A Task-Centric Framework for Measurable AI Adoption and Smart Governance in Organizations

Mehrdad Naderi

HEC Paris, Executive MSc in Innovation and Entrepreneurship
Edgewood University, Doctor of Education (EdD)
mail@mehrdadnaderi.com

Abstract

Artificial intelligence is now deployed across operational and administrative work in most large organizations, but adoption decisions are typically made at a level of analysis that is too coarse to be informative. Strategy is set at the firm level, governance is regulated at the system level, but the value, risk, and feasibility of AI concretely materialize at the task level. This paper proposes TC-MAGS, a Task-Centric Framework for Measurable AI Adoption and Smart Governance, as a conceptual design-science artifact for closing this unit-of-analysis mismatch. The framework integrates five reasoning dimensions into a single per-task decision-support pipeline: classification of tasks across human-led, AI-assisted, and rule-based automation modes; measurement readiness assessed on a four-level ladder; confidence and risk scored across six dimensions; explicit cost-of-inaction and cost-of-inefficiency reasoning; and governance checkpoints aligned to pilot, scale, and full-delegation gates. Each decision carries a justification trace and is mappable upward to leading governance instruments. None of the underlying claims is novel on its own; what this paper adds is the integration of these dimensions into one decision-support artifact suitable for operational AI governance inside organizations, including public-sector bodies and digital-service-delivery providers participating in smart-city ecosystems. The framework is presented as a conceptual artifact; no empirical validation is performed in this paper, and an agenda for future empirical instantiation is identified.

Keywords: Task-centric AI adoption; design science research; AI governance; measurement readiness; decision-support framework; operational AI governance.

1. Introduction

The diffusion of artificial intelligence into organizational work has accelerated to a point where the question is no longer whether AI will affect a given workforce, but how that effect will distribute itself across the work actually performed. Recent occupational-exposure measures suggest that roughly four-fifths of the United States workforce is in occupations where at least ten percent of work tasks are susceptible to large-language-model assistance, and nearly one-fifth are in occupations where the majority of tasks are exposed [1]. The AI Occupational Exposure index developed by Felten and colleagues [2] shows similarly broad, but unevenly distributed, reach across industries and geographies. Within any single workflow, however, AI's appropriate role differs sharply from one task to the next, a heterogeneity that Dell'Acqua and colleagues [3] characterize as a "jagged frontier" on the basis of controlled field-experimental evidence. Adoption is, in this precise sense, task-grained.

Decisions about adoption are not. Organizational strategy is typically formulated at the firm level; AI governance, as codified in instruments such as the NIST AI Risk Management Framework [4], Regulation (EU) 2024/1689 [5], and ISO/IEC 42001:2023 [6], regulates AI systems. Neither the firm nor the system is the granularity at which value and risk concretely materialize. The labor-economics tradition has long argued that work is most usefully analyzed as a bundle of tasks with distinct characteristics [7, 8], and the task-technology fit perspective in information systems established three decades ago that technology improves performance only when its functionality fits task characteristics [9]. A unit-of-analysis mismatch sits at the heart of contemporary AI adoption practice: the layer at which decisions are produced is not the layer at which their consequences unfold.

This mismatch manifests in three persistent problems. The first is invisibility: most organizations lack a task inventory granular enough to support adoption decisions, and conduct adoption discussions at the level of tools or platforms rather than discrete units of work. The second is immeasurability: even when tasks are identified, the measurement substrate required to evaluate adoption, meaning baseline data, standardized metrics, and validated indicators, is frequently absent or anecdotal, a gap noted in empirical work on organizational AI readiness [10, 11]. The third is lagging governance: existing instruments specify functions, obligations, and management-system requirements but do not provide a per-task decision substrate. The NIST AI RMF explicitly does not prescribe risk tolerance [4]; Regulation (EU) 2024/1689 imposes obligations per AI system, but a system typically implements many tasks of differing risk; ISO/IEC 42001:2023 specifies the architecture of an AI management system but leaves the operational decision logic for task-level adoption to the implementing organization. Each instrument leaves the granular operational layer underdetermined.

The mismatch is not solely a private-sector concern. Public-sector and municipal AI deployments inherit the same task-level problem with elevated accountability stakes, because the consequences of adoption fall on citizens rather than customers, and some candidate public-sector tasks may fall within regulated or high-risk categories under Regulation (EU) 2024/1689 [5]. Empirical evidence from the public-administration setting indicates that AI capability development is uneven and context-dependent, with public value outcomes mediated by organizational and task-level factors [12]. For smart-city ecosystems, the practical question is not only whether a city uses AI, but whether the organizations delivering digital public services can make AI-related task decisions measurable, accountable, and reviewable. Such organizations operate under heightened accountability obligations, integrate AI components from multiple vendors, and deliver services whose effects are felt by citizens rather than internal users.

