Shadow AI and the Monitoring Paradox: A Typological Theory of Governance Visibility Erosion

Completed Research Paper

Nora Alhelal
University of Oxford
Oxford UK
Shaqra University
Shaqra Saudi Arabia
nora.alhelal@cs.ox.ac.uk

Sadie Creese
University of Oxford
Oxford UK
sadie.creese@cs.ox.ac.uk

Michael Goldsmith
University of Oxford
Oxford UK
michael.goldsmith@cs.ox.ac.uk

Abstract

Organizations increasingly face AI capabilities that evade effective oversight even when they are embedded in systems that are formally approved and actively monitored. Existing accounts of shadow IT and shadow AI do not fully explain this pattern because they assume that once relevant activity is discovered, governability can be restored. This paper develops a typological theory of shadow AI by reconceptualizing it as a governance condition rather than an authorization status. The theory identifies two generative mechanisms, visibility erosion and authorization decay, and shows how they degrade three governance capabilities: observation, attribution, and auditability. It yields four shadow AI types: Approved but Ungoverned, Unauthorized Tools, Undisclosed Vendor AI, and Ambient AI, and introduces the Monitoring Paradox, a boundary condition under which increased monitoring intensity yields diminishing returns because the constraint lies in signal availability rather than monitoring effort. The paper reframes governance from monitoring investment to delivery architecture design.

Keywords: Shadow AI, AI governance, Monitoring Paradox, auditability, cybersecurity

Introduction

Organizations are rapidly scaling artificial intelligence: 65% now include it in corporate strategy, and leading performers allocate over 10% of technology budgets to AI initiatives while continuing to increase spending (Horton et al., 2025). Yet AI can escape effective organizational oversight even when systems are formally approved and actively monitored. IBM (2026) reports that 92% of organizations experiencing an AI-related breach lacked proper AI access controls, while 68% of breached organizations lacked AI governance to manage AI or detect shadow AI. The problem, then, is not only unauthorized adoption, which the shadow IT literature already documents, but a deeper representational failure: the mechanisms intended to monitor AI operation can cease to represent what those systems actually do at runtime.

Existing theoretical explanations do not account for this pattern because they share a restoration premise: once relevant activity is surfaced, governability can be re-established. In the shadow IT literature, discovery is treated as a precondition for governability because previously unknown artifacts can then be evaluated, integrated, monitored, or controlled (Magunduni & Chigona, 2018; Raković et al., 2020). Emerging shadow AI work largely preserves this logic by defining shadow AI as employee use of AI tools for work without organizational approval, implying that once such uses are surfaced, they can be brought under existing governance mechanisms (Puthal et al., 2025; Silic et al., 2025). Control and agency theories frame oversight similarly, as a problem of obtaining interpretable signals about behavior or outcomes and reducing information asymmetry through additional monitoring investment (Jensen & Meckling, 1976; Ouchi, 1979). Across these literatures, the shared premise is that observability can in principle be restored, and that restoring it is sufficient to re-establish governance.

Research on loosely coupled systems comes closest to this problem by foregrounding the maintenance of linkages between operational practice and oversight representations (Weick, 1976). Even this literature, however, does not fully theorize what happens when delivery architectures themselves, rather than user concealment alone, render AI capabilities structurally invisible. Many AI capabilities are not concealed by users so much as made ungovernable by the pathways through which they are delivered. Discovery and authorization alone therefore cannot restore governability without architectural and procedural changes to those delivery pathways and to the governance triggers that depend on them.

To address this gap, we ask: How do AI delivery architectures generate and sustain distinct configurations of shadow AI as a governance condition, and under what conditions can monitoring restore governability? We develop a system-level typological theory that reconceptualizes shadow AI not only as an authorization status, that is, unauthorized use, but as a governance condition in which organizational oversight cannot reliably track actual AI operation. AI delivery architectures activate interpretive, technical, behavioral, and procedural barriers that generate visibility erosion and authorization decay. Visibility erosion and authorization decay degrade three governance capabilities: observation, attribution, and auditability. The resulting capability configurations enable differential diagnosis of how governability breaks down and yield a four-type typology of shadow AI. Where delivery pathways degrade oversight signals, their redesign takes priority over increased monitoring for restoring governability. This shifts the framing of shadow AI from an authorization problem to a sociotechnical security and privacy problem: breach prevention, incident attribution, and compliance verification all encounter the same structural constraint when oversight signals are unavailable at the architectural level.

Building on these relationships, the theory identifies a boundary condition we term the Monitoring Paradox: under visibility erosion, additional monitoring intensity yields diminishing returns because the constraint lies in signal availability rather than in the amount of monitoring applied.

Theoretical Background

The restoration premise inherited from shadow systems and governance literatures provides the point of departure for the argument developed here. The synthesis draws on sources selected for their role in defining and bounding shadow AI, establishing inherited assumptions about governability, or grounding the constructs, mechanisms, and boundary conditions developed below. The review traces how shadow IT and emerging shadow AI research establish the focal phenomenon and its restoration premise; how control and agency theories, together with loose coupling, specify the monitoring and linkage assumptions on which that premise depends; and how research on AI delivery and governance grounds the architectural conditions under which those assumptions break down. Typological theory then provides the basis for organizing these constructs and relationships into ideal types and propositions.

Shadow Systems and Governance

Shadow IT refers to software or hardware present within an organization that lies outside typical review processes of the responsible IT unit and is generally not known, supported, or accepted by the official IT department (Orr et al., 2024; Raković et al., 2020; Zimmermann et al., 2017). The literature associates shadow IT with risks for information systems security and challenges to legal or contractual compliance, and with operational difficulties including undocumented data sources, inconsistent data, and potential data loss (Haag & Eckhardt, 2017). This stream refined several governance assumptions by documenting

that shadow IT includes both deliberate policy circumvention and work-enhancement uses where employees may not intend policy violations, suggesting that adoption motivations are not uniform (Silic & Back, 2014). Governance responses oriented toward identifying risks and implementing controls were developed (Magunduni & Chigona, 2018), while later work stressed that some shadow solutions support user-driven innovation, prompting governance approaches that selectively integrate rather than eliminate certain instances (Fürstenau et al., 2021; Klotz et al., 2019).

This core assumption remained largely intact even as the literature evolved: discovery ensures governability. Observation was treated as effort-dependent, on the grounds that what is monitored becomes visible. Investment in inventories, logging, and scanning was expected to identify unsupported IT solutions (Raković et al., 2020; Zimmermann et al., 2017). Because the relevant artifacts were generally inspectable once identified, they were assumed governable through management actions including coordination, integration, or monitoring (Magunduni & Chigona, 2018). This view implies that the barrier to governance is primarily one of intent or awareness: systems remain in the shadow because users conceal them or IT units fail to detect them. Once the veil is lifted, the artifact itself is readable and governable.

When this vocabulary extends to AI, emerging work defines shadow AI as employee adoption or usage of AI tools without explicit organizational approval (Chin et al., 2025; Puthal et al., 2025; Silic et al., 2025). Related guidance explicitly frames shadow AI as echoing shadow IT (Campbell & Jovanovic, 2024). These accounts foreground AI use discovery and technical monitoring, for example via Cloud Access Security Broker (CASB)-based traffic analysis, as primary governance levers, echoing earlier assumptions that making unauthorized usage visible suffices to render it governable (Campbell & Jovanovic, 2024; Puthal et al., 2025). Some accounts also identify a governance drift zone in which formal policies exist but fail to align with emergent AI tool use (Silic et al., 2025). Recent work theorizes behavioral and organizational mechanisms through which informal AI use produces disclosure and privacy risk (Sebastian, 2026), advancing the behavioral account while retaining employee use as the locus of shadow AI. This framing preserves the discovery-to-governability premise by treating all governance failures as variants of unauthorized adoption and assumes that discovery and subsequent authorization restore governability.

