Article

Not peer-reviewed version

Beyond Compliance: Skeptical Intelligence for Digital Twin Governance in Critical Infrastructure

Bechir Ben Daya *, Jean-François Audy, Mohamed Ben Daya

Posted Date: 29 May 2026

doi: 10.20944/preprints202605.2031.v1

Keywords: digital twins; skeptical intelligence; adaptive governance; critical infrastructure; smart city governance; urban digital twins; algorithmic accountability; design science research

Article

Beyond Compliance: Skeptical Intelligence for Digital Twin Governance in Critical Infrastructure

Bechir Ben Daya [1,*], Jean-François Audy [1] and Mohamed Ben Daya [2]

[1] Department of Management, Université du Québec à Trois-Rivières, Canada
[2] College of Engineering Department of Industrial Engineering, American University of Sharjah (AUS), United Arab Emirates
* Correspondence: bechir.ben.daya@uqtr.ca; Tel.: +1 4189314079

Highlights

What are the main findings?

- Current governance frameworks for Digital Twins in critical infrastructure—including smart city urban services, transportation, energy, and healthcare—remain predominantly procedural, lacking institutionalized cognitive capacities for continuous validation, proactive harm detection, and structured multi-actor deliberation.
- The Skeptical Intelligence Framework (SIF) proposes three interdependent cognitive functions—Validation (F1), Detection (F2), and Deliberation (F3)—supported by five operational principles, concrete governance artifacts, and distributed accountability roles, developed through Design Science Research and grounded in a systematic scoping review of 78 sources.

What are the implications of the main findings?

- Organizations deploying Digital Twins in multi-actor critical infrastructure—from port logistics to smart city urban management—require a cognitive governance layer that institutionalizes disciplined skepticism, epistemic humility, and structured contestation beyond compliance-oriented approaches.
- The SIF complements existing procedural frameworks (NIST AI RMF, EU AI Act) and recent stage-gate governance models for municipal AI by providing the epistemic and deliberative architecture that these frameworks leave unspecified.

Abstract

Digital twins are becoming vital decision-making infrastructures across critical infrastructure sectors such as smart city urban services, transportation, energy, and healthcare. As digital twins become autonomous and gain real-time intervention capabilities, their governance becomes increasingly essential. Yet existing governance mechanisms remain largely procedural: they emphasize compliance without operationalizing the cognitive practices required to question assumptions, detect algorithmic harms, or support legitimate multi-actor deliberation. Drawing on a systematic scoping review, this study synthesizes the literature on digital twin autonomy, algorithmic risks, epistemic foundations, and governance mechanisms. The review reveals a fundamental gap: current governance mechanisms lack institutionalized cognitive capacities for continuous validation, proactive detection of emerging harms, and structured multi-stakeholder deliberation. To address this gap, the paper proposes the skeptical intelligence framework, developed through design science research. The framework integrates three cognitive functions: validation, detection, and deliberation supported by operational principles, governance artifacts, and distributed accountability roles. The framework advances digital twin governance beyond compliance toward a model rooted in critical epistemology, reflexivity, transparency, and democratic legitimacy. It also lays the groundwork for validation of the proposed framework through implementation across multi-actor digital twin infrastructure contexts—including smart city governance, port logistics, and energy networks—where DT-mediated decisions redistribute opportunities and risks across heterogeneous stakeholders. Empirical validation in an operational setting is planned as the next phase of this research.

Keywords: digital twins; skeptical intelligence; adaptive governance; critical infrastructure; smart city governance; urban digital twins; algorithmic accountability; design science research

---

1. Introduction

Digital twins (DTs) began as simulation tools — static representations that engineers consulted to understand system behavior. That era is ending. Contemporary DTs integrate heterogeneous real-time data streams, update system representations continuously, and execute forecasting and optimization at a pace that exceeds human cognitive capacity [1,2]. The next generation goes further: cognitive DTs adapt their own objectives, federated architectures coordinate decisions across institutional boundaries without centralized human oversight, and autonomous twins in ports, energy grids, and urban systems increasingly intervene directly in operational outcomes rather than informing the humans who do [3–6]. This trajectory follows a pattern familiar from earlier technological transitions. In algorithmic trading, autonomous vehicles, and smart grid management, technical capability consistently outpaced the institutional arrangements designed to govern it. DTs in critical infrastructure are following the same arc, and the governance lag is already visible. Contemporary smart cities are increasingly understood not as municipal digitalization projects but as interconnected ecosystems of critical infrastructures—ports, energy networks, mobility systems, healthcare, urban services—whose operational coordination increasingly depends on DT-mediated decision architectures. Governing these systems therefore requires frameworks that operate across infrastructure boundaries, not within single sectors. Systems that once structured choices are now making them, in real time, across organizations that do not share the same incentives or accountability arrangements [7]. The assumptions, pipelines, proxy variables, and causal logics driving those decisions frequently remain difficult to inspect, even for well-positioned stakeholders [8,9]. When the system is hard to interrogate, safety is only part of the concern. Legitimacy becomes fragile too, particularly in multi-actor environments where the burdens of DT-mediated decisions fall unevenly and where the ability to contest decisions is itself unequally distributed [10–12].

A further complication is that DT "performance" can look impressive while harm accumulates quietly. The broader algorithmic governance literature has shown how systems can deliver strong aggregate metrics and still produce systematic disadvantage for particular groups, especially when evaluation is not disaggregated and when proxies embed historical or structural bias [13–15]. In multi-actor critical infrastructure contexts characterized by institutional fragmentation and asymmetric power relations, this is not an abstract fairness debate. Disadvantages repeat, compound, and gradually reshape participation and viability for smaller, less resourced, or less visible actors. In parallel, the justification burden rises. Explanations may help, but high-stakes decision systems increasingly require justifiability and contestability: reasons that can be challenged, revised, and, when necessary, overridden through processes that are accountable [16].

This paper emphasizes a specific governance risk that follows from these dynamics: capability erosion. A DT-enabled system may offer formal access to services, consultation channels, or appeals, while still undermining the substantive freedoms needed to use those mechanisms meaningfully [17,18]. In practical terms, a system can widen digital interfaces and narrow real agency at the same time, particularly for stakeholders with lower institutional access, lower digital literacy, or fewer resources to contest decisions [19].

Existing governance frameworks are indispensable, yet they often remain incomplete for addressing capability erosion in multi-actor DT environments. Risk management standards and regulatory regimes provide structural safeguards around documentation, risk categorization, and

 doi:10.20944/preprints202605.2031.v1

compliance boundaries [20[1], 21[2], 22[3]]. Still, the literature on artificial intelligence (AI) governance repeatedly indicates an operationalization gap: principles and procedures proliferate, while the organizational routines that make contestation, learning, and revision real are less consistently institutionalized [23]. Accountability faces a similar challenge. There is a difference between naming a responsible party and building an accountability system that can trace decisions, requires justification, allocates responsibility across the DT lifecycle, and imposes consequences when governance fails [7].

The central argument of this study is that DT governance in multi-actor critical infrastructure systems requires a cognitive layer. This layer institutionalizes vigilance that keeps assumptions explicit, surfaces emerging distortions before they harden into structural exclusion, and enables deliberation that can actually reshape objectives and constraints. To address this need, the paper proposes Skeptical Intelligence (SI) as an integrated organizational capability. SI is not "skepticism" that promotes resistance or delay. It is skepticism that promotes disciplined governance, expressed through three mutually reinforcing functions (Validation, Detection, Deliberation), supported by operational principles, and instantiated in concrete governance artifacts and accountable roles.

This work is carried out in two steps. First, a systematic scoping review synthesizes four convergent strands of literature: DTs as policy machines, documented algorithmic harms, normative foundations grounded in epistemology and human capabilities, and existing AI and DT governance frameworks. Second, the study applies Design Science Research (DSR) to develop the Skeptical Intelligence Framework (SIF) as a prescriptive governance artifact, explicitly designed to bridge the gap identified in the review.

This paper advances DT governance by reframing it as a question of epistemic justice. That is who gets to make authoritative knowledge claims, who can contest those claims, and whose uncertainties are taken seriously in DT-mediated decision environments. Building on this view, the paper introduces SI as a necessary organizational capability for governing DTs. SI supports continuous validation of DT outputs, early identification of latent and unevenly distributed harms, and structured deliberation that goes beyond procedural compliance. The paper contributes in three ways: (i) theoretically, by grounding DT governance in critical epistemology and capabilities ethics; (ii) methodologically, by demonstrating how DSR can translate normative principles into implementable governance mechanisms; and (iii) practically, by proposing adaptable governance artifacts and accountable roles that organizations can tailor to different multi-actor critical infrastructural contexts.

The remainder of the article is structured as follows. Section 2 presents the research methodology. Section 3 synthesizes the theoretical and empirical foundations of the study and identifies the central governance gap. Section 4 introduces the SIF. Section 5 discusses implications, limitations, and directions for future research. Finally, Section 6 concludes the paper.

2. Methodology

This study adopts a sequential research design combining systematic scoping review with DSR to develop a governance framework for DTs in critical multi-actor infrastructure contexts. The scoping review maps and consolidates interdisciplinary evidence across technical, socio-technical, and normative literatures to uncover convergent governance gaps that reflect a broader structural deficit. These gaps are then translated into explicit design requirements, which the DSR phase operationalizes as a prescriptive governance artifact (the SIF). Section 2.1 presents research questions; Section 2.2 details the scoping review approach, architecture, and execution; Section 2.3 describes the DSR framework development process.

[1] U.S. National Institute of Standards and Technology (NIST)

[2] European Union Artificial Intelligence Act (EU AI Act)

[3] General Data Protection Regulation (GDPR)

 doi:10.20944/preprints202605.2031.v1

2.1. Research Questions and Research Design

Consistent with Joanna Briggs Institute (JBI) guidance for scoping reviews [24], the research questions (RQs) were formulated to remain intentionally broad and exploratory, aligning concept (DT-mediated governance), context (multi-actor critical infrastructure), and affected actors. Four RQs structure both the literature synthesis (Section 3) and the subsequent design effort (Section 4):

RQ1: What theoretical and governance frameworks currently exist for understanding DTs in critical infrastructure contexts, and what challenges do they address or leave unresolved?
RQ2: What documented cases of algorithmic harm, bias, and systemic exclusion reveal governance vulnerabilities in systems where DTs mediate coordination and decision-making?
RQ3: What normative foundations, grounded in epistemology, human capabilities, and democratic legitimacy, are necessary for responsible and legitimate DT governance in multi-actor environments?
RQ4: What critical gaps persist between existing AI/DT governance frameworks and the cognitive and deliberative capacities required for transparent, inclusive, and substantively legitimate governance?

2.2. Systematic Scoping Review

2.2.1. Review Approach and Reporting Standard

A scoping review approach was selected because DT governance research is interdisciplinary and methodologically heterogeneous, spanning engineering, information systems, public administration, ethics, and socio-technical governance. In such contexts, a scoping review is appropriate for mapping concepts, identifying patterns and tensions, and identifying governance gaps rather than pursuing statistical aggregation. The review follows [24] guidance and is reported using PRISMA-ScR as a transparency framework, designed to serve dual purposes: (1) map current DT governance scholarship across disciplines, and (2) identify convergent gaps that will then inform framework design requirements.

