in the application layer, because every consumer of the index inherits the same protection.
- **Time-boxed pilot access with auto-expiry.** Pilots receive grants measured in days or weeks with automatic revocation so forgotten permissions do not become permanent back doors after the steering group moves on. The pattern fixes a failure mode that audit committees know well: a vendor integration granted read access in 2023 that still holds a valid token in 2026 because nobody owned the offboarding. Auto-expiry turns a manual hygiene task into a structural guarantee.

- **Separate eval corpora from production corpora.** Builders experiment on anonymised, synthetic, or holdout sets that cannot write to production, which prevents evaluation prompts from becoming accidental data exfiltration paths. An eval corpus that shares a connection with production is a liability, because a test rig that prompts the model to summarise sensitive text can surface that text in a place no reviewer approved. Physical separation, not a config flag, is what makes the boundary survive a rushed release.
- **Logging of every service-principal access.** Each query by the AI service account is retained per retention policy so incident response and audit can reconstruct who retrieved what, when, and under which permission snapshot. Without service-principal logging, the postmortem for a wrong-document return starts with "we cannot tell who asked," which is exactly the sentence a regulator does not accept. The log is also the evidence that the governed fast path is actually governed, not just fast.

AWS documentation on knowledge bases characterises ACL-aware retrieval as filtering, not a complete authorization substitute. Google Cloud enterprise search patterns similarly require identity-aware grounding. Source permissions alone do not detect malicious content inserted through an authorised source. Layer controls: ordinary cybersecurity, AI-specific controls, sector controls (ENISA-aligned guidance).

Pattern library (examples)

1. **Read-only policy corpus.** Current policy documents only, synced weekly with full query logs and a default five-business-day SLA once the standard form is complete, so internal assistants go live without bespoke six-month security reviews. The pattern works because the blast radius is bounded: a policy corpus changes on a known cadence and contains no client-confidential material. A professional firm used this pattern to ship an internal policy assistant in eleven working days, where the previous attempt had spent four months in a security questionnaire that never closed.
2. **Contract vault slice.** Matter-based ACLs stay intact in retrieval, download tools remain disabled, and partners see only matters they already access in the DMS when asking precedent questions. The slice respects the existing permission model rather than flattening it, which means a partner who cannot open a matter in the source system cannot surface it through the AI either. Disabling bulk export at the API level is the control that separates a discovery tool from an exfiltration tool.
3. **ERP exception fields.** A narrow schema with masked PII where possible limits the AI service principal to PO and invoice exception fields instead of the full ERP estate, which shrinks review scope and incident blast radius. Granting the whole ERP because the workflow "might need it later" is the most common reason a security review stalls for months. A narrow schema forces the business owner to name the exact fields the workflow depends on, which is the conversation that should happen before access is requested, not after.
4. **Eval replica.** An anonymised or synthetic corpus gives builders a safe experimentation surface that is physically separated from production indices so promotion requires an explicit access decision, not a config toggle mistake. The replica is where prompt engineering, retrieval tuning, and red-team cases belong, because a mistake there cannot reach a real customer or ledger. Promoting a change from eval to production is an access decision with a named approver, which is what keeps the boundary honest under deadline pressure.

Each pattern carries a default SLA once the form is complete. Security reviews accelerate when requests map to a named pattern instead of inventing a novel estate every time.