Practitioner-oriented governance and service-management approaches address AI at different levels. ITIL-oriented guidance focuses on assessing and adapting existing governance arrangements to AI-specific risks, whereas COBIT guidance organizes AI oversight across objectives for strategic alignment, risk, development and change, service operation, and performance monitoring (Finister et al., 2025; ISACA, 2025). ITIL-oriented guidance also defines shadow AI as independent use of AI capabilities without involvement of the IT organization (Finister et al., 2025), reproducing the authorization framing. ISO/IEC 20000-1 provides a complementary, technology-neutral service-management foundation through requirements for scope definition, service catalogs, configuration information, supplier relationships, and change and release management (International Organization for Standardization, 2018). Under shadow AI conditions, however, applying these governance and service-management approaches remains contingent on the relevant capability becoming identifiable and observable as an object of governance.

This discovery-to-governability premise fails to account for how AI-enabled functionality can drift outside effective oversight even when systems are formally approved and nominally monitored, as vendor-controlled updates modify capability behavior faster than governance routines trigger reassessment. Shadow AI therefore introduces a problem of structural invisibility rather than mere concealment, motivating the theoretical departure developed here.

Control, Coupling, and Agency Lenses

To pinpoint why AI destabilizes the assumption that discovery ensures governability, we examine three influential IS governance lenses. Control theory treats the feasibility of formal evaluation and control as contingent on whether behavior or output relevant to desired performance can be measured with sufficient precision (Ouchi, 1979). Behavior control is feasible when the transformation process can be specified; outcome control when outputs are unambiguously measurable. IS governance research extends this contingency logic by showing that project stakeholders configure portfolios of control modes combining formal and informal controls (Kirsch, 1997).

Principal-agent theory frames governance as a delegation problem under information asymmetry and interest divergence, treating monitoring as the mechanism through which principals limit aberrant agent

behavior, with agency costs including monitoring expenditures and residual loss (Jensen & Meckling, 1976). Positivist agency theory conceptualizes information as a purchasable commodity and formulates governance mechanisms around conditions under which the principal has information to verify agent behavior (Eisenhardt, 1989). Recent IS research extends delegation to and from agentic IS artifacts, foregrounding how verification and accountability dependencies can also be mediated through delegated digital agents (Baird & Maruping, 2021).

Loose coupling theory characterizes organizations as systems in which elements are responsive yet preserve their own identity, such that coordination can be dampened as it moves through a system (Orton & Weick, 1990; Weick, 1976). In postimplementation enterprise system settings, enacted practices can remain loosely coupled from system data, such that system use does not necessarily imply process compliance (Berente & Yoo, 2012). Related evidence from NASA highlights how enterprise controls can privilege demonstrative compliance, leaving gaps between publicly displayed processes and everyday work practices (Berente et al., 2010).

A shared baseline persists across these lenses: control modes function when behavior or outcomes are observable, coupling perspectives presume that linkages can be tightened, and agency theory frames information asymmetry as reducible through monitoring investment. These lenses encounter a boundary condition when delivery architectures limit what signals are available for oversight: a condition where sustained monitoring yields diminishing returns because the signals required for oversight are unavailable, insufficiently granular, or uninterpretable. Shadow AI strains these theoretical foundations by presenting precisely this condition.

Conceptual Foundations

The conceptual foundations specify the focal condition (shadow AI), two generative mechanisms (visibility erosion and authorization decay), and three governance capabilities (observation, attribution, and auditability). These constructs identify what oversight requires in this setting and clarify how governance visibility can degrade even when systems are nominally known, approved, or monitored.

Shadow AI

Prior work has defined shadow AI in terms of unauthorized AI tools adopted by employees without organizational approval (Chin et al., 2025; Puthal et al., 2025). Some accounts further identify a governance drift zone where formal AI policies lack alignment with everyday AI tool use (Silic et al., 2025). This framing captures one important pattern, unauthorized tool use through shadow channels, but cannot account for three additional patterns where AI escapes effective oversight. Organizations encounter AI that remains approved yet operationally ungoverned as vendor updates modify behavior between review cycles. They face AI capabilities embedded within approved platforms without capability-level disclosure, such that they cannot specify which AI features are active or which models process data. And they experience AI that becomes normalized infrastructure, diffused across workflows and treated as unremarkable background capability. These patterns share the feature that governance fails despite different relationships to authorization.

We therefore define shadow AI more broadly as a governance condition in which organizational records and oversight mechanisms cannot reliably track what AI systems are actually doing. Specifically, this occurs when governance artifacts (the organizational records and control objects that encode authorization, accountability, and expected operating conditions) and monitoring traces (the runtime-generated records that support detection and reconstruction of operation) are absent, incomplete, or insufficiently interpretable to reliably represent runtime operation. Under this condition, organizational oversight may fail even when AI services are accessed through token and API usage traces, because internal operations can remain unobservable and non-auditable (Sun et al., 2025). The 'shadow' in shadow AI thus denotes invisibility to governance oversight rather than unauthorized status per se.

Visibility Erosion

Visibility erosion refers to the progressive reduction in an organization's capacity to exercise effective oversight over AI functionality in organizational workflows, even when monitoring intensity is held

constant or increased. Erosion reflects deterioration in the availability, granularity, or interpretability of the signals on which oversight depends. Delivery architectures produce this deterioration by thinning, relocating, or externalizing governance signals, for example through encrypted browser-based access that limits inspection (Newhouse, 2025), vendor-managed updates that shift control points (Kreuzberger et al., 2023), and API-mediated services that move telemetry outside organizational reach (Pawlaszczyk et al., 2022).

Authorization Decay

Authorization decay captures the progressive degradation of policy fit between an approved governance state and a system's runtime behavior. This mechanism arises when the velocity of system evolution, including retraining triggered by concept drift and ongoing deployment updates (Kreuzberger et al., 2023), exceeds the latency of governance review cycles, which can span weeks in practice (Lahiri & Saltz, 2024). Under these conditions, point-in-time approvals become perishable because deployed systems can change after deployment and drift from prior behavior (Kapusta et al., 2025). Procedural frictions can prevent reassessment from being triggered, allowing initially compliant systems to enter ambiguous or noncompliant states without an explicit governance decision. For example, an approved enterprise AI feature may be altered through a vendor-managed update while the organization's approval record continues to describe the configuration last reviewed. Until reassessment occurs, the recorded authorization no longer fully represents runtime operation.

Governance Capabilities

We conceptualize oversight through three governance capabilities that specify what organizations must be able to do to govern AI functionality. Observation is the capability to detect and characterize AI operation at runtime with sufficient granularity for oversight. In shadow AI settings, observation can be continuous yet insufficient for governance, such as when traffic is visible but lacks information about content or decision-relevant events (Newhouse, 2025; Wang et al., 2025). Visibility erosion depresses observation when oversight-relevant signals become unavailable or uninterpretable.

Attribution is the capability to link a specific outcome to the contributing model state, data, and decision pathway in a manner that supports responsibility assignment and diagnosis. A system may be observed in the narrow sense of recorded inputs and outputs yet exhibit low attribution if recipients lack a concrete sense of how or why a particular output was produced (Burrell, 2016). Attribution is therefore audience-relative: intelligibility to a user does not entail intelligibility to organizational oversight (Berente et al., 2021). Visibility erosion undermines attribution when opacity limits explanation of the decision rule or logic producing an output (Burrell, 2016; Kroll et al., 2017).