2.2.2. Two-Tier Literature Architecture

To address the dual requirement for systematic empirical evidence and normative theoretical grounding, this study employs a two-tier literature architecture appropriate for DSR [25]. Tier 2 sources were not used to derive empirical claims, but to establish normative and epistemic evaluative criteria against which DT governance approaches are assessed.

Tier 1 (Core Empirical Corpus, n=33) was identified through systematic scoping review following PRISMA-ScR protocols, providing empirical evidence on DT implementations, documented algorithmic harms, and existing governance approaches. Database searches with explicit inclusion/exclusion criteria ensure reproducibility and minimize selection bias.

Tier 2 (Normative Foundations and Policy Instruments, n=45) was identified through purposive selection based on four explicit inclusion rules: (1) canonical theoretical texts—seminal works (500+ citations or field-defining status) establishing foundational concepts in epistemology, capabilities ethics, or democratic theory; (2) authoritative policy instruments—official regulatory frameworks and standards from recognized governance bodies (EU, NIST, ISO); (3) recent high-impact publications—post-search publications in Q1 journals addressing gaps identified in core corpus analysis; (4) cross-disciplinary bridging literature—methodological guidance enabling synthesis across disciplines.

This approach is consistent with scoping review methodology [24] and DSR principles regarding kernel theories grounding artifact design [25,26]. Systematic database searches are inappropriate for Tier 2 literature because: (1) canonical works often predate digital indexing; (2) policy instruments exist outside scholarly databases; and (3) normative philosophy literature is inconsistently indexed in technical databases. The distinction between tiers reflects epistemological differences rather than quality hierarchies. Tier 1 sources document *what is* (empirical reality); Tier 2 sources establish *what ought to be* (normative criteria) and *how to evaluate* (conceptual frameworks).

2.2.3. Cluster Derivation and Thematic Architecture

To operationalize the two-tier architecture within a tractable search strategy, the scoping review employs ten *a priori* thematic clusters (C1–C10) derived from three foundational governance requirements reflected in the RQs and literature synthesis: (1) epistemic adequacy (how DTs represent reality and make claims), (2) harm prevention and detection (how DT-mediated decisions affect stakeholders), and (3) democratic legitimacy (how governance processes enable contestation and learning).

Each requirement necessitates coverage across multiple disciplinary domains. For instance, epistemic adequacy cannot be addressed through technical literature alone as it requires engagement with philosophy of science (C1), adaptive learning theory (C2), and validation methodologies (C7). Similarly, harm detection requires empirical evidence of algorithmic failures (C3, C4, C5), normative criteria for assessing harm (C10), and institutional mechanisms for redress (C6, C8, C9).

The 10 clusters structure this multi-dimensional coverage:

- **Normative-Epistemic Foundations** (C1, C2, C10): Establish criteria for what counts as adequate knowledge, legitimate governance, and meaningful agency
- **Empirical Evidence Base** (C3, C4, C5, C9): Document how algorithmic systems produce harms, biases, and exclusions in practice
- **Technical-Operational Core** (C7, C8): Characterize DT architectures and operational deployment patterns
- **Institutional-Regulatory Context** (C6): Map existing governance frameworks and their limitations

This architecture ensures the synthesis can answer all four RQs by providing both empirical grounding (what problems exist) and normative foundations (what standards should govern evaluation), consistent with scoping review guidance for complex socio-technical systems [24,27]. As shown in Table 1, clusters differ in their evidentiary role: some provide empirical grounding by documenting governance failures (C3–C5, C9), while others contribute normative foundations for evaluating governance adequacy (C1, C2, C10).

Table 1. Cluster architecture showing theoretical derivation, RQ alignment, Section 3 mapping, and evidentiary roles.

Cluster	Cluster Label	Primary RQ(s)	Section 3 Subsection	Evidence Role (Hybrid)
C1	Epistemology & scientific methodology	RQ3, RQ4	3.3 Normative foundations	Normative–epistemic foundation (model validity, falsifiability, uncertainty)
C2	Adaptive governance & learning	RQ3, RQ4	3.3 Normative foundations	Normative–institutional foundation (learning, reflexivity, adaptation)
C3	Algorithmic bias & fairness	RQ2	3.2 Documented harm & exclusion	Empirical core evidence (documented harms and exclusion mechanisms)
C4	Generative AI & cognitive processes	RQ2, RQ4	3.2 Documented harm & exclusion	Empirical supporting evidence (cognitive and perception-related impacts)
C5	Human–automation interaction & trust	RQ2, RQ4	3.2 Documented harm & exclusion	Empirical supporting evidence (automation bias, over-reliance, trust dynamics)
C6	AI regulation & governance frameworks	RQ1, RQ4	3.4 Existing governance frameworks	Institutional baseline (procedural governance and compliance limits)
C7	DTs – core technologies	RQ1	3.1 DTs as policy machines	Technical–functional core (architectures enabling real-time intervention)
C8	DTs – governance & operations	RQ1, RQ4	3.1 & 3.4	Operational governance core (DTs as decision infrastructures)
C9	DT inclusivity & citizen participation	RQ2, RQ3	3.2 & 3.3	Normative–empirical bridge (participation limits and legitimacy gaps)

 doi:10.20944/preprints202605.2031.v1

C10	Human development & agency (capabilities)	RQ3	3.3 Normative foundations	Normative legitimacy foundation (capabilities, substantive freedom)

2.2.4. Search Strategy and Data Sources

Searches were conducted July-November 2025 in Scopus and SpringerLink, selected for complementary disciplinary coverage: Scopus provides broad multidisciplinary indexing including engineering and computer science; SpringerLink offers extensive coverage of governance, policy, and social sciences. In Scopus, searches targeted TITLE-ABS-KEY fields, retrieving top-200 relevance-ranked records per cluster (n=2,000). In SpringerLink, searches retrieved top-100 per cluster (n=1,000). These cutoffs were determined through pilot testing (diminishing marginal returns beyond these thresholds). Searches were restricted to English-language journal articles and scholarly books, with no temporal limitation as normative foundations often derive from canonical pre-2000 works while DT governance is an emerging field. Full Boolean search strings are in Appendix A. The combined search yielded 3,000 records.

2.2.5. Screening and Inclusion Logic

After deduplication (269 duplicates removed), 2,731 records underwent three-stage screening consistent with JBI guidance. Stage 1: All records underwent structured relevance assessment using a Relevance-Alignment-Knowledge (RAK) scoring rubric (scale 1–10) evaluating alignment with the review's conceptual scope, derived from the RQs (epistemology, governance, socio-technical systems, human agency). The rubric assessed conceptual relevance to DT governance (0–4 points), methodological contribution (0–3 points), and RQ alignment (0–3 points). Records scoring RAK ≤7 were excluded (n=2,570), as preliminary testing indicated scores of 8+ reliably identified sources with substantive engagement with governance concepts, leaving 160 records. Stage 2: Title and abstract screening excluded out-of-scope or predominantly descriptive publications (97 exclusions, 63 retained for full-text review). Stage 3: Full-text eligibility assessment excluded 30 records due to limited analytical contribution, conceptual redundancy, or insufficient relevance, ensuring included sources made substantive contributions to understanding DT governance issues and limitations rather than merely describing technical implementations. In total, 33 studies were included from database searching.

2.2.6. Synthesis Method

The 33 articles identified through database searching formed the core analytical corpus, examined to identify how DTs, including their AI-based components, are conceptualized and problematized in relation to governance, decision-making authority, accountability, and operational use.

In addition to the core corpus, 45 Tier 2 sources were included following the purposive selection rules detailed in Section 2.2.2: canonical theoretical texts (n=12), authoritative policy instruments (n=8), recent high-impact publications (n=18), and cross-disciplinary methodological literature (n=7). These sources provide the epistemological and normative grounding necessary for establishing evaluative criteria.

The full corpus was examined through a narrative, concept-oriented synthesis. Sources were analyzed for their conceptualization of DTs as governance instruments, documented governance challenges, normative criteria for legitimate governance, and critiques of existing frameworks. This analysis directly informed the four-strand structure of Section 3, with patterns emerging through iterative engagement with the literature. This synthesis approach prioritizes conceptual coherence and theoretical integration over systematic data extraction, enabling identification of convergent patterns across methodologically heterogeneous literature that are appropriate for exploratory scoping reviews [24]. The selection and inclusion process is summarized in Figure 1.

 doi:10.20944/preprints202605.2031.v1

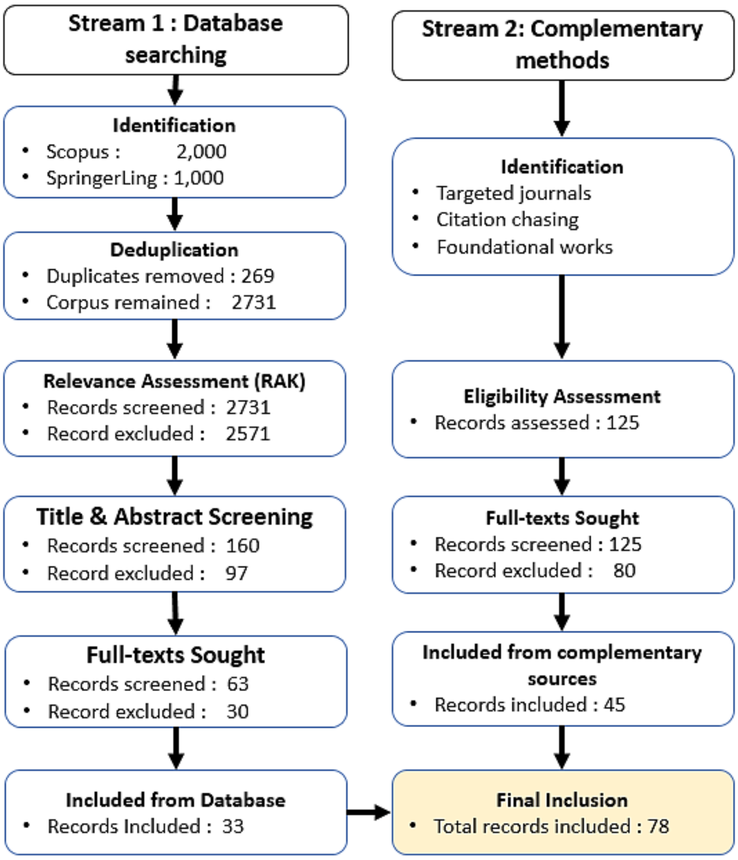

Figure 1. PRISMA-ScR flow diagram of the two-stream scoping review process.

