

"Brandon Chiazza has written the operating manual."

AJAY AGRAWAL, co-author of Prediction Machines

OPERATING INTELLIGENCE

How Enterprises Build AI That Lasts



BRANDON CHIAZZA

SILENT WAY PRESS

SAMPLE CHAPTER

Operating Intelligence

How Enterprises Build AI That Lasts

Brandon Chiazza

This file contains Chapter 6, *Data Governance for AI*, from the first edition. It opens with the MyCity, MyAI case and develops the AI Data Trust Model and the controls that separate deploying AI from governing it. Figures and references are included as they appear in the book.

The full book is available in Kindle and paperback on Amazon. The Twelve Laws of Enterprise AI, the companion self-assessment, and the other working artifacts are at operatingintelligence.org.

Silent Way Press, 2026. Paperback ISBN 979-8-9963373-0-9. Kindle ISBN 979-8-9963373-1-6.

Copyright © 2026 Brandon Chiazza. All rights reserved. This sample may be shared unaltered.

Chapter 6: Data Governance for AI

When Governance Fails in Public

A small business owner asks an AI chatbot, on the New York City government’s website, whether her landlord can refuse her Section 8 housing voucher. The chatbot says yes. Source-of-income discrimination has been illegal in New York City since 2008. The bot did not know that, or knew it and generated otherwise; in either case the answer was wrong, the user could not tell, and no public rule specified when the system would be taken offline.

AI requires governance that traditional data programs were not built to provide. AI systems learn from data and then generate new text out of the patterns they found in it, drawing inferences across sources whose authority the user cannot see. The same capabilities that make AI useful create governance challenges that table-level access controls, query reviews, and lineage tools were not designed to handle: data used to train or prompt an AI system can resurface in its outputs, biases in training data can be perpetuated and amplified rather than merely stored, and provenance—traceable in principle at every hop of a conventional pipeline—can go opaque somewhere between retrieval and generation. The regulatory environment—the EU AI Act, emerging U.S. state regulations, and sector-specific requirements—is moving faster than most organizations’ governance design [1].

The AI Data Trust Model names five pillars: Data Provenance (where the data came from and what touched it on the way), Access Control (who may reach what, enforced at retrieval), Quality Assurance

(whether the data is still accurate when the system reads it), Usage Governance (which questions and uses the system is authorized to attempt), and Auditability (records sufficient for a regulator or an internal review to reconstruct what happened). The remainder of the chapter walks through the MyCity, MyAI case as a governance failure across four of them, develops each pillar, and lays out where enforcement must live in the AI data pipeline if governance is going to operate at runtime rather than only on paper.

Why Traditional Data Governance Is Insufficient

Traditional data governance was built for deterministic systems, in which a predictable data flow produces a traceable output and access can be controlled at the level of a system or a table. AI systems challenge each of those assumptions.

A traditional business intelligence report is generated from a defined query against a defined dataset, and the governance question is straightforward—who has access, and is the query correct? A large language model may draw on dozens of sources, combine them with patterns learned during training, and produce novel text that did not exist in any source document. The governance questions multiply. Which data influenced this output? Was any sensitive information exposed? Is the output consistent with organizational policies? Can it be reproduced for audit?

These questions cannot be answered by table-level controls and query reviews. They require governance that operates at the level of data pipelines and model behavior—the kind the MyCity case turned out to lack.

MyCity, MyAI: A Public-Sector Governance Case

On October 16, 2023, New York City launched a citywide AI chatbot at chat.nyc.gov [2]. The chronology and governance read that follow draw on contemporaneous reporting, the Comptroller’s audit, and my own

case treatment of the same events [10]. I had no role in developing or operating the chatbot. Mayor Eric Adams and Chief Technology Officer Matthew Fraser unveiled it alongside the city’s first Artificial Intelligence Action Plan—a thirty-seven-action package and the first such plan published by a major U.S. city. The chatbot was the second phase of the MyCity initiative and was positioned as the friendly front door to the city’s small-business resources. It was built in partnership with Microsoft on Azure AI services, underpinned by OpenAI. It drew on more than two thousand web pages published by the city’s Department of Small Business Services. The Office of Technology and Innovation (OTI) reported spending over \$600,000 on the chatbot’s foundations [3]. A small note on the launch page warned that the bot might occasionally produce “incorrect, harmful, or biased” content; in practice the caveat was easy to miss on a page that otherwise framed the system as authoritative.