Against this background, the present paper addresses the following research question: how can organizations structure AI adoption so that decisions are produced at the task level, conditioned on measurement readiness, scored against confidence and risk, justified through economic reasoning, and gated by governance checkpoints, in a way that maps coherently to existing governance instruments?

The paper proposes TC-MAGS, a Task-Centric Framework for Measurable AI Adoption and Smart Governance, as a conceptual design-science response to this question. TC-MAGS comprises six layers operating as a pipeline: enumeration of a task inventory; classification of each task as Human-led, AI-assisted, or Rule-based automation (H/A/R); assessment of measurement readiness on a four-level ladder (M0 to M3); scoring of confidence and risk across six dimensions; explicit reasoning over cost-of-inaction and cost-of-inefficiency; and smart governance checkpoints at three gating points (G1 before pilot, G2 before scale, G3 before full delegation). The framework produces a structured per-task decision in one of five states, namely Adopt-Now, Adopt-with-Conditions, Defer-Pending-Measurement, Do-Not-Automate, or Human-Retain, each accompanied by a justification trace.

None of these underlying claims is novel on its own. The literatures cited above already establish that work is profitably analyzed task by task, that AI performance varies substantially across tasks, and that AI deployment requires governance. What this paper adds is their combination: task classification, measurement readiness, confidence and risk assessment, cost-of-inaction and cost-of-inefficiency reasoning, and governance checkpoints brought together in a single decision-support artifact, framed for operational AI governance inside organizations and explicitly mappable to the NIST AI RMF [4], Regulation (EU) 2024/1689 [5], and ISO/IEC 42001:2023 [6].

The remainder of the paper is organized as follows. Section 2 reviews five literature streams and identifies the integrative gap that motivates the artifact. Section 3 articulates seven design objectives derived from this gap. Section 4 specifies the conceptual design-science methodology. Section 5 presents the TC-MAGS framework in detail. Section 6 demonstrates the framework through an illustrative scenario in municipal HR onboarding. Section 7 provides an analytical evaluation against the design objectives. Section 8 discusses theoretical and practical implications. Section 9 acknowledges limitations and outlines an empirical agenda. Section 10 concludes.

2. Related Work

The integrative gap that motivates TC-MAGS becomes visible when several adjacent literatures are viewed in succession. Five streams each illuminate part of the problem of task-level AI adoption and governance, yet none provides the per-task decision artifact that organizations require in practice. This section reviews these streams in turn and concludes with a short synthesis.

2.1 Task-technology fit and task-based work analysis

The conceptual case for treating the task as the analytical unit has a long history. Goodhue and Thompson [9] demonstrated that a technology improves individual performance only when its functionality fits the characteristics of the task and is in fact utilized; the task, not the technology in isolation, was identified as the locus at which performance effects materialize. In labor economics, Autor and colleagues [7] proposed the task-based view of technological change, arguing that occupations are better understood as bundles of tasks with distinct cognitive and routine content. Acemoglu and Restrepo [8] extended this perspective by formalizing the displacement and reinstatement effects of automation, showing that new tasks created by technology are as consequential as the tasks it eliminates.

The arrival of large-scale machine learning has prompted a second wave of task-grained analysis. Felten and colleagues [2] developed the AI Occupational Exposure index, mapping the proximity of occupational tasks to AI capabilities. Eloundou and colleagues [1] provided task-level exposure estimates for large language models. Dell'Acqua and colleagues [3], drawing on controlled field-experimental evidence, characterized AI's productivity effects as forming a "jagged frontier": within a single workflow, some tasks benefit substantially from AI assistance while others see degraded outcomes, with little obvious pattern at the workflow level. Taken together, these works establish that the task is the appropriate analytical unit and that within-workflow heterogeneity is empirically substantial. They do not, however, supply organizations with a decision artifact that translates this analytical insight into per-task adoption choices.

2.2 AI readiness and maturity

A parallel literature has developed around organizational preparedness for AI. Jöhnk and colleagues [10] identify a multi-dimensional set of AI readiness factors spanning strategic alignment, resources, knowledge, culture, data, and governance. Uren and Edwards [11] frame technology readiness as a journey, tracing organizational progression along an adoption pathway. Both contributions operate at the organization or capability level: readiness is assessed for the organization as a whole, rather than for a specific task targeted for adoption. In the public-administration setting, van Noordt and Tangi [12] examine how AI capability development translates into public value creation, finding that outcomes are mediated by organizational and task-level factors. For municipal service providers participating in smart-city ecosystems, this is consequential: organizational-level readiness does not by itself determine whether a specific citizen-facing task is suitable for AI adoption, and the determination must be made at finer granularity than the readiness literature typically affords.