2.3. Design Science Research

The scoping review provides the evidential foundation for framework development but does not by itself produce governance solutions. A second methodological step is therefore required: translating the governance gaps identified in the literature into an institutional design that organizations can implement. This study follows an institutional design reasoning approach, in which empirical evidence and normative theory are used to derive explicit governance requirements, which in turn inform the architecture of a prescriptive framework [25,28]. This approach is appropriate because the study's contribution is not limited to diagnosing governance problems. It specifies the functions, principles, roles, and artifacts that organizations need to govern autonomous DTs legitimately and accountably in multi-actor infrastructure contexts.

The logic is explicitly traceable. The synthesis in Section 3 consolidates recurrent governance failures, epistemic risks, and legitimacy deficits into three design requirements: continuous epistemic validation, proactive harm detection, and structured deliberation with revision authority. These requirements directly inform the three governance functions of the SIF developed in Section 4. Each framework component is linked to the requirement it addresses and to the evidence strands that motivate it. The framework is demonstrated through a port infrastructure application showing how its functions would operate in practice. Formal empirical evaluation in a live infrastructure environment is the logical next step and is planned as subsequent research.

3. Literature Review

This scoping review synthesizes multidisciplinary evidence on DT and AI governance in critical infrastructures, following the two-tier literature architecture (Section 2.2.2) and cluster organization (Section 2.2.3). The review addresses the four RQs through interconnected strands: Section 3.1 examines DTs as institutional decision systems in critical infrastructures (C7-C8, RQ1). Section 3.2 documents algorithmic harms and systemic exclusion (C3-C5, C9, RQ2). Section 3.3 establishes normative foundations from epistemology, capabilities, and adaptive governance (C1-C2, C10, RQ3). Section 3.4 examines existing frameworks (C6, RQ4). Section 3.5 synthesizes convergent design requirements informing framework development in Section 4.

3.1. Digital Twins as Institutional Decision Systems in Critical Infrastructures

DTs have evolved from passive simulation into dynamic systems integrating real-time data, supporting predictive inference, and intervening in operational decisions [1,2]. Contemporary implementations span from decision-support to cognitive DTs capable of adaptive optimization [3,29], with federated architectures reflecting institutional fragmentation across multiple actors in critical infrastructure contexts [1,5]. DTs deployed in urban infrastructure integrate 3D models, geospatial data, sensors, and social sensing for planning, climate adaptation, and operational management [19,30–32].

As DTs embed in operational routines, they function less as dashboards and more as allocators and coordinators. Even when formal authority remains human, DTs structure choice spaces, redistribute constraints, and shape collective trade-offs [4,33]. In urban and infrastructure contexts, DTs operate as decision-shaping platforms influencing planning and resource allocation without being formally recognized as decision-makers [34–36]. A DT supporting exploratory planning differs fundamentally from one managing real-time allocations, where system outputs translate directly into constraints on affected actors, thereby raising accountability questions even when decision authority remains formally human [7].

DT-enabled systems increasingly function as epistemic infrastructures whose outputs shape expectations, behavior, and responsibility. [37] show that algorithmically produced representations do not merely inform decision-makers but actively reorganize strategic reasoning, accountability relations, and democratic agency. In critical infrastructures, DT representations exert influence through their capacity to stabilize particular interpretations of system behavior while sidelining alternative readings. Although these works do not examine DTs directly, they reveal a recurring structural pattern: when algorithmic representations acquire operational centrality and guide collective action, governance gaps emerge unless legitimacy, accountability, and contestability are institutionally articulated and maintained.

This evolution renders legitimacy operational. Hirschman's voice and exit concepts clarify that affected actors require meaningful channels to contest decisions and withdraw from arrangements that no longer serve them [10]. Bovens' accountability framework requires governance to institutionalize obligations to explain, justify, and face consequences [11]. In high-stakes contexts, explainability alone is insufficient without genuine contestation and trade-off justification [12,16]. Legal analyses show decision-support systems can engage rights even when authority remains formally human, requiring explicit accountability clarification [38,39]. Technically, responsible judgment presupposes epistemic robustness: verification, validation, and explicit uncertainty treatment become prerequisites for trustworthy oversight [40].

These dynamics become particularly visible in port-based DT systems managing berth allocation. Allocation decisions shape vessel schedules, fuel consumption, crew logistics, and supply-chain timing. Delays absorbable by large carriers can be existential for smaller operators. Similar patterns occur in urban mobility, energy networks, and emergency response where DT-mediated prioritization redistributes costs, risks, and benefits unevenly. In such settings, DT-mediated resource sequencing such as equipment deployment, labor scheduling, access prioritization, affects employment stability, safety exposure, and operational resilience. Governance must protect voice, exit, and accountability as substantive mechanisms enabling stakeholders to understand what is optimized, contest trade-offs, and influence redesign [41]. Critical analyses emphasize that such systems are not politically neutral: when organized around proprietary platforms, collective control weakens unless governance addresses power concentration [42].

Technical integration typically advances faster than the institutional arrangements required to manage responsibility, authority, and conflict across heterogeneous actors. This asymmetry helps

 doi:10.20944/preprints202605.2031.v1

explain why technically sophisticated DTs often remain fragile as decision infrastructures once they are embedded in real operational settings.

While AI-enabled capabilities within DTs advance rapidly, governance arrangements for responsible, accountable, and ethical use remain comparatively underdeveloped, particularly in multi-actor infrastructure contexts characterized by competing interests and asymmetric power [34]. Recent studies have begun to examine the governance implications of DT deployment in complex urban and infrastructural environments. For example, [6] propose a maturity model approach to assess governance readiness for city DT technologies, emphasizing the importance of institutional arrangements that support transparency, stakeholder participation, and accountability. Empirical studies further indicate that institutions often rely heavily on algorithmic outputs without establishing durable mechanisms for validation, challenge, and revision, while authority tends to concentrate when explicit power-sharing arrangements are absent [43]. Together, these dynamics help explain why organizations struggle to govern advanced DT systems in ways that are both operationally effective and legitimately accountable across all actors.

Similarly, an exploratory study of 99 urban DTs across three continents found that more than half had not reached series operation, and that governance capacity and authority involvement—rather than technical maturity—were the strongest predictors of scalability [44].

3.2. Documented Algorithmic Harm and Systemic Exclusion

Algorithmic systems can appear neutral while producing systematic harm. Gender Shades demonstrated stark intersectional disparities in facial recognition with errors distributed unequally across demographics [13]. Such disparities are not merely technical failures; in consequential settings, they translate into accountability gaps and liability risks [45].

For DTs in critical infrastructures, harms remain invisible when evaluation relies on aggregate metrics, stakeholder groups are not disaggregated analytically, or decision logic remains opaque. Opacity constitutes a control problem. When affected parties cannot understand, scrutinize, or contest decision bases, disadvantage becomes difficult to demonstrate and harder to correct [8,46,47].

Bias exists at multiple lifecycle stages: data collection, sampling, measurement, modeling choices, evaluation, deployment feedback loops [15]. Legal scholarship shows discrimination may arise through apparently neutral technical proxies without explicit intent [14]. In city-scale DTs, practitioners may address bias pragmatically rather than institutionalizing systematic countermeasures [48]. Empirical evidence demonstrates exclusion persists even when fairness is a stated objective. Disparities often are not monitored proactively, and corrective action relies on ad hoc complaints rather than embedded detection and redress [49,50].

Studies examining participatory arrangements in DT initiatives reinforce this pattern. [51] find that participatory mechanisms in smart-city DT initiatives remain largely consultative. Citizens may express concerns, but systematic mechanisms ensuring that such input influences decisions remain weak. Therefore, formal participation can coexist with substantive exclusion.

For example, a port DT allocating vessels using size, processing time, capacity utilization, and equipment availability appears operationally neutral. Yet smaller operators may be systematically deprioritized because vessels do not align with rewarded performance profiles. When ports report only aggregate improvements such as overall utilization and average turnaround, effects remain concealed. Detecting harms requires disaggregated outcome analysis and institutional pathways for stakeholders to contest objectives and request revision. Contestation and redress are not optional but core conditions for responsible governance where algorithmic decisions shape opportunities and outcomes [41,52].

3.3. Normative Foundations: Epistemology, Capabilities, and Adaptive Governance

What counts as legitimate governance when DTs represent complex systems imperfectly and act on those representations? Capability theory shifts attention from formal access to substantive freedom. Sen and Nussbaum argue justice should be assessed by what people can actually do and

doi:10.20944/preprints202605.2031.v1

be, not merely whether opportunities exist on paper [17,18]. In DT governance, "participation" can be offered superficially while influence remains concentrated. Recent work emphasizes governance oriented toward optimization alone may reinforce structural inequalities unless human capabilities, distributive justice, and institutional capacity are treated as core objectives [53]. [54] show narrow focus on aggregate utility sidelines distributional concerns and can legitimize substantial harm to minorities. DT governance must validate not only whether aggregate performance improves but whether systems preserve or erode human capabilities [43].

Epistemology sharpens implications. DT outputs are claims built on assumptions, boundaries, proxies, and value-laden objectives. Popper's falsificationism foregrounds disciplined conjecture testing [55]. Kuhn, Lakatos, and Feyerabend remind us that models drift, anomalies accumulate, and progress often requires reframing [56–58]. Contemporary research emphasizes digital models do not merely describe reality but actively shape it, making epistemic assumptions central to legitimacy [59]. This highlights the need for governance to interrogate the representational frames embedded in DT models themselves [34].

Transparency is insufficient for legitimacy unless it supports justification, contestation, and assumption revision [12,16]. Organizational learning and adaptive governance translate these insights institutionally. Argyris and Schön's double-loop learning clarifies that governance must not only optimize within existing objectives but question and revise those objectives [60]. Adaptive governance research emphasizes institutions integrating diverse knowledge, managing uncertainty, and enabling collective learning and revision over time [61,62]. Recent work frames legitimacy as an ongoing, reflexive process requiring epistemic humility and revision capacity under uncertainty [63–65].

Empirical work supports this: [66] demonstrate that structured collective reasoning can significantly alter participants' understanding, trust, and evaluative judgments regarding algorithmic systems. Deliberation enhances legitimacy and informed contestation only when participants have genuine epistemic access, meaningful opportunities for reasoning, and real capacity to influence outcomes. These conditions remain rarely satisfied in current DT governance mechanisms.

3.4. Existing Governance Frameworks: Procedural Strengths and Critical Gaps

Existing AI governance frameworks offer procedural robustness but lack mechanisms for sustained epistemic interrogation, proactive harm detection, and consequential deliberation. NIST AI Risk Management Framework [20] provides structured risk identification, assessment, and mitigation, offering shared vocabulary and procedures. However, it does not specify how institutions should continuously interrogate model assumptions, test validity under changing conditions, or confront divergences between stated objectives and observed outcomes [50].