Five months later, an investigation by *The Markup*, co-published with *THE CITY* and *Documented*, reported that the chatbot was confidently answering core legal questions in ways that would expose users to civil and criminal liability if followed [4]. The failures were not edge cases. Asked whether employers could take a portion of workers’ tips, the bot said yes—a violation of New York Labor Law § 196-d. Asked whether a business could refuse to accept cash, the bot indicated no regulation required cash acceptance—contradicting Local Law 34 of 2020. Asked whether landlords could refuse tenants paying through Section 8 vouchers, the bot indicated they could—even though source-of-income discrimination has been illegal in New York City since Local Law 10 of 2008. Other failures covered tenant lockouts, residential rent restrictions, funeral home pricing under the FTC Funeral Rule, food safety, and retaliation against workers who complain about sexual harassment. Some answers bordered on the absurd: a much-cited exchange in which the bot indicated an operator could serve cheese that had been bitten by a rat after “assess[ing] the extent of the damage caused by the rat.”

The administration declined to take the chatbot down. At an April 2, 2024 press conference, Mayor Adams defended continued operation —“Anyone that knows technology knows this is how it’s done. Only those who are fearful sit down and say, ‘Oh, it is not working the way

we want, now we have to run away from it all together”—and Chief Advisor Ingrid Lewis-Martin compared the situation to the early days of MapQuest [5]. Disclaimer language was rewritten more prominently. The bot’s scope was narrowed; it began declining to answer many questions and prompting users to ask only general government questions. Users were now required to actively accept the bot’s limitations before each session. The bot remained live. Julia Stoyanovich, director of NYU’s Center for Responsible AI, characterized the choice to keep the chatbot online as “reckless and irresponsible.”

The chatbot’s scope expanded in 2025 to include 311-service content, broadening the surface of questions the system was expected to handle even as the underlying accuracy concerns remained unresolved. On December 30, 2025, the New York City Comptroller released a sharply critical audit of MyCity [6]. The audit concluded the system “was poorly managed from both a project management and contract oversight perspective” and that the chatbot “appears to be unable to provide accurate or consistent information.” OTI argued in formal response that the chatbot had a 95–99% accuracy rate; the audit team noted OTI provided only a one-page summary of testing performed in 2023 and that this lacked sufficient detail to evaluate. OTI’s own user-feedback data, collected during July and August 2025 across more than 2,200 questions, showed 50 of 70 thumbs-rating respondents (71.4%) submitting negative feedback. On January 28, 2026, newly inaugurated Mayor Zohran Mamdani identified the chatbot—by then costing the administration “around half a million dollars” annually—as one example among “a number of different things we’re going to pursue for savings” [7]. The bot was taken offline on February 4, 2026. The chat.nyc.gov page was replaced with: “The Chatbot beta test has ended.”

Reading the case through the AI Data Trust Model

MyCity operated for roughly twenty-eight months under official city branding while answering legally consequential questions, and the case does not turn on any single technical flaw—there is no bug that, once fixed, would have fixed it. It turns on missing governance across four of the five pillars the rest of this chapter develops.

Provenance. The chatbot drew on the SBS web corpus but produced many answers without consistent source links, even though early descriptions told users they could verify responses through provided citations. When the source authority of an answer is not exposed to the user—and not enforced as a structural requirement of the system—the user cannot tell which claims are grounded in source material and which are model inference. Provenance must travel with the answer.

Quality assurance. There was no published accuracy threshold for launch, no published list of question categories the bot would not answer, and no published evaluation method against which user-facing claims could be reconciled. When OTI later asserted a 95–99% accuracy rate, the agency could not produce testing detail sufficient for audit evaluation; the Comptroller’s team noted only a one-page summary. The audit disputed the denominator behind the claim and recalculated August 2025 accuracy at 84.8–92.7%. The agency’s own user-feedback data measured something else again—respondent dissatisfaction, not answer accuracy—and pointed the same direction. A claimed accuracy rate that cannot be reconstructed under audit is not a governed quality control.

Usage governance. The deployed scope was unbounded by category. Users were not told before asking which domains the bot was reliable on or what categories of question fell outside its competence; they were warned only that occasional inaccuracy could occur. After failures became public, scope was narrowed reactively rather than designed in advance. Usage governance for a public-facing chatbot answering questions about regulated activity should define the categories of question the system is allowed to attempt and the categories it is required to decline—before launch, not after journalists report what it got wrong.