Auditability refers to the capability to reconstruct and evidence conformance to policies and requirements over time. It is high when durable artifacts connect approvals, configurations, versioning, and usage histories in ways that support retrospective examination. Visibility erosion weakens auditability when the artifacts connecting governance records to runtime behavior are incomplete, semantically thin, or not retained at the relevant granularity (Kreuzberger et al., 2023), such that recorded states cease to represent operational reality.

These capabilities are treated as analytically distinct and separately measurable dimensions that may vary separately while jointly defining the governance capability space, rather than a single maturity continuum. The theory does not assume causal or statistical independence, and a common barrier or barrier configuration may therefore degrade more than one capability simultaneously. The shadow AI types developed below are idealized outcome configurations within this space, sustained by characteristic barrier patterns that serve as analytical reference points even though empirical cases may exhibit hybrid or intermediate configurations.

These constructs define the capability space within which the typological theory is developed, and four premises specify the scope conditions under which that theory operates: shadow AI is a system-level governance condition rather than a property of any single tool or use case, emerging from local uses of AI-enabled functionality together with surrounding infrastructures, processes, and human practices; visibility erosion and authorization decay are structurally generated and analytically distinct, arising from delivery architectures and update practices rather than solely from gaps in monitoring intensity, and capable of

varying separately across systems; organizations can shape governance capabilities through technical architectures, process design, and role assignments; and empirical settings may exhibit hybrid or intermediate profiles and may transition between types over time.

AI Forms and Shadow Conditions

Organizational AI encompasses overlapping forms, including predictive or classificatory systems that infer outcomes or support decisions and generative systems that produce new content (Feuerriegel et al., 2024), as well as agentic systems that initiate or execute actions with varying degrees of autonomy (Baird & Maruping, 2021). No functional form is inherently shadow. Models managed through structured MLOps pipelines may be comparatively easier to register and monitor because model registries, versioned deployment artifacts, metadata stores, and monitoring components provide traceability over model versions, lineage, deployment status, and serving performance (Kreuzberger et al., 2023). Employee-adopted AI accessed through external cloud services, third-party applications, or personal devices may instead remain outside organizational knowledge (Puthal et al., 2025), while, for generative and agentic capabilities embedded within enterprise platforms, the capability-level records and monitoring data available to organizational oversight depend on what platform-owner components expose (Haki et al., 2025). Taken together, these contrasts shift the explanatory focus from functional form to delivery architecture. We theorize that, across predictive, generative, and agentic systems, shadow conditions are more likely to arise and more difficult to identify and monitor when capability identity, runtime activity, and change records are externalized or unavailable than when those signals remain accessible to organizational oversight.

A Typological Account of Shadow AI

We use conceptual typological theory building to explain how shadow AI configurations arise. Following Doty and Glick (1994), typologies constitute theory when they identify constructs, specify relationships among them, and generate falsifiable predictions; in Gregor's (2006) terms, the present account is a Type IV theory for explaining and predicting because it combines causal explanation through generative mechanisms with testable propositions about the configurations expected under stated conditions. Typological theories are suited to phenomena in which causality is configurational rather than reducible to a single linear relationship. Here, different barrier configurations may converge on the same governance-capability profile through equifinality, while differences among capability profiles provide the basis for distinguishing the ideal types. Shadow AI here is treated as a governance condition that can obtain independently of organizational recognition, while visibility erosion and authorization decay are specified as generative mechanisms whose effects are contingent on delivery architecture rather than deterministic. The constructs are treated as analytical representations of governance conditions, and the four types as conceptually derived ideal types: reference points that empirical cases may approximate to varying degrees rather than categories produced by clustering observed cases.

Building on the conceptual foundations, the core mechanism is visibility erosion: progressive degradation of governance capabilities as delivery architectures thin, relocate, or externalize required signals, persisting even when monitoring intensity increases. This operates alongside authorization decay. Barrier configurations vary in how they depress observation, attribution, and auditability unevenly, producing distinct outcome patterns in the governance capability space. The section specifies delivery architectures, identifies barrier mechanisms, derives the Monitoring Paradox, and presents the resulting four-type typology.

Delivery Architectures

Shadow AI conditions emerge through two broad delivery architectures differing in locus of control. Platform-mediated architectures deliver AI functionality through Software as a Service (SaaS) platforms and embedded features within enterprise applications, for example via vendor-managed integration and embedded features (Haki et al., 2025). Under these architectures, control points are relocated upstream to vendors, and capability enters workflows without discrete local deployment events or artifacts that conventional inventories capture. User-initiated architectures involve employees accessing external AI services for work tasks through personal accounts, browser interfaces, or consumer applications (Puthal et

al., 2025). Under these architectures, control is dispersed to individual users, and visibility depends on disclosure and behavioral signals rather than technical telemetry.

Both architectures can produce shadow AI conditions, but through different barrier configurations. Platform-mediated architectures intensify technical and procedural barriers as vendor-controlled updates and continuous delivery practices modify behavior between review cycles (Kreuzberger et al., 2023; Mahdavi-Hezaveh et al., 2021). User-initiated architectures intensify behavioral barriers as employees may avoid disclosure or fail to treat AI involvement as governance-relevant (Puthal et al., 2025), while technical barriers persist because API-mediated access produces insufficient local traces (Javadi et al., 2021; Pawlaszczyk et al., 2022). Figure 1 integrates the theory by showing how delivery architectures activate barrier mechanisms, generate visibility erosion and authorization decay, and produce distinct governance capability profiles and shadow AI types.

