

Space AI Readiness as a Proto-Institution: Governing, Operating, and Exiting AI in Physical Spaces

Yingdong Zhu¹, Wen, Xu², Xue, Wang¹, Xin, Wang^{1*}

¹*School of Marxism, Zhejiang Shuren University, Hangzhou, China*

²*OPENVIBE Ltd, Singapore*

**Corresponding Author*

Abstract: As AI migrates from screen-based applications into buildings, campuses, and commercial venues, a consistent failure mode emerges: spaces become “AI-enabled” as a display layer (dashboards, assistants, smart panels) while remaining operationally fragile. Under model drift, vendor turnover, or contested accountability, AI shifts from a productivity promise to a new source of friction and downtime. This paper introduces Space AI Readiness, defined as the durable socio-technical preconditions that allow AI to participate in everyday production, coordination, and operations within a physical space without eroding governance, responsibility, or service continuity. Building on Science and Technology Studies (STS) and Umberto Eco’s limits of interpretation, we conceptualize readiness along two coupled dimensions: (1) governance/interpretation, where AI outputs must be contestable, traceable, and ultimately subordinated to accountable human judgment; and (2) operational infrastructure, where AI must be substitutable and exitable so the space remains runnable under human takeover and across vendor or model changes. We formalize this dual-axis construct as an evaluative framework and operationalize it through an L0–L5 readiness ladder adapted from autonomy discourse, clarifying progression from conventional operations to adaptive, semi-autonomous regimes. We further theorize Space AI Readiness as a proto-institution—a shared rule-set and evidence regime that coordinates owners, operators, tenants, service providers, and visitors by reducing interpretive disputes and decision costs. Finally, we translate the framework into audit-ready deliverables (task-based operating blueprints, RACI/WBS packages, evidence chains, data-and-licensing matrices, and exit tests) that enable contracts

to bind renovation, operation, and AI governance into a verifiable loop. The paper offers a rigorous vocabulary and an implementation grammar that shifts debate from “AI adoption” to governable operation in the built environment.

Keywords: Space AI Readiness; Built Environment; Socio-Technical Governance; Auditability; Contestability; Operational Continuity; Vendor Exit; Proto-Institutions

1. Introduction

Buildings are increasingly expected to behave less like static assets and more like products: measurable, updateable, and continuously optimized through data-driven coordination. Yet the dominant logic of renovation and deployment remains episodic and project-based, delivering a finished artifact—space plus equipment—rather than a running system capable of sustained coordination, accountability, and iteration. This structural mismatch has become a primary economic constraint in commercial and institutional spaces, where operating costs are persistent while capital investments are intermittent.

The introduction of large language models and agentic workflows does not resolve this mismatch by default. Instead, it often intensifies it by introducing new forms of opacity, vendor dependence, and interpretive ambiguity. As AI systems begin to participate in everyday spatial operations—dispatching tasks, mediating access, interpreting sensor data, or recommending interventions—the question is no longer whether a space is “AI-enabled,” but whether it can remain governable when AI outputs are contested, degraded, or withdrawn.

A recurring misframing follows. Organizations typically ask whether a space uses AI—whether it deploys dashboards, assistants, or automated controls—rather than whether the space can

sustain AI participation within its operating relations. When an AI feature fails or produces disputed outputs, several governance-critical questions surface: Who is authorized to veto or override the system? What evidence supports or challenges an AI-generated decision? Can operations continue if the AI is suspended? Can vendors or models be replaced without erasing the operational memory embedded in logs, configurations, and routines? These are not questions of user experience or interface design. They are questions of governance, responsibility, and continuity.

This paper argues that the effective integration of AI into the built environment requires a shift in the unit of analysis—from tools and applications to space as a socio-technical operating system. We propose Space AI Readiness as a decision entry-point for owners and operators facing this shift. Readiness, as used here, does not primarily refer to the skill level or technical sophistication of individual actors, who can in principle acquire new tools at any time. Instead, it denotes the public, shared conditions embedded in the space itself—touchpoints, permissions, task semantics, evidence trails, and exit paths—that reduce multi-actor friction and decision costs when AI becomes part of everyday operations.