2.3 Measurement and construct validity

The word "measurable" in the framework's title carries a methodological burden. If adoption decisions are to be conditioned on the measurement readiness of a task, the constructs through which readiness, confidence, and risk are operationalized must themselves be defensibly defined. MacKenzie and colleagues [13] provide the canonical procedure for construct measurement and validation, distinguishing conceptualization, development, model specification, scale evaluation, and validation. Their procedure makes clear that the difference between a construct that supports decision-making and one that does not is the rigor of its definition and the calibration of its indicators. This concern applies directly to the M0 to M3 readiness ladder and to the six confidence and risk dimensions introduced in Section 5.

2.4 AI governance and responsible AI

The governance literature has grown rapidly alongside organizational AI adoption. Mäntymäki and colleagues [14] situate governance as the system of structures, processes, and mechanisms through which an organization steers its AI activities. Birkstedt and colleagues [15] synthesize the literature thematically and identify substantial gaps, particularly around operational mechanisms and the translation of principles into practice. Papagiannidis and colleagues [16] develop a research framework for responsible AI governance, integrating structural, procedural, and relational practices. Shneiderman [17] articulates human-centered AI principles, arguing that reliability, safety, and trustworthiness require integration of governance structures at multiple organizational levels.

Three instruments now structure responsible AI practice in operational settings. The NIST AI Risk Management Framework [4] defines four functions, Govern, Map, Measure, and Manage, to be applied iteratively across the AI lifecycle. Regulation (EU) 2024/1689 [5] establishes a risk-based regime classifying AI systems into prohibited, high-risk, limited-risk, and minimal-risk categories. ISO/IEC 42001:2023 [6] specifies the requirements for an AI management system. For public-sector bodies and municipal service providers, these instruments are particularly load-bearing: many digital public services involve heightened accountability where decisions affect citizens. Yet none specifies a per-task decision logic for adoption. The governance literature supplies the principles, structures, and obligations; it does not supply the decision substrate at which these are operationalized task by task.

2.5 Conceptual articles and design science

Because TC-MAGS is presented as a conceptual artifact rather than an empirical study, the methodological literature on conceptual and design-science research provides the appropriate scholarly basis. Design-science research [18] distinguishes itself from explanatory behavioral research by producing purposeful artifacts intended to address identified organizational problems. Peffers and colleagues [19] decompose the design-science process into six activities. Gregor and Hevner [20] clarify how to position design-science contributions, distinguishing knowledge contributions according to the maturity of problem and solution. Venable and colleagues [21] provide the FEDS framework for selecting evaluation strategies in design science, supporting ex-ante analytical evaluation prior to empirical instantiation. Jaakkola [22] identifies four approaches to conceptual articles and provides guidance for model-type contributions. Together, these sources establish that a carefully constructed conceptual artifact, grounded in cited kernel theories and communicated with explicit scope statements about what has and has not been demonstrated, constitutes a legitimate scholarly contribution.

2.6 Synthesis

Each of the five streams addresses part of the task-level AI adoption and governance problem. The task-based literatures establish that adoption is task-grained but provide no organizational decision artifact. The readiness literature characterizes organizational preconditions but does not descend to the per-task adoption decision. The measurement literature supplies procedural standards for construct definition but does not, by itself, produce a decision substrate. The governance literature offers principles, structures, and instruments but leaves the operational task-level decision logic underdetermined. The design-science methodological literature licenses the conceptual-artifact response to such an integrative gap. The gap is not the absence of any single component; it is the absence of an artifact that integrates task classification, measurement readiness, confidence and risk assessment, cost reasoning, and governance checkpoints into one per-task decision-support pipeline, mappable upward to existing instruments and suitable for operational use inside organizations, including public-sector bodies for which task-level accountability is most consequential. The five streams and the gap addressed by TC-MAGS are summarized in Table 1.

3. Research Gap and Design Objectives

Section 2 concluded that the gap motivating TC-MAGS is integrative rather than componential. The task-based literatures establish that adoption is task-grained and that within-workflow heterogeneity is