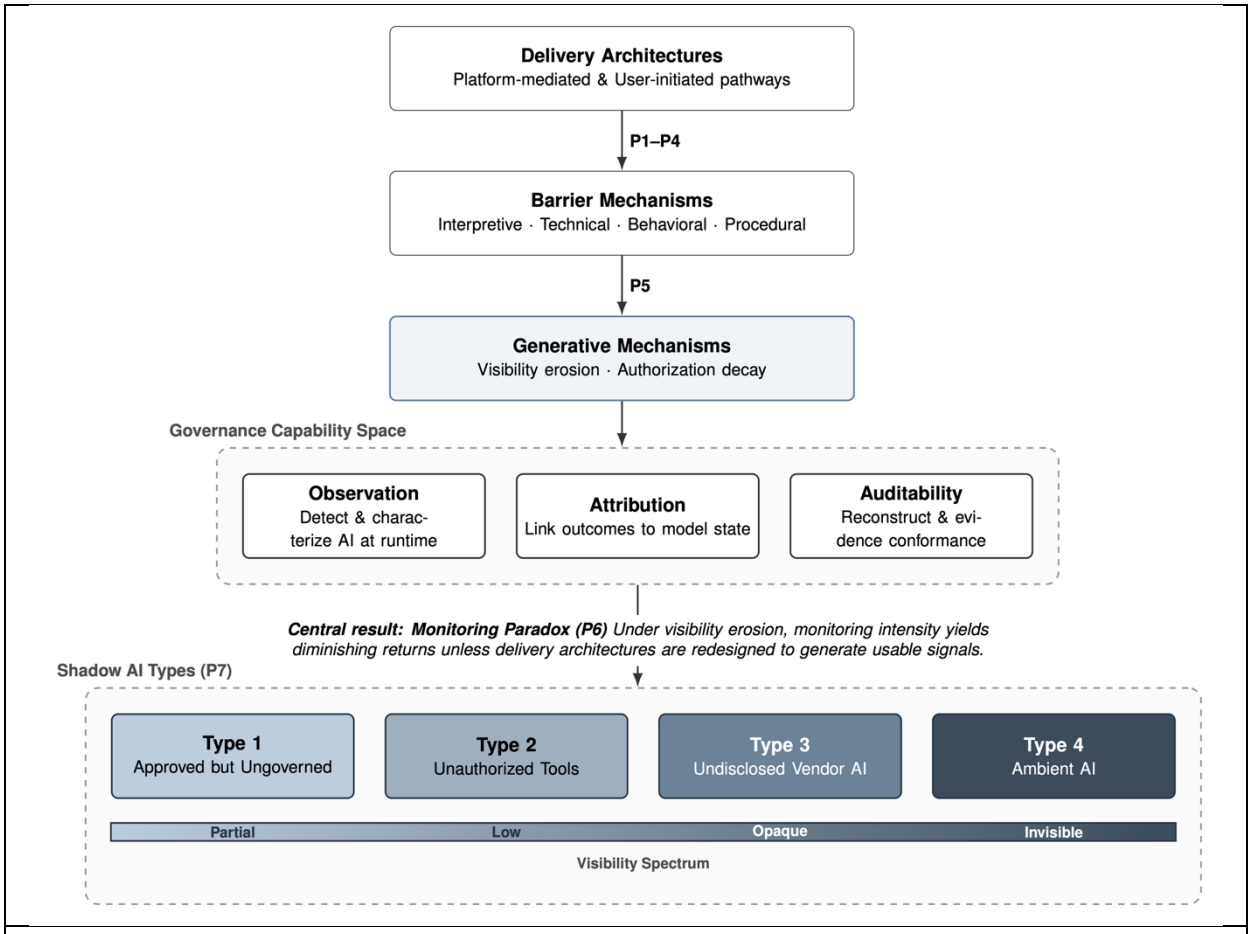

Figure 1. An Integrative Framework of Shadow AI Governance

Barrier Mechanisms

Four barrier mechanisms translate delivery architectures into capability degradation. Interpretive, technical, and behavioral barriers generate visibility erosion by degrading the availability, granularity, or interpretability of oversight-relevant traces, whereas procedural barriers primarily operate through authorization decay by weakening continuity between approvals and runtime states.

The propositions follow the theory's explanatory sequence. Propositions 1–4 specify the first-order capability effects of individual barriers, Proposition 5 explains how co-present barriers reinforce one another, Proposition 6 states the Monitoring Paradox, and Proposition 7 specifies the resulting capability configurations.

Mechanism propositions (P1–P4)

Propositions 1–4 state directional expectations relative to otherwise comparable settings in which the focal barrier is absent or materially weaker, with other delivery-architecture and governance conditions held constant.

Interpretive barriers prevent stakeholders from consistently recognizing or classifying AI-enabled behavior when the categories linking signals to governance routines destabilize. Divergent definitions of what constitutes AI allow similar features to be treated as benign enhancements rather than governed systems (Cox, 2024). Embedding intensifies this epistemic challenge: when capabilities appear through familiar interfaces, classification depends on contested interpretation rather than stable system boundaries, depressing attribution and weakening auditability.

Proposition 1: Interpretive barriers are expected to depress attribution by destabilizing classification and routing under embedding and routine enhancement framing, such that AI-enabled behavior is more likely to be treated as out of scope for AI-relevant governance even when activity is nominally visible.

Technical barriers arise when delivery architectures remove, relocate, or degrade the signals on which monitoring depends. API-mediated access to cloud AI services produces insufficient local traces, as evidence resides primarily in provider-controlled services (Javadi et al., 2021; Pawlaszczyk et al., 2022). Encryption and embedding further hide functionality behind trusted interfaces: secure channels limit inspection (Newhouse, 2025), while platform-embedded AI capabilities are mediated through provider-controlled interfaces that centralize access to prompts and outputs (Haki et al., 2025). Detection under these conditions requires engineered observability or vendor telemetry that may be unavailable or lack the granularity required for oversight (Wang et al., 2025). Additional monitoring intensity cannot recover signals that delivery architectures fail to generate.

Proposition 2: Technical barriers are expected to depress observation by thinning, relocating, or externalizing the oversight substrate, such that increased monitoring intensity does not reliably recover capability-level visibility.

Behavioral barriers arise because human disclosure and escalation form critical governance signals, yet these information flows can be systematically suppressed. Gradual rollout in familiar tools and low algorithmic literacy reduce salience, leading users to rely on outputs without labeling activity as AI-enabled (Cox, 2024). Normalization of these systems in everyday workflows and default enablement further suppress observable signals (Lomborg, 2022), depressing observation and attribution even where technical monitoring controls exist.

Proposition 3: Behavioral barriers are expected to depress observation and attribution under default enablement and normalization in familiar tools by suppressing disclosure, escalation, and governance-relevant information flow, thereby allowing visibility erosion to persist even where formal controls exist.

Procedural barriers create temporal mismatch when governance routines presuppose discrete deployment events that trigger review cycles but AI capabilities evolve continuously. Vendor-operated AI as a Service (AIaaS) changes behavior under provider control, while internally hosted models update via automated retraining pipelines (Kreuzberger et al., 2023; Pandl et al., 2021). Point-in-time approvals become perishable when behavior changes without generating the local deployment markers that would activate change-control processes. Dark launches, gradual rollouts, and long-lived feature toggles widen the timing gap between review cycles and system evolution (Hoyos et al., 2021; Mahdavi-Hezaveh et al., 2021), allowing initially compliant configurations to drift from approved states between reassessments.

Proposition 4: Procedural barriers are expected to sustain authorization decay by preventing governance routines from reliably triggering reassessment when runtime behavior changes, thereby weakening auditability through disrupted continuity between approvals, configurations, and runtime states.

Table 1 summarizes the four barrier mechanisms, their generative pathways, and the governance capabilities each primarily degrade. These first-order mappings provide the basis for the higher-order claims developed next concerning barrier co-occurrence, the Monitoring Paradox, and stable type configurations.

Barrier mechanism	Generative pathway	Capabilities depressed	How degradation operates	Proposition
Interpretive	Visibility erosion	Attribution *Auditability (secondary)*	Embedding in familiar interfaces destabilizes classification categories linking signals to governance routines → AI-enabled behavior framed as routine enhancement rather than a governed system	P1
Technical		Observation	Delivery architectures thin, relocate, or externalize oversight signals → API-mediated access produces insufficient local traces; encryption limits inspection	P2
Behavioral		Observation Attribution *(secondary)*	Default enablement and normalization in familiar tools suppress disclosure and escalation → users rely on outputs without recognizing or labeling activity as AI-enabled	P3
Procedural	Authorization decay	Auditability	Continuous updates bypass discrete deployment triggers that governance routines presuppose → point-in-time approvals become perishable	P4

Table 1. Barrier Mechanisms, Generative Pathways, and Governance Capability Effects

Delivery and governance conditions shape barrier configurations and oversight signals. Misclassification can route AI outside review; platform embedding and externalized telemetry can reduce signal availability, granularity, or interpretability; and non-disclosure, failed escalation, and normalization can prevent information from reaching governance. These barriers generate visibility erosion, weakening observation and attribution and, where reconstructive evidence is lost, auditability. Procedural barriers generate authorization decay when runtime change escapes reassessment, leaving approvals and recorded configurations misaligned with runtime operation.

Cumulative Erosion and the Monitoring Paradox

Configurational reinforcement (P5) occurs when co-present barriers interact to intensify visibility erosion and authorization decay. Two recurring interaction patterns illustrate this mechanism: a technical-behavioral combination allows embedded AI to be treated as routine infrastructure by suppressing the human signals that might compensate for thin telemetry, while an interpretive-procedural combination converts ambiguity into inaction by preventing reassessment when stakeholders cannot consistently classify a capability. In each case, one barrier degrades a governance signal while the other removes the pathway that might otherwise compensate for that loss, producing greater capability degradation than either would alone.