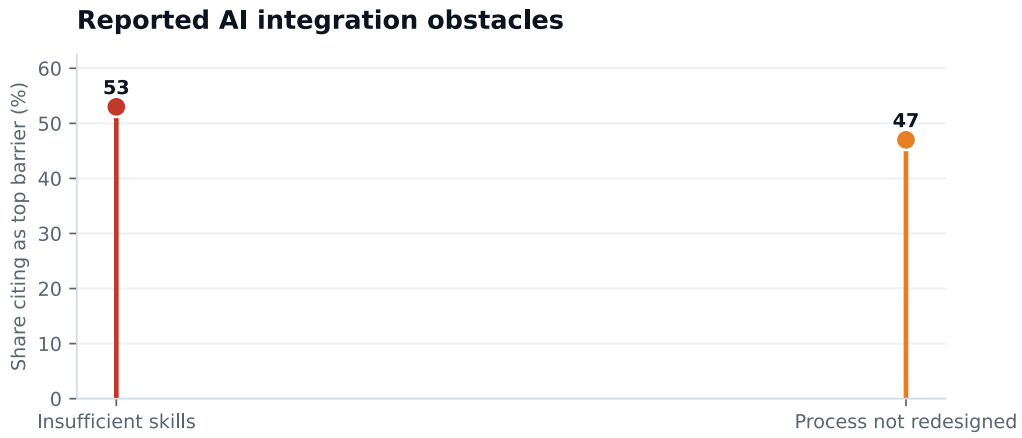

Figure 2. Skills and process redesign barriers among surveyed leaders (Deloitte survey shares). In short: access and skills sit inside integration, not beside it. **Operating case:** the critical path is often data plus process, not model selection. Source: Deloitte State of AI in the Enterprise 2026 (survey). AIMonger redraw.

Why security reviews stall

Common failure modes:

- **Over-broad requests.** Requests framed as "all SharePoint" or "entire DMS" force reviewers to size infinite risk, which defaults to months of delay or an unsafe yes. The reviewer cannot approve what they cannot bound, and the requestor cannot justify what they have not scoped. Narrowing the request to a named corpus with a declared data class is what turns an open-ended risk assessment into a decision a security team can make in days.
- **Missing data class.** When no data class is declared, security cannot judge sensitivity or residency requirements and must reject or escalate indefinitely. A data class is the single piece of metadata that lets a reviewer apply a pre-agreed rule instead of inventing an analysis from scratch. Requests without a class are the most common reason a fast-path pattern fails to apply, even when the underlying use case is low-risk.
- **No logging plan.** Reviewers cannot approve read access without knowing what queries will be retained, for how long, and who can inspect them during an incident. A logging plan is not bureaucracy; it is the evidence trail that makes a governed fast path defensible to audit. Without it, the reviewer is approving a black box, and the postmortem for any incident starts with "we cannot tell who asked."
- **No expiry date.** Permanent grants for pilot corpora accumulate silently until a former vendor integration still holds read rights to client matter archives. Expiry dates are the structural control that prevents permission sprawl from outliving the project that needed it. Auto-revocation at expiry is cheaper and more reliable than a manual offboarding checklist that nobody owns.
- **No business owner named.** Access tickets without an accountable business owner sit in IT queues because nobody can justify scope, kill criteria, or workflow priority. Naming the owner is what converts a ticket from an IT problem into a business decision with a metric and a kill criterion. Tickets without owners are also the ones that survive longest, because nobody has the authority to close them.
- **No pattern library to map against.** When every request invents a novel estate, reviewers fear irreversible risk and serial questionnaires replace the governed fast path measured in days. A pattern library gives the reviewer a pre-agreed answer to "is this shape of request safe," which is the

question that actually slows reviews. Mapping a new request to a named pattern is what makes a five-day SLA achievable rather than aspirational.

Fix with pre-approved patterns for read-only AI access to named corpora. NIST AI 600-1 recommends inventories capture data provenance, sensitive-data considerations, access modes, and human oversight roles. Access decisions are operating artefacts, not one-off tickets.

RACI for access decisions

Step	Responsible	Accountable
Request	Business owner	Business owner
Design	Technical owner	Technical owner
Approve	Security/data owner	CISO/data committee delegate
Implement	Platform	Technical owner
Review	Risk partner	Business owner

Case narrative (synthetic): the six-month demo

A company buys an AI discovery tool in January. By June it still runs on twenty sample PDFs. The blocker is not the vendor. It is a stalled request for read access to the document management system, waiting on a security questionnaire that asks questions the project team cannot answer because no data class was declared.