Table 1: Literature streams and the integrative gap addressed by TC-MAGS

Literature stream	What it covers	What it leaves open	TC-MAGS response
Task-technology fit / task-based labor economics [7 to 9, 1 to 3]	Task as unit of analysis; within-workflow heterogeneity	No organizational decision artifact	L1 plus L2 (inventory and H/A/R)
AI readiness and maturity [10 to 12]	Organization-level preconditions for adoption	No per-task decision logic	L3 (M0 to M3 as a gate)
Construct measurement [13]	Procedural standards for constructs	Not a decision substrate	Construct anchoring of M-levels and risk dimensions
AI governance and responsible AI [4 to 6, 14 to 17]	Principles, structures, instruments	No per-task operational substrate	L4 plus L6 plus governance checkpoints
Design-science methodology [18 to 22]	Justification for artifact-building contributions	Does not produce the artifact itself	Methodological frame for TC-MAGS

empirically substantial; the readiness literature characterizes organizational preconditions; the measurement literature supplies procedural standards; the governance literature articulates principles, structures, and instruments; and the design-science methodological literature licenses conceptual artifacts that respond to such gaps. None of these literatures, however, supplies an artifact that integrates these dimensions into one per-task decision-support pipeline. The task-centric view is not new by itself; its integration with measurement, risk, cost, and governance into a single decision artifact is.

This integrative gap can be operationalized as seven design objectives. Each subsequent component of TC-MAGS, specified in Section 5, traces to one of these.

DO1: Task as unit of analysis. Adoption decisions must be produced for discrete tasks, each defined as an observable unit of work with identifiable inputs, outputs, executor, and completion record [7, 9, 1]. The artifact must avoid collapsing adoption decisions into workflow-level, system-level, or firm-level verdicts.

DO2: Explicit classification across work modes. Each task must be classified as Human-led (H), AI-assisted (A), or Rule-based automation (R), with rationale grounded in observable task characteristics. This objective is motivated by the displacement-and-reinstatement framing [8] and within-workflow heterogeneity evidence [3].

DO3: Measurement readiness as a gate. Each task must be assessed for measurement readiness on a defined ladder (M0 to M3) before an adoption decision is authorized [13, 10, 11], so decisions are conditioned on the availability of a substrate for evaluation.

DO4: Confidence and risk assessed at task level. Each task must be scored across six dimensions: data quality confidence, model and output confidence, stakeholder impact, reversibility, regulatory exposure, and ethical sensitivity [4, 5, 17]. Risk and confidence are operationalized at the level at which they materialize rather than aggregated to the system level.

DO5: Cost-of-inaction and cost-of-inefficiency as twin economic signals. The framework must distinguish the expected loss from not adopting AI for a measurable task (cost of inaction) from the present-state cost of the current human or rule-based regime (cost of inefficiency). Adoption may be considered more defensible when these two costs together exceed the all-in cost of an AI-assisted regime, subject to risk and governance constraints.

DO6: Governance checkpoints with documented justification. Adoption decisions must pass governance gates at three points, before pilot (G1), before scale (G2), and before full delegation (G3), and each decision must carry a justification trace [6, 4, 14, 15, 16].

DO7: Mappability to external governance instruments. Framework outputs must be mappable upward to the NIST AI RMF [4], Regulation (EU) 2024/1689 [5], and ISO/IEC 42001:2023 [6] without

replacing them. Organizations adopting TC-MAGS are not required to choose between operational decision support and compliance with existing instruments.

Taken individually, none of these objectives constitutes a novel contribution. The contribution of TC-MAGS lies in their integration into one conceptual decision-support artifact in which each objective conditions the others: measurement readiness gates economic reasoning, risk and confidence condition governance decisions, and governance checkpoints discipline the entire pipeline through a justification trace mappable upward to external instruments. In the terms of Gregor and Hevner [20], this is a design-science Improvement contribution: a known problem addressed through a new, structured artifact.

4. Methodology: Conceptual Design Science Approach

This paper follows a conceptual design-science approach. Design Science Research (DSR), as articulated in the information-systems tradition [18, 19], differs from explanatory empirical research in its primary aim and primary output. Explanatory research seeks to describe, predict, or causally account for organizational phenomena, typically by formulating and testing hypotheses against observed data. Design science seeks to address an identified organizational problem by creating and justifying an artifact, whether a construct, model, method, or instantiation, that did not previously exist in a form sufficient to address the problem. Both modes are legitimate scholarly contributions; they differ in what they produce rather than in their standards of rigor.

The approach is appropriate for TC-MAGS because the problem identified in Sections 1 through 3 is not a causal one. The integrative gap is not a question about whether some factor predicts AI adoption success. It is, rather, the absence of an artifact: existing literatures supply partial constructs, but no integrated per-task decision-support artifact through which an organization can produce a defensible adoption decision for a specific task. Such a gap is, in design-science terms, an artifact-absence problem, and the appropriate contribution is the construction and justification of the artifact itself.

The paper maps to the design-science research methodology proposed by Peffers and colleagues [19], which decomposes the design process into six activities. The mapping of these activities to the sections of this manuscript is shown in Figure 1.