Proposition 5: When barriers co-occur, their effects are expected to be non-additive, such that capability degradation exceeds the sum of effects from each barrier operating alone.

The Monitoring Paradox (P6) shows how the non-additivity of barrier interactions constrains the effectiveness of conventional governance responses. Traditional governance assumes that expanded monitoring produces proportional improvements in oversight: additional scanning, logging, or audit activity should yield corresponding gains in detecting, attributing, and reconstructing AI operation. This proportionality assumption underpins widespread organizational investments in continuous monitoring systems and audit frequency as primary responses to shadow AI concerns. Under visibility erosion,

however, the assumption encounters a structural limit. When co-present barriers reinforce one another's degradation of governance signals, expanding inventories or scans reproduces the same blind spots at greater cost.

Monitoring can improve governability when capability-level signals remain accessible to organizational oversight, sufficiently granular and interpretable for observation and attribution, and connected to reassessment when runtime behavior changes. It yields diminishing returns when delivery architectures thin, relocate, or externalize those signals, or when procedural barriers prevent runtime change from triggering renewed review. Restoring governability under these conditions requires architectural change that makes the relevant signals available, procedural change that reinstates reassessment triggers, or both.

Proposition 6: Under visibility erosion, increases in monitoring intensity are expected to yield diminishing returns in governance effectiveness unless the architectural conditions that determine signal availability and governance triggers are changed.

Shadow AI Ideal Types

Outcome structure (P7) comprises four types derived analytically rather than by enumerating every possible combination of capability levels**.** Observation, attribution, and auditability vary continuously, while barrier patterns can depress them to different degrees. We retain a capability configuration as an ideal type when it follows from at least one theoretically specified barrier pattern, produces a distinct governance breakdown requiring a meaningfully different diagnostic or intervention response, and is approximated by at least one documented case. The case requirement provides a documented approximation of the configuration in practice; it does not provide evidence for the type's validity, prevalence, or causal claims. Within the scope of the present theory, applying these criteria yields four profiles that differ in their capability configuration, sustaining pathway, and diagnostic signature. Type 1 combines retained ownership with partial runtime observation and stale audit records when system evolution outpaces reassessment. Type 2 combines low organizational observation, locally strong attribution, and limited or absent auditability under user-initiated delivery. Type 3 preserves host-platform visibility while weakening capability-level attribution and auditability under vendor embedding. Type 4 reflects ambient embedding under barrier co-occurrence, with severe degradation across all three capabilities.

Because the resulting capability profile rather than a single barrier defines type membership, different delivery pathways and barrier combinations can converge on the same type. Personal-account use sustained by nondisclosure and unsanctioned API access sustained by externalized telemetry, for example, can both produce the Type 2 profile despite different dominant barriers. Other capability states are not assumed to disappear or collapse into the four types. They may approximate an ideal type, form hybrid, intermediate, or transitional configurations, or support a future extension if evidence establishes a distinct recurring profile, governance breakdown, and intervention requirement. Stability is therefore conditional on the continued operation of the characteristic barrier pattern.

Proposition 7: Shadow AI types are expected to manifest as stable capability configurations when characteristic barrier patterns persist, such that organizations facing similar barrier configurations are expected to exhibit similar profiles of observation, attribution, and auditability degradation.

The four types are depicted in Figure 1 and profiled in Table 2. Each type exhibits a distinct configuration across observation, attribution, and auditability, sustained by the barrier patterns specified in Table 1. The types are derived from uneven degradation in governance capabilities under different barrier configurations, rather than from authorization status alone. The capability descriptors are dimension-specific diagnostic labels rather than a single ordinal scale; what matters is the profile each type exhibits across observation, attribution, and auditability.

Each type represents a system-level governance condition sustained through the interaction of delivery architectures, barrier patterns, governance processes, and human practices, rather than a property of any single actor, technology, or incident. Documented examples make the delivery conditions and diagnostic signatures concrete. They were retained retrospectively because publicly available sources describe a relevant delivery condition and at least one governance limitation associated with the focal capability profile. The examples are illustrative rather than confirmatory, showing correspondence with documented settings and practical plausibility without constituting an empirical sample or validating the typology or its

causal claims. Where examples draw on litigation records, the descriptions refer to allegations rather than adjudicated findings of liability.

Type	Observation	Attribution	Auditability	Diagnostic signature
Type 1: Approved but Ungoverned	**Partial:** System is known, but runtime behavior is not fully captured.	**Strong:** Outcomes can be linked to assigned system owners.	**Stale:** Records diverge from current operation.	*Approval drift*
Type 2: Unauthorized Tools	**Low:** Leaves thin or indistinguishable traces.	**Locally strong:** Users know tools, but outputs cannot be linked to specific model states.	**Absent:** Formal records are limited or absent.	*Weak organizational visibility*
Type 3: Undisclosed Vendor AI	**Opaque:** Platform use is visible, but AI involvement is underspecified.	**Weakened:** Outcomes cannot be tied to a specific embedded capability.	**Ambiguous:** Records track the host platform, not the AI feature.	*Visibility without specification*
Type 4: Ambient AI	**Invisible:** AI is diffused across workflows without discrete events.	**Untraceable:** Outcomes cannot be traced to any identifiable model or data source.	**Void:** Governance artifacts do not represent actual operation.	*Governance invisibility*

Table 2. Shadow AI Types: Capability Profiles and Diagnostic Signatures

Note: Idealized reference configurations; empirical cases may exhibit hybrid or transitional profiles.

Type 1: Approved but Ungoverned. This type captures systems that remain formally approved while operationally ungoverned. The delivery architecture involves approved enterprise applications subject to vendor-controlled continuous updates (feature toggles, silent retraining, and configuration changes) that modify system behavior between governance review cycles. Observation is partial: systems remain visible in inventories and generate activity traces, but runtime behavior is not fully captured. Attribution is comparatively strong because systems retain assigned owners. Auditability is stale because governance artifacts document the system as last reviewed, while vendor-driven evolution alters the capability profile faster than governance routines trigger reassessment. Procedural barriers sustain this configuration through timing mismatches that prevent reassessment from activating when operational states diverge from approved baselines. Microsoft 365 Copilot illustrates this pattern: organizations approve Copilot as part of their existing Microsoft 365 deployment, yet Copilot inherits user permissions across SharePoint and surfaces content that oversharing policies had left accessible but practically obscured (Microsoft, 2025). The breakdown is temporal rather than spatial: the system is not hidden but changes faster than governance can track.

Type 2: Unauthorized Tools. AI use is visible to the user but invisible to the organization because employees access external AI tools that organizational monitoring cannot adequately detect. The delivery architecture places AI consumption outside organizational infrastructure through personal accounts, browser-mediated interfaces, and unsanctioned APIs that generate few host artifacts. Observation is low because these externalized channels produce signals insufficient to establish capability-level AI use. Attribution remains locally strong, since users know which services they invoked, yet weak at the capability level, as outputs cannot be linked to model states or decision pathways. Auditability is low because the delivery architecture generates limited or no formal records. Technical and behavioral barriers jointly sustain this configuration: shadow channels externalize telemetry while behavioral dynamics suppress disclosure that might compensate for thin technical signals. In 2023, after an engineer leaked sensitive internal source code through ChatGPT, Samsung issued a company-wide ban on generative AI tools (Ray, 2023). The incident constitutes a data leakage event in which proprietary information left organizational

control through a user-initiated channel and was identified after the upload, illustrating how user-initiated architectures can move sensitive data outside organizational control before governance intervention occurs.