1.1 Research Questions

Building on this reframing, the paper addresses three research questions:

RQ1. How should Space AI Readiness be defined so that it is both theoretically grounded in socio-technical governance and operable within contractual and operational settings?

RQ2. How can readiness be assessed and communicated in a manner that aligns renovation decisions, day-to-day operations, and accountability across heterogeneous actors such as owners, operators, tenants, and service providers?

RQ3. What kinds of deliverables and evidence structures are required to make readiness auditable—verifiable through documentation and traces rather than through demonstrations or promises?

1.2 Contributions

This paper makes four contributions.

First, it offers a working definition of Space AI Readiness centered on four governance-critical properties: vetoability, traceability,

replaceability, and human takeover. Together, these properties specify the minimum conditions under which AI can participate in spatial operations without undermining responsibility or continuity.

Second, it develops a theoretical synthesis between Science and Technology Studies (STS) and Umberto Eco's notion of the limits of interpretation, arguing that AI in physical spaces must be governed not merely as a technical system but as an interpretive regime whose outputs require bounded meaning, contestability, and institutional anchoring.

Third, it introduces a dual-axis analytical framework—governance/interpretation and operational infrastructure—and an accompanying L0–L5 readiness ladder to support evaluation, comparison, and roadmapping from conventional operations to adaptive, semi-autonomous regimes.

Finally, the paper translates this framework into an audit-ready implementation grammar, including task-based operating blueprints, RACI/WBS work packages, evidence chains, data-and-licensing matrices, and exit tests. These artifacts enable AI governance, renovation, and operation to be bound together contractually in a verifiable and enforceable manner.

2. Background and Related Work

Research on smart buildings, digital twins, and building management systems (BMS) has long focused on instrumentation, control loops, and optimization—typically through sensors, actuators, supervisory platforms, and efficiency-oriented KPIs. Recent “AI in the built environment” work extends this trajectory toward prediction (fault detection, demand forecasting), personalization (occupant-centric control), and higher degrees of automation (adaptive scheduling and dispatch). Across these strands, a common assumption often persists: improving model capability leads to improved operational outcomes.

However, as socio-technical scholarship has repeatedly shown, technologies do not merely add capability; they reconfigure work, accountability, and meaning-making (e. g., Suchman; Latour). This matters more—not less—when AI systems move into operational roles. Agentic workflows and large models introduce outputs that are probabilistic, context-dependent, and frequently non-transparent, which can widen interpretive disagreement and

complicate responsibility attribution. In building operations, these frictions are amplified by multi-actor ecology: owners, operators, tenants, vendors, and regulators interact through routines and documentation that must hold up under inspection, incident response, and contractual dispute.

In parallel, governance-oriented literature has produced influential cross-sector frameworks for responsible AI. ISO/IEC 42001 specifies requirements and guidance for establishing and continually improving an organizational AI management system, foregrounding governance, transparency, and risk-handling processes. NIST's AI Risk Management Framework (AI RMF 1.0) provides a lifecycle-oriented, voluntary framework to manage AI risks and promote trustworthy AI, also emphasizing governance structures, accountability, and controls. Within European building policy, the Smart Readiness Indicator (SRI) offers a vocabulary for "smart capability" in buildings—primarily around energy performance, occupant needs, and grid interaction.

Yet these strands leave a practical gap at the level where built environments succeed or fail: operational embedment under change. Owners and operators face a question that is not fully answered by model-centric building AI research, nor directly operationalized by general AI governance frameworks, nor captured by energy-centered "smartness" indicators:

What must be delivered—materially, procedurally, and evidentially—so AI can be integrated into operations without locking the space into a vendor, a model, or a non-auditable decision regime?

This paper positions Space AI Readiness as a bridge between socio-technical theory and "deliverable engineering": a concept that specifies the preconditions for AI to participate in spatial operations while preserving governance, responsibility, and continuity.

3. Theoretical Foundation: STS And The Limits Of Interpretation

We treat Space AI Readiness as a socio-technical institution-in-the-making rather than a technical maturity score. This stance is motivated by two complementary lenses: Science and Technology Studies (STS)—which explains how heterogeneous actors stabilize order through artifacts, routines, and standards—and Umberto Eco's notion of the limits of

interpretation, which frames meaning-making as bounded by rules, communities, and evidentiary practices.