Access control. Access control is the one pillar the MyCity case does not turn on. The system was public by design, available to anyone with a browser, and it was never gated on user identity; the pillar is named here only so that the omission reads as deliberate rather than overlooked. A public interface does not retire the pillar—a public system can still expose restricted sources or privileged tools, and access

control governs those—but the documented record identifies no failure involving either. What its public design did was shift the burden onto usage governance: which categories of question the system is authorized to *attempt*. That is where MyCity’s failure concentrates.

Auditability. OTI’s user-feedback data and pre-launch testing came to light only through audit, twenty-one months after public reporting of harms. A system that produces tens of thousands of user-facing answers without an evaluable audit trail of testing, accuracy, and feedback cannot be governed in production. Auditability is not a reporting afterthought; it is the precondition for any later defense of the system’s behavior.

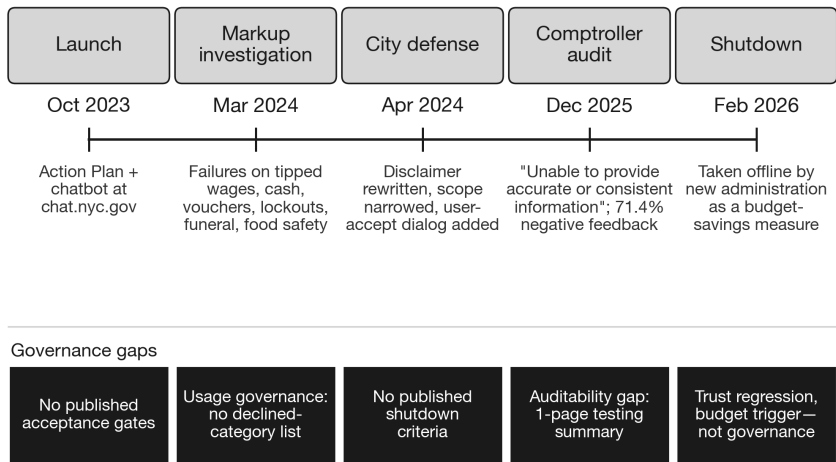
Operational trust (a sixth dimension the case foregrounds). The five pillars named so far are what the AI Data Trust Model formally requires. The MyCity case adds a sixth: the operating-decision discipline that decides when a probabilistic system in an institutional voice is allowed to keep speaking. The administration’s response treated disclaimer modifications, scope narrowing, and patches as substitutes for shutdown criteria. No public rule said under what conditions the bot would be taken offline, what evidence would trigger that decision, or who held the authority to make it. The eventual shutdown came from a budget review across administrations, not from the governance system the AI Action Plan had announced two years earlier. The system spoke in the official voice of the city while operating with the accuracy of a beta tool—the structural mismatch named by Law 11 of the Twelve Laws of Enterprise AI, *authority should match accuracy*. A confidently wrong answer delivered in an institutional voice spends down institutional credibility, and credibility does not regenerate at the rate capability does.

The narrower lesson is not that public-sector AI cannot work, or that disclaimer language is useless. It is that governance must be designed in before deployment, written down before incidents occur, and enforced through structural controls rather than through communications. The remainder of this chapter develops what that looks like.

There is a name for the pattern MyCity displays. The sociologist Diane Vaughan, in her study of the *Challenger* launch decision, called it the *normalization of deviance*: the slow drift by which what should

have been treated as a stop event becomes, through successive small accommodations, the new normal [11]. Each individual modification the city made—disclaimer rewritten, scope narrowed, user-accept dialog added, scope narrowed further—was defensible on its own terms. None of them, considered separately, would have changed the operating posture. Together they constituted a governance system in which the system’s failures had been accommodated rather than treated as signals to stop. The Comptroller’s audit, twenty-one months after the original *Markup* investigation, was the moment from outside that the institution could not provide for itself. Public-sector AI is unusual mainly in how visible this dynamic becomes; the same dynamic operates inside private enterprises with private branding and without the external accountability mechanism. The structural defense is not a better disclaimer. It is a pre-written stop rule, enforced by a named decision-rights owner, against which incremental accommodations must be tested before they become the operating norm.

FIGURE 6.1 MyCity Timeline and Governance Gap Map



The chatbot operated for ~28 months under official city branding. Shutdown came from a budget review across administrations, not from a governance trigger written before launch.

Figure 6.1—MyCity Timeline and Governance Gap Map. Five events span twenty-eight months: Launch (Oct 2023), Markup investigation (Mar 2024), city defense and disclaimer modifications (Apr 2024), Comptroller audit (Dec 2025), shutdown (Feb 2026). Each event is paired with the pillar gap it exposed. The eventual shutdown came from a budget review, not from a governance trigger.