Type 3: Undisclosed Vendor AI. Governance covers the platform but not the AI within it because embedded capabilities inherit the approval status of their host system without establishing a capability-specific governance footprint. SaaS and enterprise platforms can introduce AI within already approved workflows while disclosing neither stable model identity nor feature-level change history at the granularity required for organizational oversight. Observation is opaque: organizations can confirm platform access and host-level usage, yet cannot determine which AI capability was active, what decision logic was applied, or what changed across updates. Attribution is weakened because governance artifacts assign responsibility at the platform level but do not reliably link outcomes to a specific embedded capability, model state, or configuration. Auditability is ambiguous because approval records track the host platform rather than the embedded AI feature and its changing parameters over time. The Workday litigation illustrates this pattern: the public record describes AI and machine learning embedded in employer-facing screening tools within a visible, employer-governed application platform, while leaving the operative scoring or recommendation logic insufficiently specified for organizational attribution and audit reconstruction (Mobley v. Workday, Inc., 2024; U.S. Equal Employment Opportunity Commission, 2024). The resulting condition is visibility without specification: the platform is governed, but the embedded AI capability remains insufficiently specified for robust attribution and auditability. This attribution gap creates exposure under emerging compliance regimes, including the EU AI Act's high-risk system requirements for employment-related AI uses (European Parliament & Council, 2024), because organizations cannot evidence what the embedded capability did or how it changed.

Type 4: Ambient AI. AI capability is treated as infrastructure: ambient, default-on, and diffuse across everyday workflows. Unlike discrete tools triggering adoption decisions, the delivery architecture integrates AI within core platforms as a background condition with minimal local oversight. This type yields severe degradation across all three governance capabilities under maximal barrier co-occurrence. Deep embedding thins and externalizes signals (technical), default enablement eliminates adoption events that would otherwise trigger review (procedural), pervasive integration normalizes AI as unremarkable infrastructure (interpretive), and the absence of discrete use episodes suppresses recognition and escalation (behavioral). Otter.ai illustrates this configuration. Otter's own documentation states that its Notetaker can be connected to a user's calendar and configured to auto-join Zoom, Google Meet, or Microsoft Teams meetings, including settings in which it joins all calendar events with a valid meeting link (Otter Help, 2025a, 2025b). The subsequent Otter.ai privacy litigation further illustrates how governance failures tied to this ambient architecture can surface externally rather than through internal detection: plaintiffs allege that the notetaker joined meetings, captured conversations involving non-users, and did so without adequate notice or consent (Brewer v. Otter.ai, Inc., 2025), and related actions were later consolidated into In re Otter.AI Privacy Litigation (2025). The resulting condition is governance invisibility, in which AI ceases to register as a distinct governable entity.

Discussion

Theoretical Contributions

Our contribution is a system-level typological theory that explains shadow AI as a governance condition by linking AI delivery architectures to interpretive, technical, behavioral, and procedural barriers, tracing how those barriers generate visibility erosion and authorization decay, and representing the resulting configurations of observation, attribution, and auditability as four ideal types. The theory's central result is the Monitoring Paradox, which identifies a boundary condition under which the restoration premise that unites shadow systems and governance literatures ceases to hold.

The Monitoring Paradox identifies the operating conditions of this shared logic (Proposition 6): monitoring improves governability only where delivery architectures preserve the signals oversight consumes and runtime change triggers reassessment. When barriers co-occur, their combined effects exceed what individual barriers produce alone, blocking the pathways that might restore governance capacity (Proposition 5). This result qualifies rather than rejects inherited governance lenses: it identifies signal availability as the architectural constraint, suggesting governance research should treat it as a variable shaped by delivery architecture rather than a fixed parameter. Survey evidence that perceived

organizational compliance intensity did not predict lower shadow AI disclosure behavior (Sebastian, 2026) is consistent with the more general pattern developed here, in which intensifying conventional governance levers encounters limits set by conditions those levers do not control.

This constraint matters for formal governance. The EU AI Act requires documentation, logging, monitoring, and human oversight for relevant high-risk systems (European Parliament & Council, 2024), while the NIST AI Risk Management Framework and ISO/IEC 42001 support structured organizational AI risk management (International Organization for Standardization, 2023; National Institute of Standards and Technology, 2023). Their application depends on organizations having sufficiently granular and interpretable information about capability identity, runtime activity, and change. When delivery architectures do not provide those signals, compliance verification and risk-management processes are structurally constrained rather than merely more difficult.

Defining shadow AI as a governance condition explains why authorization and governability can decouple; approved systems may remain operationally ungoverned (Type 1), while unauthorized use may retain locally strong attribution despite weak organizational observation (Type 2). Authorization status alone is therefore insufficient for capability-level governability when delivery architectures degrade oversight signals even after discovery or approval.

The governance capability framework and typology provide the theory's diagnostic structure by showing how governability degrades and what patterns result. The three capabilities, observation, attribution, and auditability, treated as analytically distinct and separately measurable dimensions, enable diagnosis of dimension-specific deficits and move beyond binary authorization logic that treats governance as uniformly present or absent. These capabilities parallel the security functions of runtime detection, outcome attribution, and forensic reconstruction. Effective organizational oversight and accountability depend on the evidence these capabilities provide (Kroll et al., 2017). Explainability can support attribution but does not substitute for observation or auditability. The four shadow AI types express the outcome structure of this capability space as recurring configurations sustained by characteristic barrier patterns (Proposition 7), rather than a surface taxonomy of tools or authorization states.

The typology also implies transition dynamics among types, depicted in Figure 2. A primary drift pathway runs from unauthorized use (Type 2) through approval and organizational integration that establish a governed baseline but do not restore temporal alignment (Type 1), to vendor embedding (Type 3) and ambient infrastructure (Type 4).

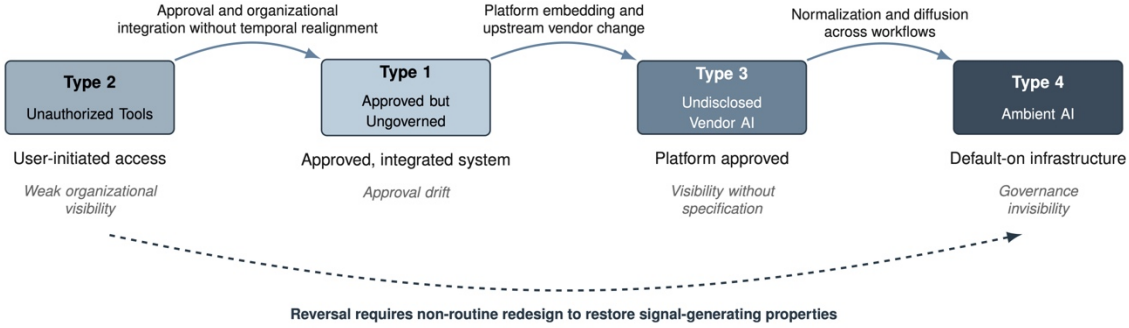

Figure 2. Asymmetric Drift Pathway Among Shadow AI Types

Forward drift occurs when approval and integration reshape procedural barriers, platform embedding and vendor-managed updates intensify technical barriers, or normalization deepens behavioral and interpretive barriers, shifting the capability profile toward a different ideal type. An organization that approves and integrates a previously unauthorized tool does not thereby resolve governance deficits if the resulting system inherits vendor-controlled update practices that outpace governance review. The pathway is structurally asymmetric because reversal requires deliberate redesign of delivery architectures and governance routines to restore oversight-relevant signals and reinstate reassessment triggers. Drift is therefore self-reinforcing, and remediation becomes progressively more costly as co-present barriers accumulate.

Practical Implications

The theory implies that monitoring intensity is not a generic remedy for shadow AI. When delivery architectures thin, relocate, or externalize governance signals, monitoring investments yield diminishing returns. Practical response begins with capability-profile diagnosis rather than tool- or status-based classification, using three questions: Can AI operation be detected and characterized at runtime (observation)? Can outcomes be linked to contributing model states and decision pathways (attribution)? Can past configurations and conformance be reconstructed over time (auditability)?