3.1 STS: from "Systems" to Actor–Networks in the Built Environment

STS rejects a simple "technology impacts society" storyline and instead highlights how socio-technical order is assembled through translation: actors align interests by enrolling devices, documents, metrics, standards, and routines into a network (e. g., Callon; Latour). In buildings, this network is inherently plural: owners and asset managers, facility teams, tenants, contractors, service providers, compliance auditors, and visitors—alongside material infrastructures such as doors, HVAC, elevators, cameras, networks, and control rooms. From this view, AI is not a detachable feature layered onto a building. It becomes a mediator that can reshape how tasks are defined (what counts as an "incident"), delegated (who acts), and validated (what counts as "evidence" of correctness). When AI generates inspection recommendations, dispatch schedules, anomaly diagnoses, or access decisions, it redistributes responsibility across people and artifacts. Readiness therefore concerns not whether AI is "accurate," but whether this redistribution remains governable, inspectable, and reversible—especially under model drift, vendor turnover, or contested accountability.

3.2 Limits of Interpretation: Why Vetoability and Evidence Are First-Class Requirements

Eco's limits of interpretation emphasizes a crucial point for operational AI: interpretation is neither arbitrary nor fully determinate; it is constrained by texts, practices, communities, and rules (Eco 1990). Large models expand the space of plausible interpretations—multiple outputs can appear reasonable under the same prompt, yet imply different actions and liabilities. In a physical space, interpretive multiplicity is not merely epistemic; it becomes contractual and ethical. Disputes about "what the model meant" can translate directly into disputes about "who is responsible."

Accordingly, a readiness regime must specify: who may treat an AI output as actionable; what constitutes admissible evidence; how disagreement is resolved; and what happens when AI is withdrawn or challenged. From this lens, we derive two boundary principles:

Interpretive governance: AI outputs must be contestable (vetoable) and traceable through an evidence regime that anchors decisions to accountable human judgment.

Continuity governance: AI must be exitable and replaceable, so meaning and responsibility do not collapse into a vendor-specific black box and the space remains runnable under human takeover.

3.3 Space AI Readiness as a Proto-Institution

Proto-institutions are emergent rule systems that arise through collaboration and can diffuse beyond a single project (Lawrence et al. 2002). We theorize Space AI Readiness as a proto-institution because it codifies a shared rule-set and evidence regime—task semantics, permissions, escalation paths, documentation, and exit conditions—that coordinates multiple actors while reducing decision costs and interpretive conflict.

Crucially, this proto-institutional framing distinguishes readiness from purely technical standards or toolkits. It integrates interpretive governance (who decides, who can veto, what counts as evidence) with continuity governance

(how to switch vendors/models, preserve operational memory, and maintain safe fallback modes). In doing so, Space AI Readiness specifies the institutional preconditions for “AI-enabled spaces” to become governable operational spaces.

4 Conceptual Framework

Space AI Readiness is defined as the stable socio-technical conditions under which a physical space can support AI participation in production, collaboration, and operations, such that AI becomes a sustained operating element rather than a one-off tool. the framework has two axes and one communication ladder.

4.1 Dual-Axis Model

Axis A—Governance/Interpretation: specifies final judgment, veto rights, responsibility chains, and evidence requirements for AI-mediated decisions. Axis B—Operational infrastructure: specifies touchpoints (network, devices, interfaces), task semantics (work orders, inspections, events), data/log exportability, modular interfaces, and takeover procedures.