The AI Data Trust Model

The AI Data Trust Model is the structured response to everything the MyCity case lacked: five pillars, each addressing a distinct dimension of AI governance risk, with gaps in four of those dimensions visible in the MyCity case.

FIGURE 6.2 AI Data Trust Model

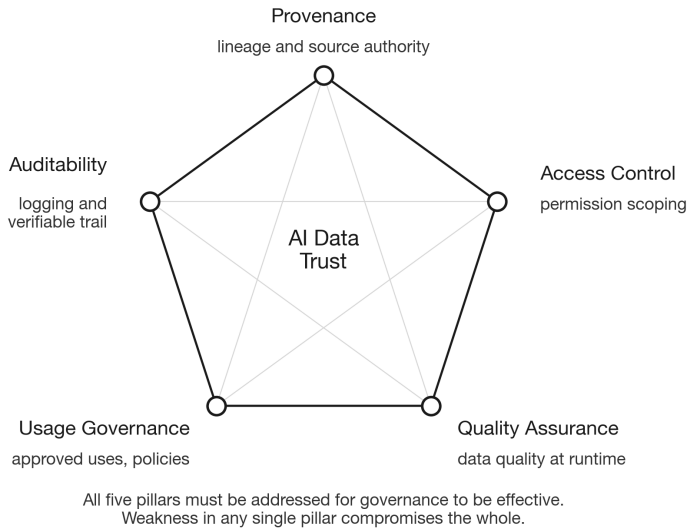


Figure 6.2—AI Data Trust Model. The five pillars—Data Provenance, Access Control, Quality Assurance, Usage Governance, and Auditability—arranged as interdependent vertices around a central trust core. The pentagon form is intentional: weakness in any single pillar compromises the whole. MyCity failed at four of the vertices at once—access control being the one pillar on which the documented record shows no failure.

The first pillar, Data Provenance, establishes the origin and journey of every data element used by AI systems. This includes documenting the sources from which data was collected, the transformations applied during processing, the models that consumed the data, and the outputs those models generated. Provenance tracking enables organizations to answer critical questions: “What data influenced this AI decision?” and “If a data quality issue is discovered, which AI outputs are affected?” The technical literature on quantitative input influence—Datta, Sen, and Zick’s framework for attributing a model’s output to specific inputs—provides one of the formal foundations for this kind of accountabil-

ity, although the production engineering required to make it operate at enterprise scale is non-trivial [9].

The second pillar, Access Control, ensures that AI systems can access only the data they are authorized to use and that the outputs they generate respect the access restrictions of the underlying data. This is more complex than traditional access control because AI systems may combine data from multiple sources with different access levels. An AI-generated summary that combines information from a public document and a confidential document inherits the access restrictions of the most restrictive source.

The third pillar, Quality Assurance, covers three checkpoints. At intake, records failing schema, freshness, or completeness checks are quarantined instead of passing through with a warning nobody reads. In the pipeline, watch for drift: the share of documents failing extraction, the age of the oldest record still being served, the fraction of retrievals returning superseded versions. At the output, the checks that matter are the ones tied to a decision—whether the answer cites a source that is still authoritative, and whether it falls within the range the workflow expects. Each check needs a named owner and a defined action on failure; a check that only logs still needs someone to act on it.

The fourth pillar, Usage Governance, controls how data is used by AI systems. This includes policies on which data can be used for model training, which data can be included in retrieval contexts, whether data can be sent to external AI APIs, and how generated outputs that contain derived data are handled. Usage governance is particularly important for organizations subject to sector-specific regulations on data use, such as HIPAA in healthcare or GLBA in financial services.

The fifth pillar, Auditability, maintains records sufficient for regulatory review, internal audit, and incident investigation. For AI systems, this means logging the inputs provided to models, the retrieval context assembled for each query, the model version and configuration used, and the outputs generated. The challenge is logging enough to investigate without paying the performance, cost, and privacy bills of logging everything.

Governance Operating Model

The natural first artifact of a governance program is a policy document, and the natural mistake is to treat the document as the program. A policy document records what an organization intends. Governance that runs is a federated structure with three tiers, and each tier holds work the other two cannot do.