Across types, intervention follows three principles: where signal loss is architectural, delivery architecture design takes priority over monitoring intensity because observability and capability-level vendor data must be built in; reassessment triggers must match update velocity because routines based on discrete deployment events fail as systems evolve; and governance artifacts must fit delivery architectures, because artifacts designed for discrete systems cannot represent capabilities entering through embedding or silent rollout. These principles yield type-specific levers corresponding to established security and privacy controls. For Type 1, contractual vendor change notifications support change-management and configuration-assurance by triggering reassessment when systems evolve. For Type 2, governed AI access channels using enterprise gateways, data loss prevention, and access policy preserve forensic and compliance traces while avoiding blanket bans that push use into less observable channels. For Type 3, procurement clauses requiring capability-level disclosure of embedded AI features extend third-party risk management beyond platform-level assurance. For Type 4, bounded governability supports risk-based prioritization by concentrating oversight on critical-path, high-stakes uses when ambient AI cannot be governed at the same granularity. Lever selection therefore follows the capability profile rather than the tool or its authorization status: the same control can restore governability in one configuration and reproduce the existing blind spot at greater cost in another. Because drift is asymmetric, the cost of remediation also depends on when a condition is diagnosed, not only on which lever is applied.

Limitations

The theory specifies a boundary condition, not a universal claim about AI governance: it does not hold that all AI governance fails or that monitoring is universally ineffective. The four types are idealized configurations rather than exhaustive categories; empirical settings may exhibit hybrid configurations and transitions over time, and researchers should interpret observed cases as approximations located by their capability profiles. The unit of analysis is the system-level governance condition rather than individual tools or adoption decisions, enabling explanation of systemic capability degradation but not accounting for micro-level variation in user behavior or local governance design choices. The theory treats its constructs as sector-general but does not model regulatory or institutional moderation of barrier salience. Future research should examine whether such conditions alter capability profiles and type prevalence, including whether EU AI Act requirements concerning system registration, provider transparency, and conformity assessment for high-risk AI systems (European Parliament & Council, 2024) reshape visibility dynamics.

Implications for Research

Empirical validation should operationalize the capabilities and test Propositions 5 and 6. Proposition 6 can be tested by comparing settings where governance artifacts represent runtime operation with those where delivery architectures thin or externalize signals, assessing whether monitoring produces proportional gains or diminishing returns. Proposition 5 requires testing whether co-present barriers produce capability degradation beyond their separate effects. Measurement should target organizational capability rather than technical features alone: trace availability and granularity (observation), linkage of outcomes to model states (attribution), and reconstruction of configurations and conformance over time (auditability). These measures should capture distinguishable dimensions rather than collapse them into a single maturity continuum; this separability does not imply causal or statistical independence.

Longitudinal designs can test whether type shifts follow changes in delivery architectures, treating shift from monitoring intensification alone as disconfirming evidence. Cross-sectional comparisons of similar AI capabilities under different architectures can test whether architecture predicts capability profiles. Case studies can assess whether approval drift, weak organizational visibility, visibility without specification, and governance invisibility appear as distinct governance breakdowns.

Conclusion

This paper explains how AI functionality can drift outside organizational oversight even when formally approved and monitored. Its system-level typological theory links delivery architecture to visibility erosion and authorization decay, the degradation of observation, attribution, and auditability, and four ideal types of shadow AI. The Monitoring Paradox specifies when the restoration premise shared across shadow systems and governance literatures reaches its limit. The theory reframes the governance challenge, asking not whether AI is known and approved but whether governance capabilities remain coupled to runtime operation. As AI enters workflows through platform embedding, continuous updates, and ambient infrastructure, that question matters for governance research, security and privacy practice, and regulatory frameworks requiring verifiable AI behavior.

References

Baird, A., & Maruping, L. M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Quarterly*, *45*(1), 315-341. https://doi.org/10.25300/MISQ/2021/15882

Berente, N., Gal, U., & Yoo, Y. (2010). Dressage, control, and enterprise systems: The case of NASA's full cost initiative. *European Journal of Information Systems*, *19*(1), 21-34. https://doi.org/10.1057/ejis.2009.47

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly, 45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Berente, N., & Yoo, Y. (2012). Institutional contradictions and loose coupling: Postimplementation of NASA's enterprise information system. *Information Systems Research*, *23*(2), 376-396. https://doi.org/10.1287/isre.1110.0373

Brewer v. Otter.ai, Inc., Class Action Complaint, No. 5:25-cv-06911-EKL (N.D. Cal. filed Aug. 15, 2025).

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1-12. https://doi.org/10.1177/2053951715622512

Campbell, M., & Jovanovic, M. (2024). Disinfecting AI: Mitigating generative AI's top risks. *Computer*, *57*(5), 111-116. https://doi.org/10.1109/MC.2024.3374433

Chin, T., Li, Q., Mirone, F., & Papa, A. (2025). Conflicting impacts of shadow AI usage on knowledge leakage in metaverse-based business models: A Yin-Yang paradox framing. *Technology in Society*, *81*, 102793. https://doi.org/10.1016/j.techsoc.2024.102793

Cox, A. (2024). Algorithmic literacy, AI literacy and responsible generative AI literacy. *Journal of Web Librarianship*, *18*(3), 93-110. https://doi.org/10.1080/19322909.2024.2395341

Doty, D. H., & Glick, W. H. (1994). Typologies as a unique form of theory building: Toward improved understanding and modeling. *Academy of Management Review*, *19*(2), 230-251. https://doi.org/10.5465/amr.1994.9410210748

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, *14*(1), 57-74. https://doi.org/10.5465/amr.1989.4279003

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L, 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Finister, J., Isaychenko, D., Karu, K., Mann, S., Moore, S. J., & Zhuravlev, R. (2025). *ITIL AI Governance Global Best Practice: White Paper* (Version 1.0). PeopleCert International Limited. https://www.itil.com/Itil-News-and-Announcements/ai-governance-white-paper

Fürstenau, D., Rothe, H., & Sandner, M. (2021). Leaving the shadow: A configurational approach to explain post-identification outcomes of shadow IT systems. *Business & Information Systems Engineering*, *63*(2), 97-111. https://doi.org/10.1007/s12599-020-00635-2

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, *30*(3), 611–642. https://doi.org/10.2307/25148742

Haag, S., & Eckhardt, A. (2017). Shadow IT. *Business & Information Systems Engineering*, *59*(6), 469-473. https://doi.org/10.1007/s12599-017-0497-x

Haki, K., Safaei, D., Magan, A., & Griffiths, M. (2025). Integrating generative AI into enterprise platforms: Insights from Salesforce. *Information Systems Journal*, *35*(5), 1497-1512. https://doi.org/10.1111/isj.12593

Horton, R., Michalski, J., Winters, S., Gunn, D., & Holland, J. (2025). *AI ROI: The paradox of rising investment and elusive returns*. Deloitte. https://www.deloitte.com/global/en/issues/generative-ai/ai-roi-the-paradox-of-rising-investment-and-elusive-returns.html

Hoyos, J., Abdalkareem, R., Mujahid, S., Shihab, E., & Espinosa Bedoya, A. (2021). On the removal of feature toggles. *Empirical Software Engineering*, *26*, Article 15. https://doi.org/10.1007/s10664-020-09902-y

IBM. (2026). *Cost of a data breach report 2026: The AI tipping point*. https://www.ibm.com/reports/data-breach

In re Otter.AI Privacy Litigation, Consolidated Class Action Complaint, No. 5:25-cv-06911-EKL (N.D. Cal. filed Dec. 5, 2025).