The EU AI Act [21] and the General Data Protection Regulation (GDPR) [22] establish requirements for data protection, risk categorization, documentation, and compliance. These instruments primarily specify what must be classified, documented, and audited, while offering limited guidance on how institutions should sustain the ongoing interpretive, monitoring, and intervention capacities required to govern DT systems effectively in real time. Existing governance frameworks thus emphasize formal procedures, risk tiers, compliance checklists, and controls, but provide little support for reassessing underlying assumptions, detecting unintended effects, or intervening when adaptive systems drift [34,67]. Empirical studies further show that even when audit mechanisms and oversight structures are in place, governance often remains performative, with organizations complying procedurally without developing the capacity to question optimization goals or revise decision criteria [47,68]. Recent reviews of urban DT deployments similarly highlight the growing integration of participatory interfaces, sensing infrastructures, and simulation tools within governance arrangements, while leaving largely unexamined whether such inputs substantively influence decision-making or redistribute authority [69].

 doi:10.20944/preprints202605.2031.v1

Advanced proposals for lifecycle-oriented and adaptive governance frameworks explicitly recognize these shortcomings [70]. While they introduce post-deployment monitoring, dynamic risk categorization, and continuous oversight, they also demonstrate that existing standards remain insufficient unless institutions possess the capacity to interpret signals, contest system behavior, and revise governing assumptions over time [71,72].

Recent contributions within smart city and urban AI governance have begun to address this operationalization challenge from complementary angles. [73] propose a stage-gate framework that sequences municipal AI governance decisions across ten procedural checkpoints, explicitly acknowledging that the framework has not been empirically tested and that its effectiveness depends on organizational culture and leadership commitment. [74] develop a social DT framework integrating participatory mechanisms and bottom-up governance roles for urban contexts. These contributions advance the procedural and participatory dimensions of DT governance, respectively. However, neither specifies the epistemic capacities—continuous assumption interrogation, proactive disaggregated harm detection, and structured deliberation with revision authority—that would make such governance substantive rather than performative.

This persistent "principles-to-practice" gap illustrates how organizations may endorse ethical guidelines without developing the tools, incentives, or routines required to enact them operationally [23]. Ecosystem-oriented approaches therefore stress that responsible governance cannot be achieved through static regulatory frameworks alone. They require coordinated institutional roles, socio-technical balance, and continuous adaptation across organizational boundaries [75,76]. Existing governance frameworks thus offer procedural strength, yet remain ill-equipped for the cognitive, institutional, and deliberative work required to govern adaptive DT systems [7]. Table 2 provides a structured comparison of these frameworks across six governance dimensions, making this gap explicit and setting the baseline against which the SIF is positioned in Section 4.

3.5. Synthesis: Three Design Requirements for DT Governance

Synthesizing the four strands yields three convergent design requirements for governing DTs in multi-actor critical infrastructures contexts (Figure 2).

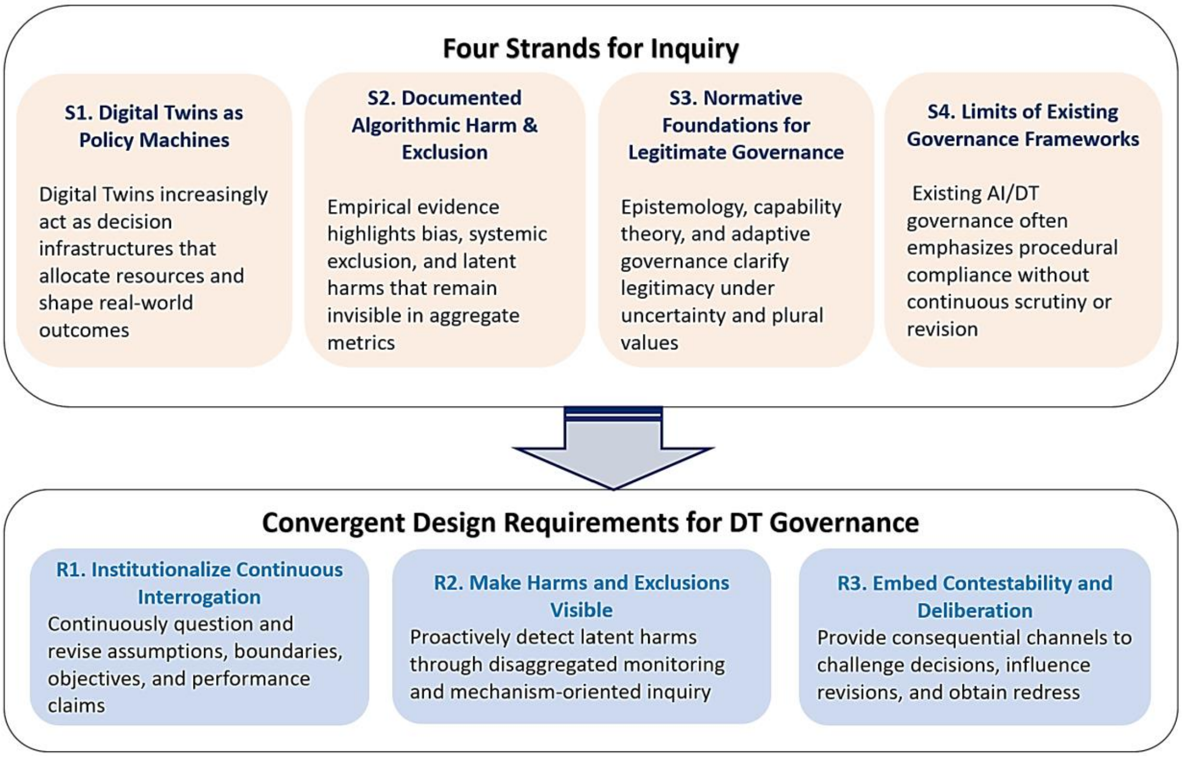

Figure 2. Convergent synthesis of literature strands and resulting design requirements for DT governance.

 doi:10.20944/preprints202605.2031.v1

First, DT governance must institutionalize continuous interrogation of assumptions, boundaries, and objectives**.** As DTs increasingly function as decision-shaping systems (Section 3.1), their outputs acquire epistemic authority that can harden through routine use. Epistemological analyses demonstrate DT models embed value-laden assumptions, proxies, and simplifications that inevitably drift as systems, data, and contexts evolve (Section 3.3). Without explicit mechanisms for verification, validation, uncertainty management, and assumption testing, DTs risk becoming authoritative yet epistemically fragile. Governance must ensure DT representations remain revisable under changing conditions rather than stabilized through procedural compliance alone [55,60].

Second, DT governance must include proactive mechanisms for detecting latent harms and systemic exclusion. Algorithmic bias literature shows optimization-driven systems can produce unequal and exclusionary outcomes while appearing neutral at aggregate level (Section 3.2). Effective governance requires institutionalized detection capacities that surface disparities early, trace how harms emerge through modeling and deployment choices, and enable corrective intervention before disadvantages entrench [8,14,15,47].

Third, DT governance must institutionalize meaningful deliberation, contestability, and revision authority. As DTs redistribute constraints and opportunities across heterogeneous actors (Section 3.1), legitimacy depends not only on transparency or formal participation but on stakeholders' substantive ability to challenge decisions, demand justification, and influence system redesign. Normative scholarship emphasizes voice without consequence risks becoming consultative rather than democratic, particularly with power asymmetries (Section 3.3). Governance must provide structured pathways for contestation, justification, and redress, ensuring deliberation can reshape objectives, metrics, and trade-offs [11,12,41,52].

These three requirements delineate a persistent governance gap. Existing frameworks provide procedural scaffolding yet do not equip institutions with epistemic, organizational, and deliberative capacities required to govern DTs as adaptive and consequential decision infrastructures [50]. These requirements form the conceptual foundation for the SIF developed in Section 4.

4. Skeptical Intelligence Framework

Section 3 identifies a clear governance impasse: as DTs increasingly operate as decision-shaping systems, existing governance approaches remain predominantly procedural and insufficient to sustain continuous scrutiny, early harm detection, and consequential contestation. The issue is therefore not whether such capacities are needed, but how organizations can institutionalize them as effective governance practices.

The SIF responds to this gap by translating the convergent requirements identified in the literature into an integrated governance architecture, developed through DSR. Rather than treating DT outputs as authoritative signals to be managed through compliance mechanisms, the framework explicitly conceptualizes them as fallible, revisable claims whose assumptions, effects, and distributive consequences must remain open to interrogation over time.

The SIF is structured around three interdependent cognitive functions: Validation (F1), Detection (F2), and Deliberation (F3), supported by five operational principles (P1–P5) and instantiated through concrete governance artifacts and accountability roles. Its purpose is not to extend existing risk management checklists, but to institutionalize structured skepticism as a governance capability for DT systems operating under uncertainty and plural interests.

4.1. From Design Requirements to Framework Architecture

The design logic follows directly from the synthesis in Section 3. If DTs operate as policy machines, then legitimacy must be actively produced and maintained through governance processes that enable understanding, justification, and contestation [12,16]. If algorithmic harms can remain hidden in aggregate metrics, then governance needs proactive detection capable of surfacing distortions before they become entrenched [14,47]. If opacity undermines control, then governance must treat interpretability and contestability as operational requirements, not optional transparency ideals [8]. Finally, if learning and revision are essential, then governance must create feedback pathways that support double-loop learning: revision of objectives, metrics, and assumptions, not only recalibration of parameters [55,60].

On this basis, the SI Framework operationalizes three requirements as an integrated architecture with four linked layers:

- Cognitive functions (F1–F3) specify what governance must continuously do;
- Operational principles (P1–P5) specify how those functions are enacted consistently;
- Governance artifacts specify what durable records and traces must exist to support scrutiny and learning;
- Accountability roles specify who has mandate to trigger scrutiny, interpret evidence, deliberate trade-offs, and authorize revisions.

This separation is deliberate: functions specify governance purposes; principles specify enactment rules; artifacts and roles make both auditable and resistant to governance capture.

4.2. Core Cognitive Functions: Validation, Detection, and Deliberation

4.2.1. F1 – Validation: Epistemic Adequacy and Justified Alignment

Validation keeps the DT's epistemic foundations contestable and revisable. It asks whether the DT's representational assumptions, data foundations, causal structure, and objectives remain adequate for the decisions being made [9]. Importantly, validation here is not framed as a one-time certification. In critical multi-actor infrastructure contexts, "correctness" is rarely a stable achievement. What governance can insist upon, instead, is continuous, disciplined testing and documented justification for why the current model, data, and objectives remain acceptable in light of observed system behavior and decision consequences. This framing emphasizes validation as an ongoing governance practice rather than a terminal act of verification.

Validation therefore combines scheduled reviews (e.g., annual assumption audits) with event-triggered reviews when anomalies accumulate, when stakeholders raise credible concerns, or when operational conditions shift materially. Testing, evaluation, verification, and validation approaches implemented across the DT lifecycle — from design to operation — support this lifecycle orientation [77]. SI extends that logic by tying validation explicitly to contestability and revision authority, so that evidence can lead to change rather than merely documentation.