Table 1. Dual-Axis Model

Dimension	Key question	Observable evidence (examples)
Governance/Interpretation	Are AI outputs vetoable and traceable?	Veto/rollback workflow; role-based authorization; decision logs
Governance/Interpretation	Is responsibility assignable across actors?	RACI for AI-mediated tasks; incident handling records
Operational infrastructure	Can humans take over when AI fails?	Takeover playbooks; drills; minimal runnable mode
Operational infrastructure	Can AI be replaced without collapse?	Modular interfaces; data/log export tests; vendor-switch plan
Operational infrastructure	Is data treated as a first-class deliverable?	Data & licensing matrix; archival policy; audit access

4.2 L0–L5 Readiness Ladder (Communication Device)

Inspired by the communicative value of the autonomous driving L0–L5 ladder, we propose

an L0–L5 readiness ladder for spaces. This is not a claim of equivalence; it is a structured language to express an evolution from traditional delivery to governable, semi-autonomous operation.

Table 2. L0–L5 Readiness Ladder (Communication Device)

Level	Space condition	AI participation	Audit focus
L0 Traditional	Renovation delivered; operations decoupled	None/ad hoc tools	No evidence regime
L1 Digitized	Network & basic touchpoints available	Information display & search	Touchpoint coverage; logging
L2 Taskified	Work orders/inspections/events are task-native	AI assists dispatch & checking	Closed-loop tasks; permissions; trace
L3 Coordinated	Cross-actor standardized processes	AI supports review & retrospection	Veto rights; responsibility chain; drills

L4 Semi-autonomous	Key operations can auto-link within rules	AI runs within constraints; supervise humans	Exit/replacement; incident audit
L5 Adaptive	Continuous optimization with governance	AI strategy iterates under oversight	Long-term controllability & ethics

5. From Concept to Deliverables: an Audit-Ready Implementation Grammar

In the built environment, socio-technical governance becomes actionable only when it is encoded in deliverables, verified through evidence, and enforced through acceptance. Renovation projects and operational handovers are precisely the moments when such encoding can succeed—or fail—because contractual scope, responsibilities, and performance criteria are being fixed. For this reason, we translate Space AI Readiness into an audit-ready implementation grammar: a structured package of documents, artefacts, and tests that can be embedded into procurement processes and EPC (engineering–procurement–construction) contracts.

We use “audit-ready” in a strict sense. A readiness claim is audit-ready when (i) it can be checked against explicit criteria, (ii) it is supported by traceable evidence rather than demonstrations, and (iii) it includes re-testability under drift, vendor change, or degraded conditions. the package below is designed to align renovation delivery with operational continuity by making governance and exit conditions part of what is delivered and accepted.

5.1 Baseline Package (Mandatory)

The baseline package constitutes the minimum viable readiness. It binds renovation and operations into a single verifiable loop by specifying (a) what is delivered, (b) how it is accepted, and (c) what evidence must exist for later dispute resolution, incident review, and vendor replacement.

5.1.1 Readiness Assessment Report

A structured diagnosis that establishes the “as-is” state and the target state. It includes:

a dual-axis assessment (governance/interpretation × operational infrastructure),

an L0–L5 readiness positioning and justification, and a prioritized gap register with risk severity, dependency mapping, and remediation sequencing.

Audit function: provides an explicit baseline for acceptance, procurement scope, and future re-testing.

5.1.2 Task-based Operating Blueprint

A translation layer from policy statements to operations. It decomposes “time rules” and “operating rules” into dispatchable tasks with: task semantics (what counts as completion), inputs/outputs and handoff points, escalation thresholds (when to stop/override), and evidence expectations (what logs, forms, or sensor traces must be recorded).

Audit function: turns governance into operational routines that can be checked and trained.

5.1.3 Outsourcing Work Packages: WBS+RACI+Acceptance Criteria

A procurement-ready decomposition of work so that multi-vendor delivery remains governable. Each work package specifies:

WBS scope (what is in/out),

RACI roles (accountable vs. responsible vs. consulted), and acceptance criteria mapped to test steps and evidence artefacts.

Audit function: prevents “scope ambiguity” and makes responsibility attributable across contractors.

5.1.4 Touchpoint & Data Baseline

A documented inventory of operational touchpoints and data routes that determine whether AI can be governed and exited. It covers: networks and segmentation,

screens/terminals and user touchpoints, access control and identity/permission boundaries,

device interfaces and control APIs,

logging and telemetry requirements, and data export paths, ownership, and retention policies.

Audit function: ensures operational memory is not trapped in a proprietary stack and that evidence remains portable.