International Organization for Standardization. (2018). *Information technology—Service management—Part 1: Service management system requirements* (ISO/IEC Standard No. 20000-1:2018). https://www.iso.org/standard/70636.html

International Organization for Standardization. (2023). *Information technology - Artificial intelligence - Management system* (ISO/IEC 42001:2023). https://www.iso.org/standard/81230.html

ISACA. (2025). *Leveraging COBIT for Effective AI System Governance*. ISACA. https://www.isaca.org/resources/white-papers/2025/leveraging-cobit-for-effective-ai-system-governance

Javadi, S. A., Norval, C., Cloete, R., & Singh, J. (2021). Monitoring AI services for misuse. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 597-607. https://doi.org/10.1145/3461702.3462566

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*(4), 305-360. https://doi.org/10.1016/0304-405X(76)90026-X

Kapusta, A. S., Jin, D., Teague, P. M., Houston, R. A., Elliott, J. B., Park, G. Y., & Holdren, S. S. (2025). A framework for the assurance of AI-enabled systems. *Proceedings of SPIE: Vol. 13476. Assurance and Security for AI-enabled Systems 2025*, 134760C. https://doi.org/10.1117/12.3056719

Kirsch, L. S. (1997). Portfolios of control modes and IS project management. *Information Systems Research*, *8*(3), 215-239. https://doi.org/10.1287/isre.8.3.215

Klotz, S., Kopper, A., Westner, M., & Strahringer, S. (2019). Causing factors, outcomes, and governance of shadow IT and business-managed IT: A systematic literature review. *International Journal of Information Systems and Project Management*, *7*(1), 15-43. https://doi.org/10.12821/ijispm070102

Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (MLOps): Overview, definition, and architecture. *IEEE Access*, *11*, 31866-31879. https://doi.org/10.1109/ACCESS.2023.3262138

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, *165*(3), 633-705. https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/

Lahiri, S., & Saltz, J. (2024). The need for a risk management framework for data science projects: A systematic literature review. *International Journal of Information Systems and Project Management*, *12*(4), 41-57. https://doi.org/10.12821/ijispm120403

Lomborg, S. (2022). Everyday AI at work: Self-tracking and automated communication for smart work. In S. Pink, M. Berg, D. Lupton, & M. Ruckenstein (Eds.), *Everyday automation* (pp. 126-139). Routledge. https://doi.org/10.4324/9781003170884-11

Magunduni, J., & Chigona, W. (2018). Revisiting shadow IT research: What we already know, what we still need to know, and how do we get there? *2018 Conference on Information Communications Technology and Society (ICTAS)*, 1-6. IEEE. https://doi.org/10.1109/ICTAS.2018.8368735

Mahdavi-Hezaveh, R., Dremann, J., & Williams, L. (2021). Software development with feature toggles: Practices used by practitioners. *Empirical Software Engineering*, *26*, Article 1. https://doi.org/10.1007/s10664-020-09901-z

Microsoft. (2025). *Secure and govern Microsoft 365 Copilot: Foundational deployment guidance*. Microsoft Learn. https://learn.microsoft.com/en-us/microsoft-365/copilot/secure-govern-copilot-foundational-deployment-guidance

Mobley v. Workday, Inc., 740 F. Supp. 3d 796 (N.D. Cal. 2024).

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Newhouse, W. (2025). *Addressing visibility challenges with TLS 1.3 within the enterprise* (NIST SP 1800-37). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1800-37

Orr, S. G., Bonyadi, C. J., Golaszewski, E., Sherman, A. T., Peterson, P. A. H., Forno, R., Johns, S., & Rodriguez, J. (2024). Shadow IT in higher education: Survey and case study for cybersecurity. *Cryptologia*, *48*(1), 26-90. https://doi.org/10.1080/01611194.2022.2103754

Orton, J. D., & Weick, K. E. (1990). Loosely coupled systems: A reconceptualization. *Academy of Management Review*, *15*(2), 203-223. https://doi.org/10.5465/AMR.1990.4308154

Otter Help. (2025a, July 2). *Automatically add Otter Notetaker to your meetings*. https://help.otter.ai/hc/en-us/articles/13674910923671-Automatically-add-Otter-Notetaker-to-your-meetings

Otter Help. (2025b, July 2). *Manage your calendar meeting events*. https://help.otter.ai/hc/en-us/articles/13676368852631-Manage-your-calendar-meeting-events

Ouchi, W. G. (1979). A conceptual framework for the design of organizational control mechanisms. *Management Science*, *25*(9), 833-848. https://doi.org/10.1287/mnsc.25.9.833

Pandl, K. D., Teigeler, H., Lins, S., Thiebes, S., & Sunyaev, A. (2021). Drivers and inhibitors for organizations' intention to adopt artificial intelligence as a service. *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)*, 1769-1778. https://doi.org/10.24251/HICSS.2021.215

Pawlaszczyk, D., Bochmann, M., Engler, P., Klaver, C., & Hummert, C. (2022). API-based evidence acquisition in the cloud – A survey. *Open Research Europe*, *2*, 69. https://doi.org/10.12688/openreseurope.14784.1

Puthal, D., Mishra, A. K., Mohanty, S. P., Longo, A., & Yeun, C. Y. (2025). Shadow AI: Cyber security implications, opportunities and challenges in the unseen frontier. *SN Computer Science*, *6*(5), 405. https://doi.org/10.1007/s42979-025-03962-x

Raković, L., Sakal, M., Matković, P., & Marić, M. (2020). Shadow IT—A systematic literature review. *Information Technology and Control*, *49*(1), 144-160. https://doi.org/10.5755/j01.itc.49.1.23801

Ray, S. (2023, May 2). Samsung bans ChatGPT among employees after sensitive code leak. Forbes. https://www.forbes.com/sites/siladityaray/2023/05/02/samsung-bans-chatgpt-and-other-chatbots-for-employees-after-sensitive-code-leak

Sebastian, G. (2026). Digital shadow AI risk theory (DART): A framework for managing data disclosure and privacy risks of AI tools at work. *Technological Forecasting & Social Change*, *229*, 124697. https://doi.org/10.1016/j.techfore.2026.124697

Silic, M., & Back, A. (2014). Shadow IT—A view from behind the curtain. *Computers & Security*, *45*, 274-283. https://doi.org/10.1016/j.cose.2014.06.007

Silic, M., Silic, D., & Kind-Trüller, K. (2025). From shadow IT to shadow AI—Threats, risks and opportunities for organizations. *Strategic Change*. https://doi.org/10.1002/jsc.2682

Sun, G., Wang, Z., Zhao, X., Tian, B., Shen, Z., He, Y., Xing, J., & Li, A. (2025). Invisible tokens, visible bills: The urgent need to audit hidden operations in opaque LLM services. *arXiv*. https://doi.org/10.48550/arxiv.2505.18471

U.S. Equal Employment Opportunity Commission. (2024, April 9). *Brief of the Equal Employment Opportunity Commission as amicus curiae in support of plaintiff and in opposition to defendant's motion to dismiss*, *Mobley v. Workday, Inc.*, No. 3:23-cv-00770-RFL (N.D. Cal.)

Wang, C. L., Singhal, T., Kelkar, A., & Tuo, J. (2025). MI9: An integrated runtime governance framework for agentic AI. *arXiv*. https://doi.org/10.48550/arxiv.2508.03858

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, *21*(1), 1-19. https://doi.org/10.2307/2391875

Zimmermann, S., Rentrop, C., & Felden, C. (2017). A multiple case study on the nature and management of shadow information technology. *Journal of Information Systems*, *31*(1), 79-101. https://doi.org/10.2308/isys-51579