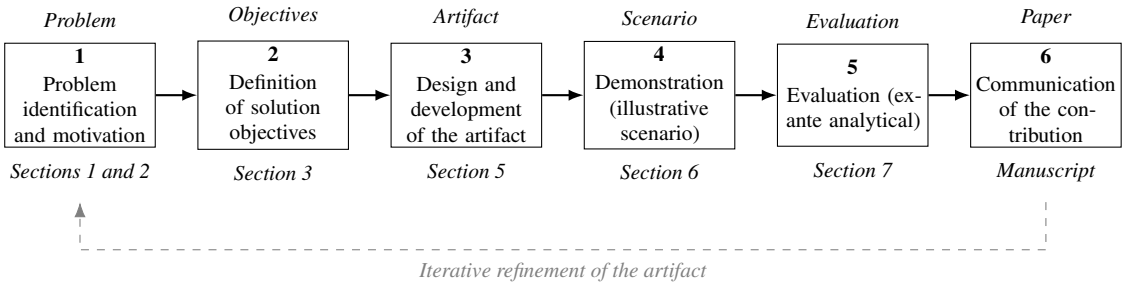

Figure 1: Mapping of the DSRM activities [19] to the sections of this manuscript

Problem identification and motivation are executed in Sections 1 and 2 through the unit-of-analysis mismatch and the integrative gap identified across five literatures. Definition of solution objectives is executed in Section 3 through the seven design objectives. Design and development is executed in Section 5, which specifies the framework's six layers, core constructs, decision logic, justification trace, and mappings to external instruments. Demonstration is executed in Section 6 through an illustrative scenario in municipal HR onboarding; the scenario is presented as a worked thought experiment to show how the framework's layers produce differentiated decisions, not as a case study, pilot, or empirical instantiation. Evaluation is executed in Section 7 as ex-ante analytical evaluation against the design objectives. Communication is the present manuscript.

The artifact produced is, in the typology offered by Hevner and colleagues [18], principally a method/model artifact. It comprises a set of constructs, namely task, the H/A/R classification, the M0 to M3 measurement readiness ladder, the six confidence and risk dimensions, cost-of-inaction and

cost-of-inefficiency, the G1/G2/G3 governance checkpoints, and the five decision-output states, together with a procedure that prescribes how these constructs combine to produce a per-task adoption decision accompanied by a justification trace.

The knowledge contribution is positioned using the framework of Gregor and Hevner [20]. TC-MAGS occupies the Improvement quadrant: the problem is known and substantial, and the solution is new in its structured integration of components previously scattered. The choice to present the artifact as a conceptual article is supported by Jaakkola [22], who identifies four approaches: theory synthesis, theory adaptation, typology, and model. TC-MAGS is best characterized as a model with two embedded typologies (H/A/R and M0 to M3).

Evaluation is ex-ante and analytical in the sense distinguished by Venable and colleagues [21]. The framework is assessed against its design objectives for internal coherence, theoretical grounding, and practical plausibility, rather than against observed outcomes in a deployed setting. FEDS explicitly recognizes such evaluation as a legitimate stage prior to empirical instantiation. Empirical validation is deferred to future work (Section 9). The claims made here are design-level claims about conceptual coherence and theoretical grounding, not performance claims; the paper does not claim that TC-MAGS has been implemented, tested, or proven to improve organizational outcomes.

5. Proposed Framework

TC-MAGS is a conceptual method/model artifact responding to the integrative gap identified in Section 2 and to the seven design objectives stated in Section 3. The framework comprises six layers operating as a pipeline: each layer consumes the outputs of the preceding layer and contributes a distinct reasoning dimension to the per-task decision-support output produced by Layer 6. The contribution lies in the integration of these layers into one coherent decision substrate. No single layer is novel in isolation; what TC-MAGS provides is the structure in which these dimensions condition one another and produce a defensible, traceable adoption decision.

5.1 Core Constructs

Task. A discrete, observable unit of work with identifiable inputs, outputs, an executor, and a record of completion. The task is the smallest unit at which adoption decisions are produced in TC-MAGS.

Task inventory. A structured enumeration of tasks within a designated scope, such as a process, role, service, or organizational unit. The inventory is the foundational artifact on which subsequent layers operate.

H/A/R classification. Human-led (H) tasks are those requiring judgment, tacit knowledge, ethical or relational stakes, or accountability that should remain with a human agent. AI-assisted (A) tasks are those in which probabilistic AI provides support, such as drafting, summarization, classification, recommendation, or pattern detection, subject to human validation. Rule-based automation (R) tasks are deterministic, codifiable, and low in ambiguity, suitable for workflow rules rather than AI.