FIGURE 6.3 Governance Operating Model

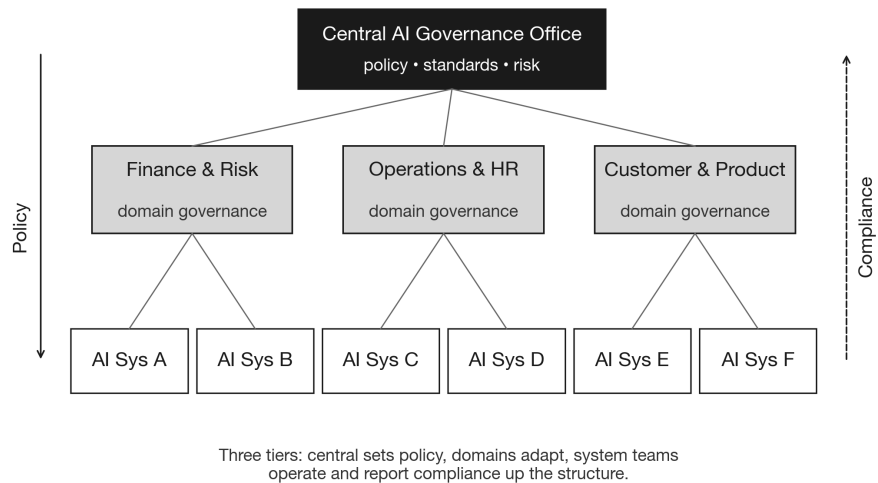


Figure 6.3—Governance Operating Model. Three tiers. A central AI governance office holds enterprise policy, the regulator interface, and the governance technology platform. Domain governance teams hold sector-specific risk and the data inventory. AI system teams hold control implementation, audit-log generation, and evaluation evidence. Policy flows down; compliance evidence flows up. The signature failure is the inverse—policy that never reaches the system team and evidence that never reaches the central office.

The central AI governance office owns enterprise policy and the regulator-facing posture. It writes the standards the rest of the organiz-

ation implements, classifies AI use cases by risk tier, owns the governance technology platform, runs the enterprise-wide risk register, and is the named counterparty when a regulator calls.

The domain governance teams translate enterprise policy into operating reality for their business unit. They maintain the data inventory the policy applies to, run use-case-level risk assessments, and absorb the sector-specific regulatory exposure (HIPAA in clinical data, GLBA in financial services, FERPA in education) that the central office cannot speak to in detail.

The AI system teams implement governance at the system level—the controls that show up in code, the audit logs that feed the governance dashboard, and the evaluation runs the governance review consumes. Governance that does not reach the system team is governance that does not reach production.

What that looks like in practice is narrower and more specific than a policy document suggests. In a county contract-management design, we had to turn several governance conditions into behavior the application could enforce, and three of them show how far apart similar-sounding rules turn out to be. An inactive or suspended supplier was a blocking condition. An active, unresolved bid protest prevented one particular transition—negotiation to board approval—and no other. An insurance mismatch was flagged and did not block anything.

Each required different code. The protest rule was the instructive one: the presence of a protest record was not the condition. Its *active and unresolved* status was, and so was the specific transition being attempted. A control built without those two qualifications either stops work it should permit or permits work it should stop, and both failures look like the control is working.

An agent operating in that workflow inherits all of it. It might explain why a contract cannot advance, find the protest record, or draft material for whoever must resolve it. What it cannot do is talk its way past the transition, and the system must be built so that a well-reasoned explanation of why proceeding seems sensible carries no authority whatsoever. The insurance case is the subtler one, because the agent must report the discrepancy without silently converting an advisory flag into a hard stop, and the workflow continuing must not

be recorded as the discrepancy being resolved. That gives a governance team something to inspect rather than assert: does the system block the supplier condition, does the protest prevent that specific transition, does the insurance flag stay visible while work proceeds? Three questions with executable answers.

Policy flows downward through this structure while compliance evidence flows upward, and the signature failure is the inverse of both: policy that never reaches the system team, and compliance evidence that never reaches the central office. A program in that state looks compliant from above and is indefensible under audit.

Shutdown and Escalation Criteria

The MyCity case illustrates why shutdown criteria must be explicit and enforceable. Without such criteria, shutdown can become a political or budgetary decision made under pressure. The requirement is stated in Law 10 of the Twelve Laws of Enterprise AI—*every AI system needs a written rule under which it will be taken offline*. A system without one has been launched but not governed.

Every AI system operating in production should have written shutdown and escalation criteria covering five areas. NIST’s AI Risk Management Framework organizes risk management around four functions—govern, map, measure, and manage—with governance applied across the other three [8]. It remains the clearest public articulation I know of the discipline a system without written criteria is missing.