4.2.2. F2 – Detection : Harm Visibility and Early-Warning Governance

Detection searches deliberately for emerging harms, biases, and effects that may undermine stakeholders' capabilities in DT-mediated decisions. It does not wait for crises or complaints, partly because many harms are diffuse and because less powerful stakeholders often lack the resources to raise effective complaints. Detection operationalizes what the bias literature has long implied: fairness failures are often invisible unless the institution chooses to look for them, with the right disaggregation and the right questions [14,15,47].

Practically, detection treats DT decisions and outcomes as hypotheses about how the system behaves and who benefits. It monitors outcomes across stakeholder categories and contexts (e.g., by vessel size category, operator type, route class, or geography), and it investigates mechanisms that link design choices (objective functions, weightings, constraints) to observed disparities. Opacity becomes a detection risk: the more black-boxed the decision logic, the harder it is to diagnose mechanisms and the easier it is for systematic disadvantage to persist without remedy [8].

4.2.3. F3 – Deliberation : Legitimacy, Contestability, and Redress

Deliberation provides the institutional arena where evidence and values are brought together to interpret findings from validation and detection, negotiate trade-offs, and authorize revisions. This is not a consultation exercise. It is meant to be consequential. In the SIF, deliberation is the mechanism that produces legitimacy under plural values and asymmetric power, by ensuring that affected

 doi:10.20944/preprints202605.2031.v1

stakeholders can understand the system's rationale, contest objectives, and shape revisions when harms or misalignments are identified [11].

Research on algorithmic governance and public decision-making highlights an important legitimacy dynamic. Perceived legitimacy depends on how decisions are justified and whether such justification is experienced as sufficient and fair, especially in public decision contexts [12]. Moreover, in high-stakes systems, deliberation must support contestability, not merely explanation, aligning with arguments that justifiability and contestability constitute core legitimacy requirements [16]. Finally, deliberation should not end at "discussion." It requires pathways for redress so that individuals and groups can contest harmful outcomes and obtain remedies, thereby strengthening human agency in AI-mediated environments [39,52].

4.3. Operational Principles P1–P5

The SIF introduces five cross-cutting principles as design rules that make the three functions operable as routine practice.

P1 *explicit fidelity thresholds*. Divergence between DT representations and observed system behavior is treated as a governance trigger. This prevents institutions from normalizing anomalies as "noise" when they may signal model drift or objective misalignment [77]. By formalizing thresholds for acceptable divergence, P1 requires organizations to treat sustained discrepancies as actionable signals that warrant investigation, justification, or intervention rather than routine exceptions to be absorbed operationally.

P2 *systematic assumption logging and interrogation*. DT's intellectual foundations are made durable, inspectable, and revisable. Assumptions that disappear into code become politically and operationally untouchable, and then governance becomes symbolic [9]. This failure mode is also noted in responsible AI frameworks that stress documentation without sustained interpretive capacity [50].

P3 *structured stakeholder integration.* Validation and detection processes incorporate diverse forms of knowledge ensuring that deliberation extends beyond internal technical experts. Without such integration, a system may remain technically sophisticated while drifting away from legitimate social objectives and from the capabilities and interests of affected stakeholders [17,19,41].

P4 *clear accountability and role distribution*. Mandates are assigned, so no single actor monopolizes validation, detection, and deliberation. This operationalizes accountability as a structured practice distributed across the lifecycle rather than a label attached to one role [11], echoing governance analyses that stress enforceable role differentiation as a condition for responsible AI oversight [71].

P5 *adaptive revision and learning*. Findings lead to changes in models, objectives, metrics, and routines. This implements double-loop learning and adaptive governance in DT contexts [60], and aligns with recent DT governance frameworks emphasizing continuous revision rather than static compliance [33].

Together these principles address a recurring AI governance failure mode: organizations adopt ethical principles but struggle to translate them into implementable routines, tools, and incentives. SI's operational principles layer is meant to be that translation bridge, consistent with critiques of the "operationalization gap" in AI ethics practice [23,76]. In urban contexts, systematic reviews confirm that DT governance remains weakly evidenced despite the rapid proliferation of smart city technical deployments [78].

4.4. Governance Artifacts and Accountability Roles

To remain auditable and durable, the SIF is implemented through a limited set of governance artifacts and roles that concretely embody its functions and principles.

- **Assumption Log**: A structured version-controlled record of DT assumptions (boundaries, causal relations, proxies, objectives, and constraints). It supports validation by keeping foundational modeling choices visible, inspectable, and open to epistemic scrutiny, thereby preserving their

 doi:10.20944/preprints202605.2031.v1

revisability over time [9]. It also addresses failures where assumptions remain implicit and effectively insulated from challenge [50].

- **Bias and Impact Assessment Reports**: Periodic syntheses of detection findings, including disaggregated outcome patterns and plausible mechanisms generating systematic advantage or disadvantage. Their purpose is to make harm visibility routine and to counter the tendency of aggregate metrics to conceal unequal effects [14,47].
- **The Decision Journal**: A traceable record of significant DT-mediated decisions, including options considered, evidence used, stakeholder inputs, and the reasons for final choices. This supports accountability and institutional learning by enabling reconstruction and contestation [11].
- **Confidence Panel Dossiers**: Evidence packages prepared for deliberative bodies, enabling structured review and informed contestation. These dossiers are necessary where information is technically dense or fragmented; under opacity, interpretive governance becomes fragile and requires explicit mediating practices to sustain intelligibility and contestability [8].
- **Accountability roles**: Roles distribute responsibility and reduce capture by assigning explicit institutional mandates to the framework's three cognitive functions. A DT Owner maintains the system technically and implements revisions authorized through governance processes. A Cognitive Governor leads Validation (F1), convenes assumption reviews, and triggers deeper scrutiny when fidelity thresholds are breached. A Bias and Impact Auditor leads Detection (F2) and has authority to request additional analyses where disparities or risks are identified. Stakeholder Representatives participate in Deliberation (F3) with a mandate that is consequential rather than merely advisory. An External Reviewer periodically assesses the SI process itself for blind spots and capture risk. Finally, a Governance Committee or Confidence Panel serves as the institutional locus for deliberation and revision authority.

This role architecture operationalizes the separation between validation, detection, and deliberation, ensuring that no single actor monopolizes epistemic, evaluative, and decision authority. It aligns with AI governance scholars emphasizing that accountability requires structured oversight and enforceable consequences rather than transparency claims alone [7]. Redress mechanisms can be formalized so that contestation has practical effect and harms can be remedied [52].

Figure 3 provides a summary of the proposed SIF, and Table 2 provides a comparison between SIF and existing governance frameworks.

Table 2. Comparative Analysis of Governance Frameworks.

Governance dimension	NIST AI RMF [20]	EU AI Act [21]	GDPR [22]	SIF
Primary focus	Risk management & mitigation processes	Regulatory classification & compliance requirements	Data protection & individual rights	Cognitive capacity for continuous structured interrogation
Assumption interrogation	Limited; focuses on risk identification not assumption validation	Minimal; assumes technical risk classification is sufficient	Minimal; assumes GDPR compliance is sufficient	Central; treats all assumptions as falsifiable conjectures requiring continuous interrogation
Algorithmic bias detection	Mentioned but no operational depth; assumed risk identification is sufficient	Yes; prohibits high-risk uses & requires impact assessments but limited ongoing detection	No	Central; requires proactive, disaggregated detection & bias auditing (F2)
Stakeholder voice & deliberation	Limited; presupposes consultation, not deliberation with power	Limited; focuses on transparency & notification, not deliberation with power to reshape governance	Data subject rights & some rectification rights	Central; enables genuine deliberation with power to reshape governance (F3)
Adaptive revision of objectives & governance structure	Limited; assumes compliance suffices, not fundamental learning	Prescriptive categories don't easily accommodate paradigm change	Minimal; assumes legal compliance suffices	Central; treats anomalies as learning opportunities driving revision (P5)

 doi:10.20944/preprints202605.2031.v1

Scope of application	AI systems broadly	High-risk AI applications specifically	Personal data processing	DT-mediated decision infrastructures in critical multi-actor contexts
Evaluation mechanism	Procedural auditing & maturity assessment	Compliance auditing & regulatory approval	Legal & formal auditing	Relational; includes cognitive, ethical, & participatory dimensions

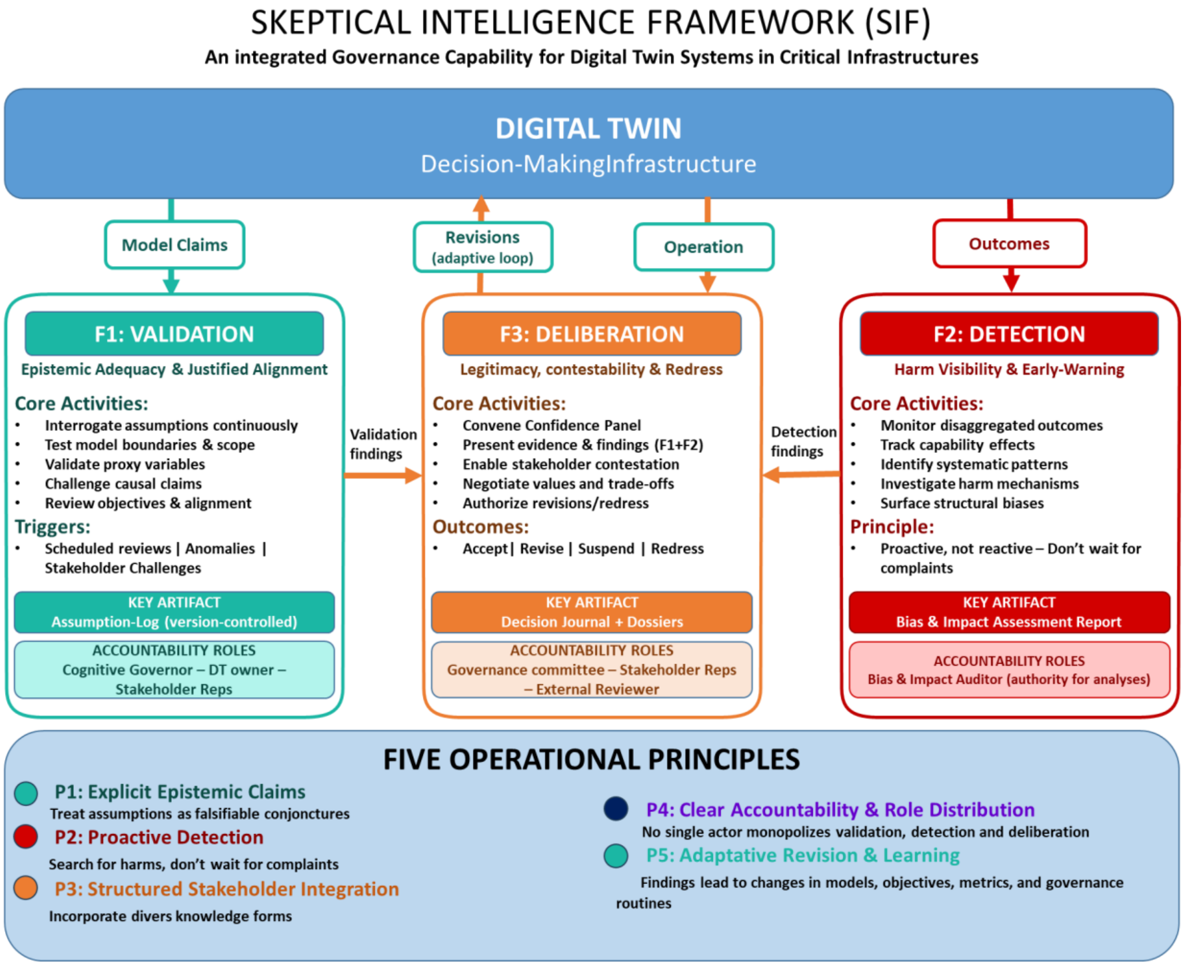

Figure 3. Skeptical intelligence framework for DT governance.