Measurement readiness (M0 to M3). M0 indicates the absence of any baseline. M1 indicates anecdotal or qualitative evidence only. M2 indicates that a quantitative metric exists but is not standardized. M3 indicates a standardized, validated, repeatable metric with baseline data sufficient for comparison.

Confidence and risk dimensions. Six dimensions: data quality confidence; model/output confidence; stakeholder impact; reversibility; regulatory exposure; ethical sensitivity.

Cost-of-Inaction (CoI-A) and Cost-of-Inefficiency (CoI-E). Two distinct economic quantities. CoI-A is the expected loss from not adopting AI for a measurable task. CoI-E is the present-state cost of the current human or rule-based regime. The two are conceptually distinct: a task may have high CoI-E but low CoI-A, or vice versa. No quantitative formula is prescribed; the reasoning is qualitative or structured-comparative.

Smart governance checkpoints. G1 (before pilot): data adequacy, ethical review, scope definition, accountability assigned. G2 (before scale): pilot evidence, residual risk, operational readiness. G3 (before full delegation): audit trail, fallback design, ongoing oversight.

Decision-support output states. Five states: Adopt-Now (task is ready under current conditions); Adopt-with-Conditions (authorization subject to mitigations); Defer-Pending-Measurement (adoption deferred until readiness improves); Do-Not-Automate (risk, regulatory, or ethical properties preclude adoption); Human-Retain (the task should remain human by design intent).

5.2 Layered Architecture

Figure 2 shows the six-layer architecture of TC-MAGS, the three governance checkpoints (G1 to G3), and the five decision-output states. Each layer is specified below by its purpose, input, output, and theoretical grounding.

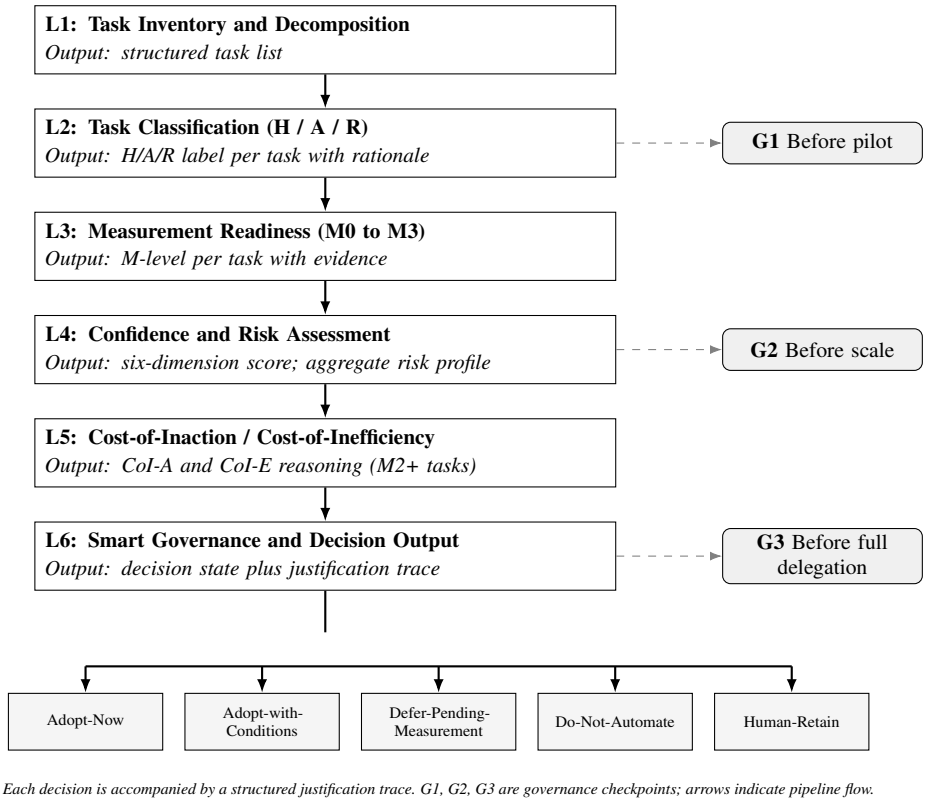

Figure 2: Layered architecture of TC-MAGS, with governance checkpoints (G1 to G3) and the five decision-output states

L1: Task Inventory and Decomposition. Renders the work within scope visible at the granularity at which adoption decisions can be made. Input: process maps, role descriptions, system logs, SOPs, stakeholder accounts. Output: a structured task list satisfying the four properties of the task construct. Grounded in [7, 9]. Feeds L2.

L2: Task Classification (H/A/R). Assigns each task to one of three work-mode categories. Input: the task list from L1. Output: an H/A/R label per task with rationale referencing codifiability, data availability, feedback loop, stakes, tacit-knowledge content, and reversibility. Grounded in [8, 3]. Feeds L3.

