Inference-Time AI Data Control

Why Securing AI and Ensuring AI Performance are Inseparable

Executive Summary

Enterprise AI systems are now embedded in core business workflows, decision-making processes, and customer-facing operations. The quality, reliability, and risk profile of AI-generated outcomes have direct and material impact on business performance.

Yet enterprises lack a control plane capable of governing how enterprise data contributes to AI-generated outcomes in real time. Existing controls focus on models, users, infrastructure, or compliance artifacts. None directly regulate the data contribution layer that increasingly determines AI behavior. This gap manifests daily as hallucinations caused by missing context, bias introduced by duplicated data, unreliable results driven by conflicting information, and compliance failures that occur inside AI systems rather than at their boundaries.

AI Data Control defines a new category of platforms designed to address this gap. AI Data Control regulates which enterprise data fragments participate in AI-generated outputs at inference time by evaluating multiple control signals: policy, relevance, semantic meaning, business context, and usage patterns. By doing so, it enables enterprises to both promote intended AI outcomes and protect against unintended consequences.

The Core Problem: Static-Dynamic Mismatch

Enterprise AI architectures face a fundamental control gap: statically defined system prompts cannot effectively direct the use of dynamically changing data. A fixed instruction set (the system prompt) acts as a static map for a terrain (the data) that shifts daily.

Enterprise AI systems are increasingly built around retrieval-augmented generation (RAG), agentic AI and tool-using agents, and multi-source context assembly across documents, APIs, and databases. In these architectures, the same model can produce radically different outcomes depending on which data fragments are assembled into its context window. The primary determinant of AI behavior is not the model, but which data participates in generation, how often it appears, and how fragments relate semantically.

A solution valid for version 2.1 may break in version 3.0. Updates can make context that was appropriate for today's query irrelevant for tomorrow's—even when both queries use identical tokens. When an agent attempts to reconcile rigid instructions with fluid data contexts, open-loop architectures fail in predictable ways.

The Imperative for Closed-Loop Control

Bridging the gap between static instructions and dynamic data requires closed-loop control architecture. This moves the system from "fire-and-forget" to a governed, self-correcting utility:

- **Contextual Drift Mitigation:** Continuous feedback keeps agent behavior aligned with current business logic as data evolves
- **Real-Time Validation:** A governance layer validates relevance and policy adherence before data enters the context window
- **Iterative Correction:** The control loop detects deviations and triggers corrective adjustments, ensuring automatic recovery from errors

The Boundary Conditions for AI Data Control

AI Data Control is grounded in technical realities unique to AI systems that constrain how closed-loop control can be achieved. These boundary conditions are non-negotiable architectural requirements that any solution must satisfy.

Fragment-Level Data Use

AI models cannot process massive datasets or lengthy documents in their entirety due to context window limitations. Information is pre-processed: stripped of formatting and sliced into discrete, manageable chunks. While easier to describe as "sentences, tables, and charts," a data fragment can be any binary sequence.

This creates fundamental challenges: AI reasoning operates on a stitched-together mosaic of isolated fragments. Original relationships between fragments are lost during chunking. Business context, ownership, and use-policy become ambiguous at the fragment level. Moreover, from copy-paste, "oversharing," and versioning, over 95% of fragments exist duplicated across multiple documents and sources.

> *Controls designed for documents, files, or databases cannot govern AI systems that operate on fragments extracted from those containers.*

Security-Only Controls Create AI Failure Modes

Controls that simply remove or redact data without understanding semantic meaning or business context create informational gaps that increase hallucination likelihood and degrade output reliability. AI systems are probabilistic—when context is incomplete, they infer. Security-only controls often increase risk rather than reduce it.

When a security tool blocks data without understanding its relevance, the AI doesn't know data is missing. It generates responses based on incomplete information, confidently presenting answers that may be wrong. Research consistently shows that aggressive security filtering degrades AI answer quality. Enterprises face a false choice between security and accuracy when the real goal is achieving both.

Duplicate Data Bias

Duplicate or near-duplicate data fragments within an AI context window skew probabilistic weighting toward overrepresented viewpoints, reinforce unintended bias, and artificially increase confidence in incorrect conclusions. If the same information appears three times from three sources, the model treats it as three independent confirmations rather than one fact repeated.

This problem is invisible to document-level or classification-based controls. Traditional deduplication operates on files; AI bias operates on fragments. A document may be unique while containing paragraphs duplicated across dozens of other documents.

Conflicting and Junk Data

Authorized data can still be harmful if ambiguous, inconsistent, outdated, or low value. Such data increases inference cost without proportional benefit, pollutes context with noise, and produces incoherent or contradictory outputs. These are data contribution problems, not access problems—the data is legitimately accessible but shouldn't participate in the current inference.