4.5. Application and Evaluation Strategy

Consistent with DSR, SI evaluation can be structured as a phased program: conceptual evaluation (internal coherence and theoretical grounding), prototype implementation (feasibility, usability, organizational fit), and outcome evaluation (observed changes in governance and stakeholder outcomes). Outcome evaluation should combine quantitative indicators (distribution of delays, incident rates, environmental compliance patterns) with qualitative evidence (stakeholder perceptions of voice, fairness, and learning), reflecting the socio-technical character of legitimacy in DT-mediated governance [12].

In a port context, a DT may generate berth and crane allocation recommendations throughout the day. SI evaluation would proceed by (i) validating fidelity using prediction and override metrics, with explicit thresholds triggering event-based review (F1); (ii) detecting distributional impacts through disaggregated monitoring of waiting times and service levels across stakeholder categories (F2); and (iii) assessing whether evidence leads to consequential revisions of objectives and constraints through documented deliberation and stakeholder voice (F3). This operationalizes evaluation beyond technical accuracy by directly examining harm visibility, contestability, and revision behavior, consistent with the governance-quality emphasis.

 doi:10.20944/preprints202605.2031.v1

5. Discussion, Implications, and Future Research

Following the systematic scoping review (Sections 2-3) that identified three convergent design requirements, Section 4 translated these into the SIF through DSR (Section 2.3). The framework operationalizes structured skepticism through three interdependent functions (F1-F3), five principles (P1–P5), and concrete governance artifacts and accountable roles. This section discusses theoretical contributions, practical implications, limitations, and future research directions.

5.1. Theoretical Contributions

First, SI distinctively operationalizes epistemological traditions as organizational governance commitments. Unlike existing frameworks that reference falsificationism, paradigm critique, and research-program evaluation rhetorically [20,21], SI translates them into specific functions (F1 Validation), principles (P1 fidelity thresholds, P2 assumption logging), and artifacts (Assumption Log). DT models and objectives become contestable conjectures requiring continuous stress-testing rather than stabilization through routine use. These safeguards prevent organizations from confusing model persistence with model adequacy. This approach also addresses epistemic opacity as a governance hazard rather than technical inconvenience: opacity weakens institutions' ability to justify decisions, diagnose harms, and learn credibly from anomalies [8]. SI therefore emphasizes deliberation as a cognitive governance function, not a communications layer. Without interpretability, "participation" becomes decor and legitimacy becomes fragile.

Second, SI integrates empirical evidence on algorithmic harm with normative commitments to capabilities and democratic legitimacy as a single governance problem. Detection (F2) surfaces latent burdens and exclusions; Validation (F1) interrogates whether they originate in data representativeness, boundary choices, or objective functions; Deliberation (F3) decides what should change. This advances beyond transparency-focused legitimacy [12] by treating contestability as a governance requirement with enforceable revision authority, not a desirable afterthought [16].

Third, SI advances adaptive governance theory by operationalizing adaptive operational principles in technical systems. Multi-actor deliberation becomes constitutive of "good performance" rather than consultative input about achieving pre-defined objectives. This matters because DTs increasingly function as infrastructural decision devices, requiring governance capable of revising objectives, constraints, and representations under uncertainty and conflict. The framework decomposes accountability from desirable aspiration into auditable practice through enforceable responsibilities and traceable artifacts (Assumption Log, Bias and Impact Reports, Decision Journal, Confidence Panel Dossiers) that structure how responsibility is assigned, exercised, and reviewed. This resonates with recent analyses of accountability in socio-technical and distributed AI systems [7].

Positioned within recent governance scholarship on urban and critical infrastructure systems, the SIF addresses a gap that adjacent contributions leave open. Stage-gate frameworks for municipal AI governance provide procedural sequencing for responsible deployment but do not specify the cognitive content—epistemic interrogation, disaggregated detection, consequential deliberation—that each gate should enact [73]. Social DT frameworks integrate participatory mechanisms and bottom-up governance roles but do not address how participation can be sustained when algorithmic opacity limits what stakeholders can actually scrutinize [74]. The SIF contributes the epistemic and deliberative architecture that makes both procedural sequencing and participatory inclusion substantive rather than ceremonial.

5.2. Practical Implications

SI requires governance as integral to DT architecture and institutional design from inception, not as post-deployment compliance. This reframes what implementation teams consider "core work": governance becomes ongoing intellectual practice with dedicated authority and resourcing, not periodic auditing.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 May 2026 doi:10.20944/preprints202605.2031.v1

At the organizational level, the most immediate implication is role formalization. SI requires, at minimum, a Cognitive Governor responsible for leading Validation (F1) and maintaining the Assumption Log, and a Bias and Impact Auditor responsible for proactive Detection (F2) and producing Bias and Impact Reports. These roles are not symbolic; they need explicit mandates, access to decision forums, and authority to trigger reviews when thresholds are exceeded or harms detected. P4 becomes crucial because it prevents governance capture by any single actor, particularly in settings where operational incentives may favor throughput and stability over scrutiny and revision.

At the process level, SI implies three operational routines:

1. *Routine validation and event-triggered review*: Fidelity thresholds (P1) trigger structured validation when model reality divergence becomes meaningful, preventing normalization as operational noise.
2. *Disaggregated monitoring as standing practice:* Detection (F2) explicitly searches for systematic advantage or disadvantage across stakeholder groups, time windows, and operational contexts proactively, not reactively.
3. *Deliberation with revision authority and redress:* Deliberation (F3) operates through intelligible evidence packages and documented decision pathways, with authority to mandate DT changes to objectives, constraints, or data practices. Contestability requires credible redress mechanisms preserving human agency when systems cause harm [52].

In port contexts, implications are especially concrete because DT-mediated decisions are frequent and distributive. SI offers a way to institutionalize questions that otherwise remain informal and politically awkward. Which stakeholder groups are systematically delayed? Which objective weights drive those delays? Who has standing to challenge them? What would count as sufficient justification for maintaining them?

The same governance questions apply in smart city contexts where urban DTs mediate resource allocation, service prioritization, and planning decisions across populations with unequal access and voice. Which neighborhoods receive slower emergency response because of how the DT weights demand patterns? Which citizen groups are effectively excluded from participatory interfaces that presuppose digital literacy? Do municipal authorities have structured mechanisms to detect these disparities before they compound? SI provides the institutional architecture for posing and acting on such questions systematically, regardless of the specific infrastructure domain.

SI also addresses the "operationalization gap" that many organizations encounter when implementing AI ethics. In practice, principles proliferate while implementable routines remain thin. SI's artifacts and roles design prevents that drift by turning ethical and legitimacy commitments into inspectable, repeatable organizational work [50].

5.3. Limitations

The framework's primary limitation is that it is theoretically grounded but not yet empirically validated in a real DT governance environment. Until SI is implemented in an operational DT within a critical multi-actor infrastructure context, claims about effectiveness should be treated as plausible but unproven. A second limitation concerns organizational capacity. SI demands skilled analytical work, structured deliberation, and sustained attention to disaggregated impacts; some organizations will simply lack the maturity, staffing, data access, or political latitude to do that consistently.

A methodological limitation concerns literature selection. While the systematic scoping review followed explicit inclusion criteria (Section 2.2), the 45 Tier 2 sources were purposively selected based on four explicit rules (Section 2.2.2), potentially introducing theoretical framing bias despite transparency measures.

A third limitation is the opacity threshold problem. If key DT components are fundamentally non-interpretable, governance can still operate through outcome monitoring and contestation, but diagnosing mechanisms becomes harder and justification more fragile. This limitation is not merely technical but institutional: opaque systems increase dependence on trust, and trust is exactly what governance is supposed to avoid treating as an implicit default. Also, power asymmetries can deform

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 May 2026 doi:10.20944/preprints202605.2031.v1

deliberation. Even well-designed structures for voice may be undermined if dominant actors control what evidence is surfaced, what revisions are acceptable, or what counts as "reasonable" dissent.

Finally, generalizability is bounded. The framework is designed for critical multi-actor infrastructure contexts where DT-mediated decisions have significant distributive and accountability implications. Applying SI to smaller-scale DT deployments will likely require adaptation, particularly regarding the scope and intensity of deliberation and external review.

5.4. Future Research Directions

Future research should treat the SIF as the starting point of an empirical and comparative program. The immediate next step is implementation in an operational port DT environment, with iterative refinement based on observed friction points, institutional resistance, and stakeholder dynamics. This step is foundational: it allows researchers to determine whether the framework's three governance functions operate as designed or degrade into performative rituals under operational pressure. This is a risk that governance research on comparable technological transitions has repeatedly documented.

The natural extension of this implementation lies in federating the port Digital Twin with the city's in a sustainable mobility perspective. Congestion generated by freight transport serving the port is a shared governance challenge, involving common infrastructures—road corridors, rail networks, periurban logistics zones—whose coordination exceeds the institutional boundaries of the port authority. This transition to a federated port-city architecture represents a privileged testbed for the SIF: it confronts the framework's three cognitive functions with actors holding distinct mandates, temporalities, and legitimacies, and allows researchers to assess whether skeptical intelligence mechanisms transfer beyond the organizational perimeter of a single actor.

A third direction concerns how governance demands evolve as DT autonomy deepens. Current deployments already strain existing institutional arrangements. As cognitive twins adapt their own objectives and federated architectures coordinate decisions across organizational boundaries, the pressure on governance mechanisms will intensify. Research should treat governance capacity as a moving target rather than a fixed design problem.

A fourth direction concerns cross-sector transfer beyond the port-city continuum: energy systems, healthcare systems, and supply chains. Comparative analysis should identify which governance design elements remain stable across sectors and which require context-specific adaptation, building toward a generalizable theory of institutional capacity for autonomous infrastructure governance.

Finally, the framework raises measurement questions that warrant dedicated methodological attention. How do we distinguish genuine validation from ceremonial compliance? What evidence demonstrates that detection surfaces harms that would otherwise remain hidden? How can deliberation be shown to be consequential rather than legitimacy theatre? Answering these questions requires mixed-method designs combining process tracing, disaggregated distributive metrics, and qualitative legitimacy indicators, as well as quasi-experimental comparisons across sites implementing different subsets of SI mechanisms.