In July, leadership nearly cancels “AI” as a failure. The postmortem should say: permissions critical path unmanaged.

Prevention: permission backlog on the steering agenda from week one; pre-approved read-only pattern; five-day SLA once the form is complete.

Connecting permissions to shadow AI

When official access takes months, staff photograph documents into consumer tools. Permission delay is a shadow-AI manufacturing process. Boards that want less shadow AI must accelerate governed access, not only write bans.

the advantage is not broad access. It is a fast, narrow, repeatable access pattern that lets the business move without manufacturing shadow AI.

EU and sector context

Regulation (EU) 2024/1689 (AI Act) does not replace document-level permissions, but it raises the cost of getting them wrong. High-risk use cases require documented data governance, human oversight, and traceability. Even for lower-risk internal assistants, GDPR and sector rules (finance, health, gaming, professional secrecy) mean “copy the whole archive into a vector store” is not a viable architecture.

Data class means: a label that says how sensitive a document is and which tools may touch it. **On the P&L:** class 1 public marketing may use any approved assistant; class 4 client confidential may use only the governed contract vault pattern with logging.

Boards should ask the CISO to publish data classes before the AI steering group approves corpora. Without classes, security reviewers default to “no” because they cannot size the risk.

Integration checklist (technical owner)

Before production retrieval goes live, the technical owner should sign off:

1. **Documented ACL model.** Groups, matter-level rules, and row-level filters are written down and reviewed so engineers and auditors share the same picture of who may see what at query time. An ACL model that lives only in a senior engineer's head is a key-person risk and an audit failure waiting to happen. Documentation also forces the conversation about edge cases, such as inherited permissions and broken inheritance, before they surface as a wrong-document return.

2. **Tested permission sync.** Add, remove, and permission-change events are exercised on a schedule so stale indices do not serve documents a user lost access to yesterday. Permission sync is the most common silent failure in retrieval systems, because the index looks healthy while quietly diverging from the source of record. A scheduled test that revokes a test user's access and confirms the document disappears from their results is the cheapest way to catch drift before a regulator does.

3. **Minimum corpus scope.** The service principal is scoped to the smallest corpus that satisfies the workflow so a retrieval bug cannot expose unrelated matters or ledgers. Granting the whole estate because a narrower scope takes more effort to define is how an otherwise safe workflow becomes an enterprise-wide incident. Minimum scope is a security control that also accelerates review, because a smaller corpus is easier for a reviewer to say yes to.

4. **Query log retention.** Logs are retained per policy and exportable for incident review so "we cannot tell who asked" is never the postmortem conclusion. Retention is a dual-purpose control: it supports incident response and it proves the governed fast path is actually governed. Logs that exist but cannot be exported in an incident are functionally absent, because the person who needs them is rarely the person who configured retention.

5. **Eval isolation from production.** Evaluation indices cannot write back to production corpora or trigger side effects, which blocks the common mistake of test rigs becoming exfiltration paths. An eval rig that shares a connection with production is a liability, because a red-team prompt can surface sensitive text in a place no reviewer approved. Physical isolation, not a config flag, is what keeps the boundary honest under deadline pressure.

6. **Holdout at multiple permission levels.** Holdout queries run as users with different ACL profiles so wrong-document returns surface before external customers or regulators do. A holdout set that runs only as an administrator tests almost nothing, because an admin can see everything and therefore the test cannot detect a privilege leak. Running holdouts as junior, partner, and external-counsel profiles is what turns a green dashboard into actual evidence.

7. **Incident path for wrong-document return.** A named owner, severity class, and notification rule exists before go-live so the first leakage event triggers response instead of a steering-deck debate. The first wrong-document return is a question of when, not if, and a firm that has rehearsed the response will contain it in hours rather than weeks. Pre-assigning the severity class also prevents the common failure of under-rating an incident because nobody wanted to escalate.