Pre-launch acceptance gates. What accuracy, refusal-rate, citation-coverage, and safety thresholds must the system clear before public release? Who tested against these thresholds, on what evidence base, and where is the record kept? OTI’s later inability to produce evaluable pre-launch testing—only a one-page summary—is what turned the 95–99% accuracy assertion into an audit finding.

Incident classification. What categories of failure exist (factually wrong answer, restricted-data exposure, unsafe recommendation, scope violation, cost anomaly, public-trust event), and what does each

category trigger in terms of internal notification, user notification, scope narrowing, pause, or shutdown?

Decision rights. Who can pause retrieval, narrow scope, disable a tool, take the system offline, or restore service? These decisions should not require an off-topic press conference. They should be exercised by named owners against pre-defined criteria.

Review cadence. What recurring evaluation, user-feedback review, audit, and red-team activity is scheduled, who is accountable for reading the results, and how do the results travel back to system owners and executive sponsors?

Escalation triggers. What threshold of failed evaluations, negative user feedback, regulator interest, or external reporting moves the system from internal monitoring to executive escalation? These triggers should be quantitative where possible: a refusal-rate spike, an evaluation-drift threshold, a cost anomaly, a negative-feedback ratio.

A system that carries all five in writing can be taken offline as an ordinary operating decision. Without them, the decision can be driven by pressures unrelated to system performance, as MyCity’s eventual shutdown through a budget review illustrates.

Artifact: AI Production Shutdown Decision Record

Field	Required content
System scope	Use case, user population, allowed and prohibited question categories
Acceptance gates	Accuracy threshold, refusal-rate target, citation-coverage target, safety thresholds, location of testing evidence
Incident classes	Defined categories triggering notification, user warning, scope narrowing, pause, or shutdown
Decision rights	Named owners who can pause, narrow scope, disable tools, take the system offline, restore service
Review cadence	Evaluation, user-feedback review, audit, red-team schedule and accountable owner

Field	Required content
Escalation triggers	Quantitative thresholds (negative-feedback rate, evaluation drift, cost anomaly, external reporting) moving from monitoring to executive escalation
External-report response	Process for triaging journalistic, regulator, or audit reporting; published response timeline

Enforcement Points Across the AI Data Pipeline

Governance fails when policy is captured in documents that the runtime does not read. Effective governance places enforcement points along the AI data pipeline so that policy becomes runtime constraint rather than aspiration. Six enforcement points are common across enterprise AI systems.

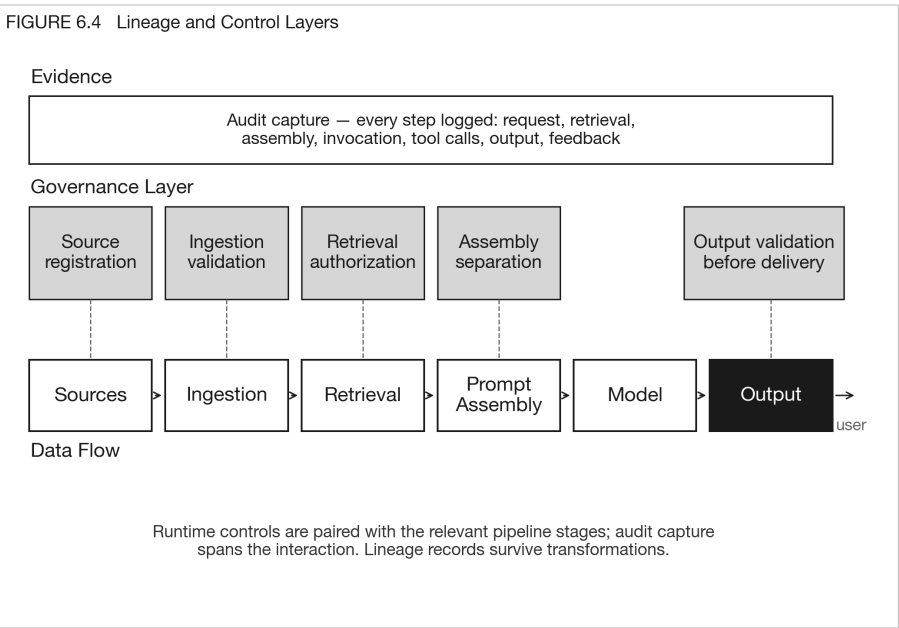


Figure 6.4—Lineage and Control Layers. Data flows from sources through ingestion, retrieval, and prompt assembly to the model and its output. The

governance layer places runtime controls along the pipeline—source registration, ingestion validation, retrieval authorization, assembly separation, and output validation before delivery—and an audit band spans the whole interaction. A program that places policy only at the top of the pipeline leaves the runtime unprotected.