L3: Measurement Readiness Assessment. Determines whether each task possesses the measurement substrate to support a defensible adoption decision. Input: the classified task list with available evidence on existing metrics. Output: an M0 to M3 score per task with evidence. Grounded in [13, 10, 11]. Feeds L4 and is consulted directly by L6's decision logic.

L4: Confidence and Risk Assessment. Scores each task across the six confidence and risk dimensions and produces an aggregate profile. Input: task description, data inventory, regulatory mapping, stakeholder analysis. Output: six-dimension scoring and aggregate risk profile, with an

indication of which risk tier under [5] the task most plausibly occupies. Grounded in [4, 5, 17]. Feeds L5 and L6.

L5: Cost-of-Inaction / Cost-of-Inefficiency Reasoning. Surfaces the twin economic signals bearing on adoption. Input: readiness profile from L3, risk profile from L4, and available operational data. Output: qualitative or structured-comparative estimates of CoI-A and CoI-E. Applied only to M2+ tasks; for M0/M1, cost reasoning is acknowledged as unsupported and deferred. Feeds L6.

L6: Smart Governance and Decision-Support Output. Integrates the outputs of the preceding layers and produces the per-task decision. Input: H/A/R, M-level, risk profile, cost reasoning. Output: one of the five decision states with a justification trace and an indication of the applicable governance checkpoint. Grounded in [6, 4, 14, 15, 16].

5.3 Decision Logic

The decision logic by which Layer 6 combines the prior layers' outputs is conceptual and explanatory rather than algorithmic. It can be expressed narratively as a small number of routing patterns. A task classified A or R, scored at M2 or M3, with an aggregate risk profile within the organization's stated tolerance and with clear accountability assigned, may receive Adopt-Now once G1 has been satisfied. A task with otherwise promising characteristics but with a confidence or risk dimension that exceeds tolerance, for example high stakeholder impact with only moderate model confidence, may receive Adopt-with-Conditions, with conditions specifying the human-in-the-loop arrangement, monitoring frequency, or scope limitation. A task scored at M0 or M1 generally receives Defer-Pending-Measurement: without a measurement substrate, an adoption decision cannot be defensibly evaluated. A task with unacceptable scores on regulatory exposure, ethical sensitivity, or reversibility receives Do-Not-Automate. A task whose human content is the value proposition itself, such as empathy, judgment, or relational accountability, receives Human-Retain as a matter of design intent.

This decision logic is presented as conceptual routing, not as a validated algorithm. Organizations applying TC-MAGS will calibrate the boundaries between these states to their own risk tolerance, regulatory environment, and operational context.

5.4 Justification Trace

Each decision-output state is accompanied by a structured justification trace recording the inputs to the decision: task description and inventory reference; H/A/R classification with rationale; M-level score with supporting evidence; six-dimension confidence and risk profile; cost reasoning (where applicable); governance checkpoint status; named human reviewer or accountable role; and final decision state.

The justification trace serves three functions. It supports auditability: a third party reviewing an adoption decision can reconstruct the basis on which it was made. It supports organizational accountability: the named accountable role makes governance an executed practice rather than a paper commitment. And it supports what may reasonably be called technology culture: by making the reasoning behind each adoption decision explicit, the trace embeds a culture of measurement, deliberation, and justification into AI adoption practice itself, rather than leaving these as abstract values asserted at the policy level. For public-sector and digital-service-delivery organizations where decisions affect citizens, the trace also provides the evidentiary basis on which transparency obligations can be discharged.

5.5 Mapping to Governance Instruments

TC-MAGS is designed to operate as a complement to existing governance instruments, not as a substitute for them. Three mappings are sketched below.

Mapping to the NIST AI Risk Management Framework [4]. Layer 4 operationalizes the MAP function (context, categorization, impact mapping). Layers 3 and 5 jointly operationalize MEASURE (the readiness ladder and cost reasoning produce the metric basis on which trustworthiness characteristics can be evaluated). Layer 6 operationalizes MANAGE through the decision-state output and the checkpoint structure. The whole pipeline is bounded by GOVERN, which sets policy, accountability, and risk-

tolerance parameters. NIST explicitly does not prescribe risk tolerance [4]; TC-MAGS provides the granular task-level substrate at which an organization's tolerance is operationalized.

Mapping to Regulation (EU) 2024/1689 [5]. Layer 4's regulatory-exposure dimension supports task-level identification of where a task falls within the Regulation's four-tier risk taxonomy. Because a single AI system typically implements many tasks of differing risk, task-level identification supports proportionate compliance: high-risk task uses may trigger additional obligations, including risk-management, documentation, human-oversight, or conformity-related requirements, depending on the organization's role under the Regulation. TC-MAGS does not replace legal compliance; it provides the granular task-level analysis on which proportionate compliance is built.