This checklist is boring. It is also what separates a demo from something audit can defend.

Worked example: contract discovery in a professional firm

Workflow: "Have we dealt with this indemnity clause before?" **Systems:** Document management with matter-level ACL **Pattern:** Contract vault slice (#2) **Technical design:** Nightly sync of permitted matters; retrieval applies user identity at query time; no bulk export API **Human gate:** Partner review on any draft that cites a prior agreement **Metrics:** median days to permission decision; query success rate on holdout set

Access request form fields: workflow, data class, systems, fields/documents, read/write, users/service principals, retention, logging, eval vs prod, expiry, business justification, kill condition.

Metrics

- **Median days to permission decision.** Track how long priority workflows wait for access approval so the steering group sees permissions as tempo, not a hidden IT backlog. The metric converts an invisible delay into a number the board can escalate against, which is the only way access becomes a first-class steering item. A rising median is the earliest signal that the pattern library is breaking down or that reviewers are defaulting to bespoke questionnaires.
- **Access-related AI blockers.** Report what share of portfolio blockers are permission decisions so leadership stops blaming model quality when the critical path is ACL design. The metric reframes the conversation from "the AI is not good enough" to "the access decision is overdue," which is a problem an executive can actually unblock. Without this number, programme status decks understate the real constraint by attributing delay to model tuning.
- **Expired grants revoked on time.** Measure whether time-boxed pilots actually lose access on schedule, because orphan grants are a common source of audit findings and shadow exfiltration. The metric tests whether auto-expiry is real or whether someone has been silently extending grants to avoid a renewal ticket. On-time revocation is the control that keeps the permission estate from accumulating debt the way a codebase accumulates technical debt.
- **Access incidents.** Count wrong-document returns, over-broad queries, and sync failures so permission architecture improves from telemetry instead of anecdote alone. An incident count above zero is not necessarily bad; an incident count of zero on a system with no telemetry is the real warning sign. The metric gives the technical owner evidence to prioritise fixes that otherwise compete with feature work for engineering time.

Any AI-critical access request older than the published SLA appears on the next steering agenda with a named blocker. Silence is not an acceptable status.

Evidence base: access and skills dominate timelines more than model choice

Source	Finding	Critical-path reading
Deloitte 2026	Skills and process integration barriers dominate reported obstacles	Permissions and data readiness sit inside integration
NIST AI 600-1	Inventories should capture provenance, access modes, oversight	Access decisions are operating artefacts
McKinsey 2025	High performers redesign workflows and scale faster	Redesign assumes real systems of record are reachable
Eurostat 2025	Large enterprises adopt AI at much higher rates than small (55.0% vs 17.0%)	Mid-market must industrialise permission patterns
AWS / Google Cloud docs	ACL-aware retrieval is designed dependency	Filtering is not full authorisation substitute

Plain read: there is no universal "days to ACL" benchmark across sectors. Instrument your own median days-to-permission-decision and report it beside AI portfolio status.

Methodology note

Vendor documentation describes intended architecture, not your firm's observed SLA. Eurostat is representative EU enterprise survey, not ROI. Do not claim a universal percentage of pilots "stall on access" without firm-specific telemetry.

Counter-position: broad access accelerates innovation

Another board might argue that temporary broad access lets teams learn faster. Broad unmanaged access accelerates incidents. The design pattern is pre-approved, least-privilege, time-boxed read paths with logging, fast and narrow, not permanent open estates. Innovation belongs in eval replicas and bounded corpora, not in exfiltration risk.