Source registration. Authoritative sources are registered, classified, and version-controlled. Unregistered sources do not enter the retrieval index. New sources require ownership and review before activation. Stale or superseded sources are removed or quarantined automatically.

Ingestion validation. Documents are checked for sensitivity classification, license restrictions, hidden instructions, and quality signals before they are indexed. Anything that fails ingestion validation is logged and routed for review rather than silently accepted.

Retrieval authorization. Each retrieval call carries the requesting user’s identity, role, entitlements, and purpose. The retrieval engine filters before returning context. Documents are not exposed to the model if the user is not authorized to see them.

Prompt assembly. System instructions, retrieved context, and user input are concatenated with explicit separation. Trusted instructions are marked. Untrusted content (retrieved documents, tool outputs) is treated as data, not as instructions. Response structure is enforced through output schemas rather than request-level prose; whether a citation supports its claim, and whether a refusal was warranted, are checked at the validation and evaluation steps that follow.

Output validation. Model output passes through factuality checks where feasible, citation checks, sensitive-information filters, policy compliance checks, and structured-output validation before it reaches the user or downstream system. Disclaimers belong here, but disclaimers are communication, not enforcement. This is Law 6—a *disclaimer is not a control*—and Chapter 14 builds its security architecture on the distinction.

Logging and audit capture. Each step—request, retrieval, prompt assembly, model invocation, tool calls, output, feedback—generates structured records sufficient to reconstruct the interaction for evaluation, incident investigation, and audit. The Comptroller reported receiving only a one-page testing summary, limiting its ability to evaluate the system’s quality controls.

Figure 6.4 shows these enforcement points as a paired layer above the data flow. A governance program that places policy only at the top of the pipeline—for example, a model-card review at procurement—leaves the runtime unprotected.

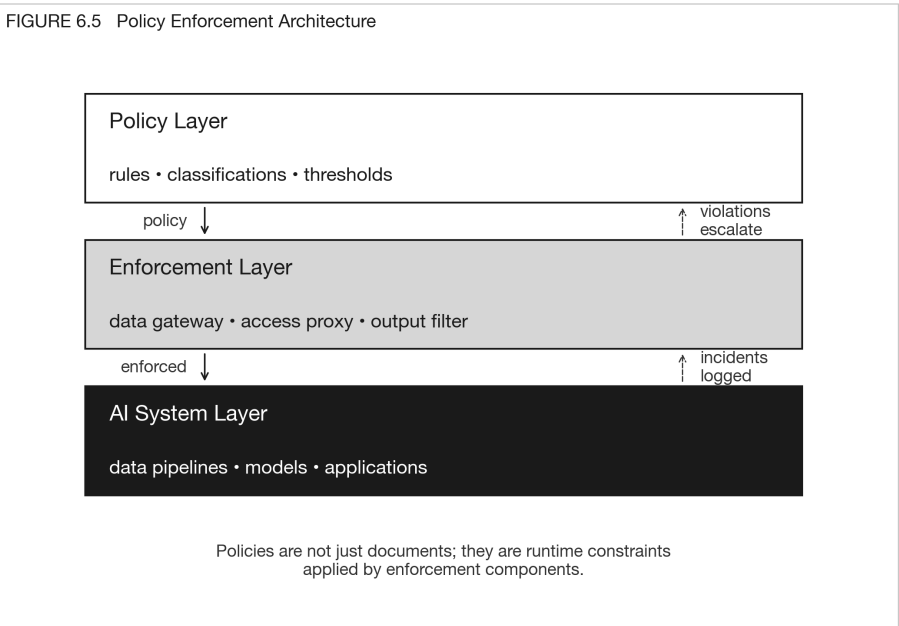


Figure 6.5—Policy Enforcement Architecture. Three stacked layers: Policy (rules, classifications, thresholds), Enforcement (data gateway, access proxy, output filter), and AI System. Policy flows downward to constrain behavior; violations flow upward through feedback paths that feed the next policy update.