6. Conclusions

DTs are moving from analytical representations to operational decision-shaping systems whose outputs redistribute time, risk, and opportunity across stakeholders with unequal voice and power. In such settings, governance cannot be reduced to procedural compliance or technical performance. What is needed is an institutional capacity for structured skepticism in the form of continuous validation of assumptions and objectives, proactive detection of latent harms and exclusions, and deliberation that is both meaningful and capable of mandating revision.

This paper proposed the SIF as a governance framework designed to meet that need. By integrating epistemological rigor with empirical sensitivity to algorithmic harms and a commitment to democratic legitimacy, SI specifies how organizations can make DT governance auditable,

 doi:10.20944/preprints202605.2031.v1

contestable, and adaptive over time. The framework's central claim is not that skepticism solves governance. It is that skepticism can be designed, resourced, and institutionalized so that DTs remain revisable when evidence changes, when harms emerge, and when stakeholders contest the meaning of "good performance."

The work now needs to be moved into empirical testing. If SI is implemented across critical multi-actor infrastructure contexts and evaluated rigorously, it can help shift DT governance from documentation-heavy procedures toward practices that are epistemically disciplined, socially legitimate, and institutionally accountable. The framework's applicability extends across critical multi-actor infrastructure contexts—from port logistics to smart city governance and energy networks—wherever DT-mediated decisions shape outcomes for stakeholders with unequal voice and power.

Acknowledgments: The authors used generative AI tools exclusively for language editing and formatting assistance. No generative AI tools were used for scientific content generation, data analysis, interpretation, or research design.

Appendix A. Thematic Clusters (C1-C10) and Search Queries

C1: Epistemology & Scientific Methodology
("design science" OR "scientific discovery" OR "logic") AND
("research" OR "information systems" OR "epistemology") AND
("methodology" OR "validation")
C2: Adaptive Governance & Learning
("adaptive governance" OR "adaptive management" OR "accountability") AND
("social-ecological systems" OR "conceptual framework" OR "governance framework") AND
("resilience" OR "responsibility" OR "learning")
C3: Algorithmic Bias & Fairness
("algorithmic bias" OR "unfairness" OR "gender bias") AND
("machine learning" OR "data" OR "AI systems") AND
("accuracy" OR "fairness" OR "intersectionality" OR "discrimination")
C4: Generative AI & Cognitive Processes
("generative AI" OR "AI chatbot" OR "large language models") AND
("cognitive processes" OR "digital well-being" OR "mental health") AND
("perception" OR "impact" OR "implications" OR "students" OR "psychological")
C5: Human-Automation Interaction & Trust
("human-automation" OR "human-machine" OR "human-AI interaction") AND
("trust" OR "trustworthiness" OR "automation bias") AND
("reliance" OR "acceptance" OR "digital twins")
C6: AI Regulation & Governance Frameworks
("artificial intelligence" OR "AI") AND
("regulation" OR "risk management" OR "governance") AND
("framework" OR "compliance" OR "trustworthy AI" OR "responsible AI")
C7: Digital Twins - Core Technologies
("digital twins" OR "virtual twins") AND
("decentralized" OR "complex dynamical systems" OR "complex systems") AND
("architecture" OR "modeling" OR "systems design")
C8: Digital Twins - Governance & Operations
("digital twins" OR "digital twin") AND
("governance" OR "governance framework" OR "regulation") AND
("autonomous" OR "cognitive" OR "human in the loop" OR "supply chain" OR "city" OR "policymakers")
C9: Digital Twins - Inclusivity & Citizen Participation

 doi:10.20944/preprints202605.2031.v1

("digital twins" OR "urban digital") AND
("invisible citizens" OR "inclusivity" OR "urban") AND
("futures" OR "rethinking" OR "citizens")
C10: Human Development & Agency
("development as freedom" OR "human development" OR "capabilities approach") AND
("society 5.0" OR "open models" OR "decision-making") AND
("digital twins" OR "agency" OR "well-being" OR "governance")

References

1. San, O.; Pawar, S.; Rasheed, A. Decentralized digital twins of complex dynamical systems. Sci. Rep. 2023, 13, 20087. https://doi-org/10.1038/s41598-023-47078-9
2. Tao, F.; Xiao, B.; Qi, Q.; Cheng, J.; Ji, P. Digital twin modeling. J. Manuf. Syst. 2022, 64, 372–389. https://doi.org/10.1016/j.jmsy.2022.06.015
3. Kalaboukas, K.; Rožanec, J.; Košmerlj, A.; Kiritsis, D.; Arampatzis, G. Implementation of cognitive digital twins in connected and agile supply networks—An operational model. Appl. Sci. 2021, 11, 4103. https://doi.org/10.3390/app11094103
4. Sajadieh, S.M.M.; Noh, S.D. From simulation to autonomy: Reviews of the integration of artificial intelligence and digital twins. Int. J. Precis. Eng. Manuf.-Green Technol. 2025, 12, 1597–1628. https://doi-org/10.1007/s40684-025-00750-z
5. Klar, R.; Fredriksson, A.; Angelakis, V. Digital twins for ports: Derived from smart city and supply chain twinning experience. IEEE Access 2023, 11, 71777–71799. https://doi.org/10.1109/ACCESS.2023.3295495
6. Haraguchi, M.; Funahashi, T.; Biljecki, F. Assessing governance implications of city digital twin technology: A maturity model approach. Technol. Forecast. Soc. Change 2024, 204, 123409. https://doi.org/10.1016/j.techfore.2024.123409
7. Novelli, C.; Taddeo, M.; Floridi, L. Accountability in artificial intelligence: What it is and how it works. AI Soc. 2024, 39, 1871–1882. https://doi.org/10.1007/s00146-023-01635-y
8. Carabantes, M. Black-box artificial intelligence: An epistemological and critical analysis. AI Soc. 2020, 35, 309–317. https://doi.org/10.1007/s00146-019-00888-w
9. Wagg, D.J.; Burr, C.; Shepherd, J.; Conti, Z.X.; Enzer, M.; Niederer, S. The philosophical foundations of digital twinning. Data-Cent. Eng. 2025, 6, e12. https://doi.org/10.1017/dce.2025.4
10. Hirschman, A.O. Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States; Harvard University Press: Cambridge, MA, USA, 1972.
11. Bovens, M. Analysing and assessing accountability: A conceptual framework. Eur. Law J. 2007, 13, 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x
12. de Fine Licht, K.; de Fine Licht, J. Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. AI Soc. 2020, 35, 917–926. https://doi-org/10.1007/s00146-020-00960-w
13. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. Proc. Mach. Learn. Res. 2018, 81, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html
14. Barocas, S.; Selbst, A.D. Big data's disparate impact. Calif. Law Rev. 2016, 104, 671–732.
15. Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. A survey on bias and fairness in machine learning. ACM Comput. Surv. 2021, 54, 1–35. https://doi.org/10.1145/3457607
16. Henin, C.; Le Métayer, D. Beyond explainability: Justifiability and contestability of algorithmic decision systems. AI Soc. 2022, 37, 1397–1410. https://doi.org/10.1007/s00146-021-01251-8
17. Sen, A. Development as Freedom; Oxford University Press: Oxford, UK, 1999.
18. Nussbaum, M.C. Creating Capabilities: The Human Development Approach; Harvard University Press: Cambridge, MA, USA, 2011.
19. Ye, X.; Goodchild, M. Toward ethical GeoDesign in the Urban Digital Twin era. J. Plan. Educ. Res. 2025, 45, 721–725. https://doi-org/10.1177/0739456X251359233

20. NIST. Artificial Intelligence Risk Management Framework; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2023. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
21. Raj, P.; Christakis, T. EU AI Act: Final Text Published in the Official Journal of the EU with Interactive Table of Contents; HAL Open Science, 2024. https://hal.science/hal-05462628/
22. Voigt, P.; Von dem Bussche, A. The EU General Data Protection Regulation (GDPR): A Practical Guide, 1st ed.; Springer: Cham, Switzerland, 2017. https://doi-org/10.1007/978-3-319-57959-7
23. Morley, J.; Kinsey, L.; Elhalal, A.; Garcia, F.; Ziosi, M.; Floridi, L. Operationalising AI ethics: Barriers, enablers and next steps. AI Soc. 2023, 38, 411–423. https://doi.org/10.1007/s00146-021-01308-8
24. Joanna Briggs Institute. JBI Manual for Evidence Synthesis; Joanna Briggs Institute, 2024. https://jbi.global/critical-appraisal-tools
25. Gregor, S.; Hevner, A.R. Positioning and presenting design science research for maximum impact. MIS Q. 2013, 37, 337–355. http://www.jstor.org/stable/43825912
26. Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S. A design science research methodology for information systems research. J. Manag. Inf. Syst. 2007, 24, 45–77. https://doi.org/10.2753/MIS0742-1222240302
27. Arksey, H.; O'Malley, L. Scoping studies: Towards a methodological framework. Int. J. Soc. Res. Methodol. 2005, 8, 19–32. https://doi-org/10.1080/1364557032000119616
28. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design science in information systems research. MIS Q. 2004, 28, 75–105. https://doi.org/10.2307/25148625
29. Xu, H.; Omitaomu, F.; Sabri, S.; et al. Leveraging generative AI for urban digital twins: A scoping review on the autonomous generation of urban data, scenarios, designs, and 3D city models for smart city advancement. Urban Inform. 2024, 3, 29. https://doi-org/10.1007/s44212-024-00060-w
30. Dembski, F.; Wössner, U.; Letzgus, M.; Ruddat, M.; Yamu, C. Urban digital twins for smart cities and citizens: The case study of Herrenberg, Germany. Sustainability 2020, 12, 2307. https://doi.org/10.3390/su12062307
31. Ye, X.; Du, J.; Han, Y.; Newman, G.; Retchless, D.; Zou, L.; Ham, Y.; Cai, Z. Developing human-centered urban digital twins for community infrastructure resilience: A research agenda. J. Plan. Lit. 2022, 38, 187–199. https://doi-org/10.1177/08854122221137861
32. Weil, C.; Bibri, S.E.; Longchamp, R.; Golay, F.; Alahi, A. Urban digital twin challenges: A systematic review and perspectives for sustainable smart cities. Sustain. Cities Soc. 2023, 99, 104862. https://doi.org/10.1016/j.scs.2023.104862
33. Kalaboukas, K.; Kiritsis, D.; Arampatzis, G. Governance framework for autonomous and cognitive digital twins in agile supply chains. Comput. Ind. 2023, 146, 103857. https://doi-org/10.1016/j.compind.2023.103857
34. Kreuzer, T.; Papapetrou, P.; Zdravkovic, J. Artificial intelligence in digital twins—A systematic literature review. Data Knowl. Eng. 2024, 151, 102304. https://doi-org/10.1016/j.datak.2024.102304
35. Argota Sánchez-Vaquerizo, J. Urban Digital Twins and metaverses towards city multiplicities: Uniting or dividing urban experiences? Ethics Inf. Technol. 2025, 27, 4. https://doi-org/10.1007/s10676-024-09812-3
36. El-Agamy, R.F.; Sayed, H.A.; AL Akhatatneh, A.M.; et al. Comprehensive analysis of digital twins in smart cities: A 4200-paper bibliometric study. Artif. Intell. Rev. 2024, 57, 154. https://doi-org/10.1007/s10462-024-10781-8
37. Cerina, R.; Rouméas, É. The democratic ethics of artificially intelligent polling. AI Soc. 2025, 40, 3209–3223. https://doi-org/10.1007/s00146-024-02150-4
38. Mureddu, F.; Paciaroni, A.; Pavelka, T.; Pemberton, A.; Remotti, L.A. Rights and responsibilities: Legal and ethical considerations in adopting local digital twin technology. In Decide Better; Raes, L., et al., Eds.; Springer: Cham, Switzerland, 2025. https://doi-org/10.1007/978-3-031-81451-8_11
39. Suffia, G. How to regulate a Digital Twin City? Insights from a proactive law approach. In Proceedings of the 24th Annual International Conference on Digital Government Research, Gdańsk, Poland, 11–14 July 2023; pp. 122–128. https://doi.org/10.1145/3598469.3598482
40. Sel, K.; Hawkins-Daarud, A.; Chaudhuri, A.; et al. Survey and perspective on verification, validation, and uncertainty quantification of digital twins for precision medicine. npj Digit. Med. 2025, 8, 40. https://doi-org/10.1038/s41746-025-01447-y