CIO decision criteria

1. **Permission backlog with owners and SLAs.** Every priority workflow appears on a backlog with named owners and SLA dates so the board can escalate blockers instead of discovering quarter-long delays in week twelve. The backlog is the artefact that converts permissions from an IT ticket into a steering item with tempo. A backlog that does not exist is the clearest signal that the AI portfolio is fictional, because there is no record of what is waiting on whom.
2. **Published pattern catalogue.** At least four pre-approved read patterns are published so security reviews map requests to known corpora instead of reinventing estate-wide risk each time. The catalogue is what makes a five-day SLA achievable, because a reviewer can apply a pre-agreed answer to a recognisable shape of request. Without patterns, every request is a novel risk assessment and the fast path collapses back into a serial questionnaire.
3. **Monthly median days-to-decision.** The CIO reports median days-to-permission-decision monthly beside portfolio status so finance sees access tempo alongside model and integration spend. Reporting the number monthly, rather than annually, is what gives the steering group a leading indicator instead of a postmortem. A trend line also distinguishes a one-off slow decision from a structural bottleneck that needs executive escalation.
4. **Tested ACL propagation.** Permission add, remove, and change events are tested on a published cadence so retrieval indices cannot silently diverge from the system of record. Silent divergence is the most common cause of wrong-document returns, because the index reports healthy while serving stale permissions. A scheduled test that revokes a test user and confirms the document disappears is the cheapest control that catches drift before a regulator does.
5. **Eval corpus separated from production.** Builders work only on isolated eval replicas until an explicit promotion decision grants production read scope under a named pattern. Physical separation, not a config flag, is what keeps an experimental prompt from becoming an accidental exfiltration path. The promotion decision is an access event with an approver, which is what makes the boundary survive a rushed release.
6. **Executive escalation on SLA breach.** Any AI-critical access request older than the published SLA appears on the next steering agenda with a named blocker and executive escalation path. Escalation is the mechanism that prevents a single cautious reviewer from stalling a programme for a quarter without visibility. Publishing the SLA before the breach is what makes escalation enforceable rather than performative.
7. **Shadow-AI retirement link.** Permission acceleration connects to the shadow-AI retirement plan so governed fast paths win back the use cases staff currently run in consumer tools. The link reframes access speed as a shadow-AI control, not only an enablement task, because staff photograph documents into consumer tools when the official path takes months. A fast narrow pattern retires shadow use by being easier than the workaround, which a ban alone never achieves.

Appendix: access request form (fields)

Complete one form per access grant. Leave the Entry column blank until the request is filed.

Field	Entry
Workflow	
Data class	
Systems	
Fields or documents	
Read or write	
Users or service principals	
Retention	
Logging	
Eval versus production	
Expiry	
Business justification	
Kill condition	

Plain read: empty cells are intentional. An access grant without a kill condition and expiry is a permanent entitlement.

Closing position

Data permissions are a critical path for safe production AI.

Fund models if you want demos. Fund access architecture and decision rights if you want results that can survive audit.

References

1. 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
2. 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
3. McKinsey & Company / QuantumBlack, "The State of AI: Global Survey 2025." https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai
4. Eurostat, "20% of EU enterprises use AI technologies," 11 December 2025 (20.0% in 2025; 13.5% in 2024). https://ec.europa.eu/eurostat/en/web/products-eurostat-news/w/ddn-20251211-2
5. Eurostat, "Use of artificial intelligence in enterprises," Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises
6. 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
7. Gartner, "Gartner Says More Than 80% of Enterprises Will Have Used Generative AI APIs or Deployed Generative AI-Enabled Applications by 2026," press release, 11 October 2023. https://www.gartner.com/en/newsroom/press-releases/2023-10-11-gartner-says-more-than-80-percent-of-enterprises-will-have-used-generative-ai-apis-or-deployed-generative-ai-enabled-applications-by-2026
8. AWS documentation on metadata filtering / ACL-aware retrieval for knowledge bases. https://docs.aws.amazon.com/bedrock/latest/userguide/kb-advanced-parsing.html

9. Google Cloud documentation on ACL / identity-aware enterprise search and grounding patterns. https://cloud.google.com/generative-ai-app-builder/docs/acl

10. ENISA publications on AI cybersecurity. https://www.enisa.europa.eu/topics/artificial-intelligence

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