5.1.5 AI Agent Configuration & Operating Procedures

A version-controlled operational specification for how AI is allowed to act. It includes:

policy/prompt and configuration versioning (with change control),

permission boundaries and action constraints,

escalation and veto rules (who can suspend/override),

incident handling playbooks, and audit logs and trace templates linking outputs to actions and

responsible humans.

Audit function: makes “AI behavior” inspectable as governance, not as a vendor promise.

5.1.6 Acceptance and Re-testing Plan

A lifecycle acceptance regime rather than a one-off demonstration. It includes:

initial acceptance tests tied to criteria and evidence artefacts,

periodic re-tests for drift and operational change, and explicit exit/takeover drills (suspension of AI, fallback to human operation, vendor/model switch rehearsal).

Audit function: operationalizes continuity and ensures readiness survives turnover and drift.

5.2 Data-and-Licensing Matrix and Exit Tests (Critical Hardening)

Procurement audit failures in high-constraint domains show that capability outsourcing turns into sovereignty outsourcing when data rights and evidence chains are not defined early. the structural lesson for buildings is clear: if logs, configurations, and task histories are not contractually deliverable and exportable, vendor replacement becomes infeasible. Accordingly, readiness requires a data-and-licensing matrix and an exit test (minimal runnable mode+human takeover drill+migration window and checklist).

Table 3. Data-and-Licensing Matrix and Exit Tests (Critical Hardening)

Data object	Owner (default)	License scope	Export format/frequency	Retention	Acceptance evidence
Work-order logs	Owner/Operator	Ops & audit; no external training by default	CSV/JSON daily	≥ 24 months	Export sample+replay check
Device configs	Owner	Maintenance & replacement allowed	Config bundle quarterly	Lifecycle	Restore test+migration drill
Agent policy versions	Owner/Operator (by contract)	Project operations; reuse by agreement	Version package per release	≥ 24 months	Rollback drill+version audit

6. Discussion: Why Readiness Is about Lowering Decision Costs

A persistent managerial temptation is to treat readiness as a bundle of “AI capability” that can be purchased—better models, more sensors, more dashboards, more automation. Our claim is different. Space AI Readiness is not an attribute of a tool; it is a public condition of a space. It reduces decision costs across heterogeneous actors by stabilizing how AI outputs become actionable, contestable, and reversible within everyday operations.

In multi-actor environments, operational failure rarely comes from a single wrong prediction. It emerges when actors cannot quickly answer governance questions under pressure: Who is accountable? Who is authorized to override? What evidence is admissible? What is the rollback path? When these questions are unresolved, AI can increase throughput while simultaneously increasing uncertainty—because speed amplifies the consequences of ambiguity. Readiness lowers this ambiguity by supplying shared task semantics, permission boundaries, and evidence trails that prevent interpretive disputes from escalating into operational paralysis.

This point also clarifies why individual AI use

does not automatically create collective benefit. Staff may adopt personal assistants that raise local productivity, but coordination in physical spaces requires a shared evidence regime and institutionalized fallback: standardized traces, clear veto rights, consistent logging, and rehearsed takeover procedures. Without these collective conditions, AI becomes a new dependency layer: when outputs are contested, drift occurs, or vendors change, the space loses continuity because operational memory is not portable and responsibility is not anchored.

Seen in this light, readiness is closer to institutional infrastructure than to IT deployment. It is the difference between (i) demonstrating an AI feature and (ii) maintaining governable operation under routine turnover, degraded performance, and contractual dispute.

6.1 Implications for Owners and Operators

Investment logic: pay for optionality, not demos. Owners should treat readiness investments as protection of cash flow and operational continuity. the primary return is not “more AI”; it is lower coordination overhead, fewer escalation loops, and preserved service levels when AI is suspended or replaced. Readiness turns AI from a brittle productivity lever into a controllable component of operations.

Procurement logic: contract for evidence and portability.

Procurement should explicitly treat data, logs, configurations, and exportability as first-class deliverables. Contracts can require evidence chains for critical tasks (what was decided, by whom, based on what traces) and mandate export tests to ensure operational memory is not trapped inside a proprietary platform. In practice, this means specifying: logging schemas, retention policies, configuration/version control, ownership of operational datasets, and vendor-neutral interfaces where feasible.