Why Existing Control Models Fail

IT tools for data governance and access control evolved separately, despite their common directive to manage enterprise data. Individually they lack needed capability to satisfy the boundary conditions above, while connections between them are too weak for closed-loop control.

Control Category	Primary Function	AI Data Control Gap
Governance / GRC	Define policies, accountability structures	Design-time only; cannot enforce at inference
AI Firewalls	Block outputs at edge, filter traffic	Resource-based; no data understanding
DLP / Classification	Detect/block based on data type	Static, context-blind; creates gaps
MLOps	Model performance and stability	Model-centric; ignores data contribution
DSPM	Discover what data exists where	Discovery only; cannot control use

Governance and GRC

AI governance and GRC frameworks define principles, policies, and accountability structures. They are essential, but they are not operational control planes. Governance operates at design time and review cycles; GRC is retrospective and compliance-centric. Neither can shape AI behavior in real time, nor can they adapt enforcement based on observed relevance, meaning, or business context.

AI Firewalls and Access Control

Access control governs *who* may access *which resources,* based *on assumed relationships* between those resources and the data they contain. But AI systems are not simply conduits for data. They recombine data fragments across sources to generate new outputs and drive

decisions. The role of an agent may change with each user query. Controlling access to a source system does not control how data is used once it enters an AI context.

Data Loss Prevention and Classification

DLP and data classification are static tools using keywords and patterns to slot data into pre-defined types. They detect data and block, redact, or alert, based on category of data, while being blind to context of use. They lack precision, coloring all data in a document based on any "sensitive" values in it. The same sentence in separate documents can receive ambiguous treatment. These tools protect boundaries, but they cannot shape outcomes.

Defining AI Data Control

AI Data Control is an operational control plane that regulates how enterprise data contributes to user-directed, AI-generated outputs at inference time. Boundary conditions defined in the previous section preclude unilaterally "stopping the bad" without awareness of the consequences. **Promotion and protection are inseparable in AI systems**.

What AI Data Control Is

Unlike existing governance and security controls, AI Data Control is explicitly dual-purpose—it promotes desired outcomes while protecting against risk. It is data-centric rather than user- or model-centric. It operates at the level of data fragments and contextual contribution. It evaluates multiple control signals simultaneously. And it shapes outcomes rather than merely blocking inputs.

To meet these requirements, AI Data Control evaluates and balances several signals:

- **Policy:** Existing access, classification, regulatory, and contractual *requirements*
- **Relevance:** Whether a data fragment meaningfully contributes to the user's query and intended outcome
- **Semantic Meaning:** Understanding duplication, conflict, and coherence between fragments
- **Business Context:** How data relates to business processes, workflows, and requirements
- **Usage-Derived Signals:** Observed outcomes that reveal misalignment and gaps between policy requirements and implementation

What AI Data Control is Not

AI Data Control cannot be absorbed into existing categories without distortion. It is not governance that defines requirements but cannot enforce them in real time. It is not security that prevents leakage but cannot preserve meaning. It is not MLOps that improves models but doesn't regulate inference. It is not DSPM that discovers where data exists but cannot control its use.

> *AI Data Control operates below governance and above access controls, at the data-to-outcome layer that existing tools do not reach.*

Broader Implications

Data control for AI, and AI control of data are two sides of the same coin. By enabling precise identification, traceability, and direct data-centric control of data-in-motion and data-in-use, the capabilities brought by AI Data Control close critical gaps left by existing controls.

Incident Response

False positive alerts now exceed 95 percent because today's incident detection and response processes are fundamentally open-loop. Security tools generate indicators of compromise based on behavior and anomalies, but determining whether confidentiality, integrity, or availability has actually been compromised requires significant human effort to correlate logs and reconstruct data flows. Fragment-level tracing enables closed-loop incident response where the data involved in any operation is immediately identifiable.

App Performance

Data moves through ETL/ELT pipelines, BI tools, caches, search indexes, message queues, and SaaS integrations, many of which have only partial lineage visibility or incompatible metadata models. Optimizing performance requires manual mapping of data flow, coupled with iterations of trial-and-error, or as one customer described it, "100 people on a conference call." Fragment-level tracking enables precise performance optimization without manual reconstruction.

Storage

Data sprawl, long managed through coarse file- or table-level catalogs, becomes tractable through fragment-level cataloging and real-time tracing. Enterprises gain precise visibility into what data exists, where it resides, and how it propagates across AI workflows.

Conclusion

AI has fundamentally changed how data drives innovation and creates risk. Control models built for documents, systems, and users are no longer sufficient. The critical question is no longer who accessed what, but **which data influenced which outcome, and why**.

Caber is pioneering AI Data Control to answer that question. We're defining the new operational control plane that enables enterprises to promote intended AI outcomes while protecting against unintended consequences. As AI becomes inseparable from business operations, AI Data Control will become a foundational requirement for responsible, effective, and competitive use of AI.