References

[1] European Parliament, “Artificial Intelligence Act,” Regulation (EU) 2024/1689, 2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

[2] Office of the New York City Mayor, “Mayor Adams releases first-of-its-kind plan for responsible artificial intelligence use in NYC government” [Press release], October 16, 2023. <https://www.nyc.gov/mayors-office/news/2023/10/mayor-adams-releases-first-of-its-kind-plan-responsible-artificial-intelligence-use-nyc>

[3] S. Shibu, “New York City’s AI chatbot keeps getting facts wrong, 6 months and \$600,000 after launch,” *Entrepreneur*, April 5, 2024. <https://www.entrepreneur.com/business-news/nycs-first-ai-chatbot-keeps-getting-important-things-wrong/472280>

[4] C. Lecher, “NYC’s AI chatbot tells businesses to break the law,” *The Markup*, co-published with *THE CITY* and *Documented*, March 29, 2024. <https://themarkup.org/artificial-intelligence/2024/03/29/nycs-ai-chatbot-tells-businesses-to-break-the-law>

[5] C. Lecher, K. Honan, and M. Puertas, “Malfunctioning NYC AI chatbot still active despite widespread evidence it’s encouraging illegal behavior,” *The Markup*, April 2, 2024. <https://themarkup.org/artificial-intelligence/2024/04/02/malfunctioning-nyc-ai-chatbot-still-active-despite-widespread-evidence-its-encouraging-illegal-behavior>

[6] Office of the New York City Comptroller, *Audit Report on the New York City Office of Technology and Innovation’s MyCity System*, December 30, 2025. <https://comptroller.nyc.gov/reports/audit-report-on-the-new-york-city-office-of-technology-and-innovations-mycity-system/>

[7] C. Lecher and K. Honan, “Mamdani to kill the NYC AI chatbot we caught telling businesses to break the law,” *The Markup* and *THE CITY*, January 30, 2026; updated February 4, 2026. <https://themarkup.org/artificial-intelligence/2026/01/30/mamdani-to-kill-the-nyc-ai-chatbot-we-caught-telling-businesses-to-break-the-law>

[8] NIST, “AI Risk Management Framework (AI RMF 1.0),” NIST AI 100-1, January 2023. <https://doi.org/10.6028/NIST.AI.100-1>

[9] A. Datta, S. Sen, and Y. Zick, “Algorithmic Transparency via Quantitative Input Influence,” *Proc. IEEE S&P*, 2016. <https://doi.org/10.1109/SP.2016.42>

[10] B. Chiazza, “MyCity, MyAI: New York City’s AI Chatbot Experiment, 2023–2026,” PUBPOL 5019 Case Study, Cornell Jeb E. Brooks School of Public Policy, 2026.

[11] D. Vaughan, *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. University of Chicago Press, 1996. Vaughan’s concept of the *normalization of deviance*—the gradual organizational process by which signals of risk are reinterpreted as acceptable variation—is the most direct social-science analogue to the MyCity disclaimer-modification pattern.

PRAISE FOR OPERATING INTELLIGENCE

“Generative AI models are undoubtedly amazing, but this refreshingly honest book puts them in their place. Unless your organization redesigns work around AI capabilities, captures organizational context, upskills your people, and puts the AI system into full production, you won't get economic value. It's all hard work, but the sooner you get started the sooner you'll succeed with AI.”

THOMAS H. DAVENPORT

Distinguished Professor, Babson College; Fellow, MIT Initiative on the Digital Economy and Stanford Institute for Human-Centered AI

“In *Power and Prediction*, we argued that the real value of AI emerges not just from better point solutions, but from redesigning systems that properly utilize machine prediction and human judgment. *Operating Intelligence* is an unusually practical guide to that system-level work, showing executives how to move beyond demos toward governed context, proof of value, and operating-model change that creates and compounds value. For leaders seeking to turn clever neural nets into durable enterprise advantage, Brandon Chiazza has written the operating manual.”

AJAY AGRAWAL

Professor, University of Toronto, and Founder, Intrepid Growth Partners; co-author of *Prediction Machines* and *Power and Prediction*

“Cheap prediction leads to the question: which decisions do you redesign? *Operating Intelligence* provides answers based on Chiazza's real-world experience in AI strategy.”

AVI GOLDFARB

Rotman School of Management



ABOUT THE AUTHOR

Brandon Chiazza is CEO of Modali Consulting and teaches at Cornell's Brooks School of Public Policy and Yeshiva's Katz College of Science and Health. Formerly CTO of the NYC Mayor's Office of Contract Services, he is a co-inventor on multiple U.S. patents covering AI-assisted workflows.