Governance logic: operationalize veto and takeover as routine.

Veto rights and rollback procedures should not live only in policy documents. They should be operationalized through daily permissions, escalation thresholds, and periodic drills—just like fire safety or incident response. Readiness implies that “human final judgment” is not symbolic; it is enacted via enforceable permission boundaries and rehearsed takeover modes.

Operations logic: manage AI as a lifecycle, not a feature.

Spaces should assume drift, staff turnover, and vendor change as normal conditions. Readiness therefore supports lifecycle practices: continuous monitoring of model behavior, change control for prompts/policies, scheduled re-tests, and explicit decommissioning pathways. This reduces the hidden cost of “silent degradation,” where systems remain deployed but are no longer trusted or used.

6.2 Implications for Research

Shifting the research object: from adoption to governable operation.

Much built-environment AI research still evaluates models and applications. Space AI Readiness suggests a complementary agenda: studying how governance and continuity mechanisms determine whether AI participation can persist in operational settings.

Empirical validation through longitudinal, multi-site designs.

Future work can test and refine the dual-axis framework and L0–L5 ladder using longitudinal studies across building types (co-working spaces, campuses, malls, hospitals, factories) and across phases (renovation, commissioning, steady-state operation, vendor transition). Key outcomes include incident rates, escalation frequency,

downtime under AI suspension, and dispute resolution time.

Proto-institution diffusion and the politics of evidence.

If readiness behaves as a proto-institution, it should diffuse through templates—contracts, procurement checklists, acceptance tests, and audit routines. Research can examine how these templates travel, who controls them, and how they reshape labor, accountability, and “what counts as evidence.” This is especially important where AI outputs are used to justify interventions that affect safety, access, or resource allocation.

Measurement and instrumentation of decision costs.

A concrete research contribution would be operational metrics for decision costs: number of handoffs per incident, time-to-veto, time-to-rollback, time-to-reconstruct causality, and the proportion of actions supported by admissible evidence. These metrics can connect governance claims to measurable operational performance without reducing readiness to model accuracy.

7. Conclusion

Space AI Readiness reframes the question of AI in the built environment from whether a space adopts AI to whether AI participation is governable, auditable, and exitable over time. Grounded in Science and Technology Studies and the limits of interpretation, the concept treats AI not as a neutral tool but as an interpretive and institutional mediator whose outputs reshape responsibility, evidence, and authority in physical space. By translating readiness into audit-ready deliverables and acceptance regimes—such as task-based operating blueprints, evidence chains, data-and-licensing matrices, and exit/takeover drills—the framework offers a concrete way to align renovation, everyday operations, and accountability across heterogeneous actors. The broader normative claim is that AI in physical spaces should be designed for continuity and contestability: spaces must remain runnable under human control, decisions must remain open to dispute through evidence, and vendors and models must remain replaceable without erasing operational memory.

References

- [1] Callon M (1986) Some elements of a sociology of translation: domestication of

- the scallops and the fishermen of St Brieuc Bay. In: Law J (ed) *Power, action and belief: a new sociology of knowledge?* Routledge, London, pp 196–233.
- [2] Eco U (1990) *the limits of interpretation*. Indiana University Press, Bloomington
- European Commission (2020) *Smart Readiness Indicator (SRI) for buildings*. Publications Office of the European Union, Luxembourg
- ISO/IEC (2023) *ISO/IEC 42001:2023 Artificial intelligence—Management system*. International Organization for Standardization, Geneva.
- [3] Latour B (2005) *Reassembling the social: an introduction to actor-network-theory*. Oxford University Press, Oxford
- Lawrence TB, Hardy C, Phillips N (2002) *the institutional effects of interorganizational collaboration: the emergence of proto-institutions*. *Acad Manage J* 45(1):281–290.
- [4] NIST (2023) *AI Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology, Gaithersburg
- Suchman L (1987) *Plans and situated actions: the problem of human-machine communication*. Cambridge University Press, Cambridge.