 doi:10.20944/preprints202605.2031.v1

41. Kallina, E.; Singh, J. Stakeholder involvement for responsible AI development: A process framework. In Proceedings of the 4th ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, Arlington, VA, USA, 30 October–1 November 2024; pp. 1–14. https://doi.org/10.1145/3689904.3694698
42. Makanadar, A. Digital surveillance capitalism and cities: Data, democracy and activism. Humanit. Soc. Sci. Commun. 2024, 11, 1533. https://doi-org/10.1057/s41599-024-03941-2
43. Allen, D.; Hubbard, S.; Lim, W.; et al. A roadmap for governing AI: Technology governance and power-sharing liberalism. AI Ethics 2025, 5, 3355–3377. https://doi-org/10.1007/s43681-024-00635-y
44. Hiller, J.; Mansour, M.; Kremer, N.; Crampen, D.; von Behren, S. Mapping Urban Digital Twins Across Regions: An Exploratory Study of Maturity, Implementation Status, and Authority. Smart Cities 2026, 9, 49. https://doi.org/10.3390/smartcities9030049
45. Raposo, V.L. When facial recognition does not 'recognise': Erroneous identifications and resulting liabilities. AI Soc. 2024, 39, 1857–1869. https://doi.org/10.1007/s00146-023-01634-z
46. Afreen, J.; Mohaghegh, M.; Doborjeh, M. Systematic literature review on bias mitigation in generative AI. AI Ethics 2025, 5, 4789–4841. https://doi-org/10.1007/s43681-025-00721-9
47. Balayn, A.; Lofi, C.; Houben, G.J. Managing bias and unfairness in data for decision support: A survey of machine learning and data engineering approaches to identify and mitigate bias and unfairness within data management and analytics systems. VLDB J. 2021, 30, 739–768. https://doi-org/10.1007/s00778-021-00671-8
48. Lam, W.Y.; Wernholm, E. Algorithmic Bias in City-Scale Digital Twin Data Management Processes: A Qualitative Exploration. Master's Thesis, Lund University, Lund, Sweden, 2023. https://lup.lub.lu.se/student-papers/search/publication/9121472
49. Hung, T.W.; Yen, C.P. Predictive policing and algorithmic fairness. Synthese 2023, 201, 206. https://doi-org/10.1007/s11229-023-04189-0
50. Díaz-Rodríguez, N.; Del Ser, J.; Coeckelbergh, M.; de Prado, M.; Herrera-Viedma, E.; Herrera, F. Connecting the dots in trustworthy artificial intelligence: From ethical principles to requirements and regulation. Inf. Fusion 2023, 99, 416–438. https://doi.org/10.1016/j.inffus.2023.101896
51. White, G.; Zink, A.; Codecá, L.; Clarke, S. A digital twin smart city for citizen feedback. Cities 2021, 110, 103064. https://doi.org/10.1016/j.cities.2020.103064
52. Fanni, R.; Steinkogler, J.; Posada, A. Enhancing human agency through redress in artificial intelligence systems. AI Soc. 2023, 38, 537–547. https://doi.org/10.1007/s00146-022-01454-7
53. Bhanye, J. Flood-tech frontiers: Smart but just? A systematic review of AI-driven urban flood adaptation and associated governance challenges. Discov. Glob. Soc. 2025, 3, 59. https://doi-org/10.1007/s44282-025-00190-9
54. Satz, D.; White, S. What is wrong with inequality? Oxford Open Econ. 2024, 3, i4–i17. https://doi.org/10.1093/ooec/odad040
55. Popper, K.R. The Logic of Scientific Discovery; Hutchinson: London, UK, 1959.
56. Kuhn, T.S. The Structure of Scientific Revolutions; University of Chicago Press: Chicago, IL, USA, 1962.
57. Lakatos, I. The Methodology of Scientific Research Programmes: Philosophical Papers, Volume 1; Cambridge University Press: Cambridge, UK, 1978.
58. Feyerabend, P.K. Against Method: Outline of an Anarchistic Theory of Knowledge; Verso: London, UK, 1975.
59. Shah, M. The logics of discovery in Popper's evolutionary epistemology. J. Gen. Philos. Sci. 2008, 39, 303–319. https://doi-org/10.1007/s10838-008-9070-0
60. Argyris, C.; Schön, D.A. Organizational Learning: A Theory of Action Perspective; Addison-Wesley: Reading, MA, USA, 1978.
61. Folke, C.; Hahn, T.; Olsson, P.; Norberg, J. Adaptive governance of social-ecological systems. Annu. Rev. Environ. Resour. 2005, 30, 441–473. https://doi.org/10.1146/annurev.energy.30.050504.144511
62. Pahl-Wostl, C. A conceptual framework for analysing adaptive capacity and multi-level learning processes in resource governance regimes. Glob. Environ. Change 2009, 19, 354–365. https://doi.org/10.1016/j.gloenvcha.2009.06.001

 doi:10.20944/preprints202605.2031.v1

63. Alfaqeeh, M.; Zakiyah, N.; Postma, M.; et al. Setting priorities for healthcare interventions in Indonesia: A comprehensive conceptual framework. Int. J. Equity Health 2025, 24, 327. https://doi-org/10.1186/s12939-025-02668-z
64. Ishkhanyan, A. The sovereignty-internationalism paradox in AI governance: Digital federalism and global algorithmic control. Discov. Artif. Intell. 2025, 5, 123. https://doi-org/10.1007/s44163-025-00374-x
65. Kneuer, M.; Wimmer, M.A.; Bahms, G. Digital democratic participation in local governance from an urban-rural perspective. In Citizen Participation in Local Governance; Kössler, K., Schläppi, E., Eds.; Springer: Cham, Switzerland, 2026. https://doi-org/10.1007/978-3-031-99824-9_14
66. Arnesen, S.; Broderstad, T.S.; Fishkin, J.S.; et al. Knowledge and support for AI in the public sector: A deliberative poll experiment. AI Soc. 2025, 40, 3573–3589. https://doi-org/10.1007/s00146-024-02104-w
67. Ibitoye, A.O.; Nkwo, M.S.; Orji, R. Rethinking responsible AI from ethical pillars to sociotechnical practice. AI Ethics 2025, 5, 6207–6223. https://doi-org/10.1007/s43681-025-00809-2
68. Moro-Visconti, R. Is artificial intelligence a new stakeholding agent? Hum.-Intell. Syst. Integr. 2025. https://doi-org/10.1007/s42454-025-00069-9
69. Nica, E.; Popescu, G.H.; Poliak, M.; Kliestik, T.; Sabie, O.M. Digital twin simulation tools, spatial cognition algorithms, and multi-sensor fusion technology in sustainable urban governance networks. Mathematics 2023, 11, 1981. https://doi.org/10.3390/math11091981
70. Kumar, R.; Sporn, K.; Waisberg, E.; et al. Navigating healthcare AI governance: The comprehensive algorithmic oversight and stewardship framework for risk and equity. Health Care Anal. 2025. https://doi-org/10.1007/s10728-025-00537-y
71. Israr, M.; Khan, S.; Gupta, V.; Sharma, A.; Tiwari, D. Responsible AI governance and regulation. In A Compendium of Responsible Artificial Intelligence; CRC Press: Boca Raton, FL, USA, 2025; pp. 183–216. https://doi-org/10.1201/9781003514497
72. Shamsuddin, R.; Tabrizi, H.B.; Gottimukkula, P.R. Towards responsible AI: An implementable blueprint for integrating explainability and social-cognitive frameworks in AI systems. AI Perspect. Adv. 2025, 7, 1. https://doi-org/10.1186/s42467-024-00016-5
73. Yigitcanlar, T.; David, A.; Marasinghe, R.; Senadheera, S.; Hossain, T.; Ye, X.; Taeihagh, A. Governing urban AI from the frontline: A stage-gate framework for municipal algorithmic decision-making. Smart Cities 2026, 9, 81. https://doi.org/10.3390/smartcities9050081
74. Qanazi, S.; Leclerc, E.; Bosredon, P. Integrating Social Dimensions into Urban Digital Twins: A Review and Proposed Framework for Social Digital Twins. Smart Cities 2025, 8, 23. https://doi.org/10.3390/smartcities8010023
75. Singh, P.; Ul Haq, A.; Presser, M. Towards a responsible AI adoption/adaptation (RAA) ecosystem: Vision and model to keep socio-technological balance. In Global Internet of Things and Edge Computing Summit (GIECS 2025), Communications in Computer and Information Science; Presser, M., et al., Eds.; Springer: Cham, Switzerland, 2026; Volume 2719. https://doi-org/10.1007/978-3-032-09555-8_5
76. El-Deeb, S.; Jahankhani, H.; Hussien, O.A.A.M.; Arachchige, I.S.W. Towards responsible AI: Exploring AI frameworks, ethical dimensions and regulations. In Market Grooming: The Dark Side of AI Marketing; Emerald: Bingley, UK, 2024; pp. 139–157. https://doi-org/10.1108/978-1-83549-001-320241007
77. Waters, G. Testing, evaluation, verification and validation (TEVV) of digital twins: A comprehensive framework. arXiv 2025, arXiv:2507.04555. https://doi.org/10.48550/arXiv.2507.04555
78. Varzeshi, M.; Fien, J.; Irajifar, L. Integrating smart city technologies and urban resilience: A systematic review and research agenda for urban planning and design. Smart Cities 2026, 9, 2. https://doi.org/10.3390/smartcities9010002