Mapping to ISO/IEC 42001:2023 [6]. The task inventory and risk profile from Layers 1 and 4 can populate the AI system inventory and impact assessment required by an AI management system. The decision-support outputs of Layer 6 can populate operational planning and control records. The justification traces satisfy documented-information requirements. The G1/G2/G3 structure supports the continual-improvement orientation of the management system.

Across the three mappings, the role of TC-MAGS is to provide a structured task-level reasoning surface on which these instruments can be operationalized. This mapping is intended as a governance-screening aid, not as a legal classification or compliance determination; jurisdictional applicability and final compliance determinations remain the responsibility of the implementing organization and its legal counsel.

Mapping to national or sector-specific AI governance instruments in other jurisdictions can be treated as a future extension; this paper does not invent jurisdiction-specific requirements that it cannot verify.

5.6 Summary Tables

Tables 2 through 5 summarize the H/A/R classification criteria, the M0 to M3 ladder, the confidence and risk dimensions, and the governance checkpoints respectively.

Table 2: H/A/R classification criteria

Criterion	Human-led	AI-assisted	Rule-based automation
Codifiability	Low; requires judgment	Partial; AI augments judgment	High; fully codifiable
Data availability	Often sparse or contextual	Sufficient for probabilistic support	Sufficient for deterministic rules
Feedback loop	Slow, qualitative, relational	Moderate; requires human validation	Fast, deterministic, verifiable
Stakes	High ethical, relational, or accountability stakes	Moderate stakes manageable with human review	Low ambiguity; low stakes per instance
Tacit knowledge	High	Moderate; AI complements explicit knowledge	Low
Reversibility	Often low; errors hard to correct	Moderate; errors caught by human review	High; rules can be amended or rolled back

6. Illustrative Scenario: Municipal HR Onboarding

The scenario in this section is hypothetical and illustrative. It demonstrates how the six layers of TC-MAGS produce differentiated decisions within a single workflow, in a setting relevant to public-sector and digital-service-delivery organizations operating within smart-city ecosystems. No real organization, dataset, interview, or measured outcome underlies the scenario. It is not a case study, pilot, or empirical validation; it is a worked thought experiment in the sense recognized by the design-science methodological literature [19, 22] for the demonstration activity prior to empirical instantiation.

Table 3: M0 to M3 measurement readiness ladder

Level	Meaning	Typical evidence	Adoption implication
M0	No baseline	No metric defined; no historical record	Defer; build measurement substrate first
M1	Anecdotal or qualitative evidence only	Reports, interviews, informal accounts	Defer; quantify before authorizing adoption
M2	Quantitative metric exists but is not standardized	Local metric; variable definition across instances	Eligible for cost reasoning and adoption with caution
M3	Standardized, validated, repeatable metric with baseline	Definition in data dictionary; fixed cadence; baseline available	Fully eligible for adoption decision

Table 4: Confidence and risk dimensions

Dimension	Key question	Governance implication
Data quality confidence	Are the data underlying this task reliable and representative?	Conditions readiness scoring and pilot scope
Model/output confidence	How reliable are AI outputs for this task type?	Conditions whether human-in-the-loop is mandatory
Stakeholder impact	How materially does the task affect affected parties?	Raises the threshold for adoption and oversight
Reversibility	Can errors be corrected after the fact?	Low reversibility constrains automation level
Regulatory exposure	What governance obligations apply to this task?	Determines applicable tier and oversight duties
Ethical sensitivity	Are there fairness, dignity, or autonomy concerns?	May warrant Human-Retain or Do-Not-Automate

Table 5: Smart governance checkpoints

Checkpoint	Timing	Main question	Typical output
G1	Before pilot	Are data adequacy, scope, ethical review, and accountability assigned?	Pilot authorization with stated scope and metrics
G2	Before scale	Does pilot evidence support broader rollout under acceptable residual risk?	Scale authorization with operational readiness package
G3	Before full delegation	Are audit trail, fallback design, and ongoing oversight in place?	Delegation authorization with monitoring regime

The setting is a municipal digital-services department or smart-city operations unit responsible for onboarding new staff into roles spanning citizen-service platforms, urban-data operations, IT support, and digital public-service delivery. The onboarding process involves several organizational actors, including HR, IT, security, departmental managers, and compliance officers, and combines administrative, technical, training, and relational components. Such a setting is representative of operational environments in which task-level AI adoption decisions arise with elevated accountability stakes, because the workforce being onboarded will subsequently make decisions affecting citizens.

Seven tasks constitute the onboarding workflow under consideration: T1, collect onboarding documents and required forms; T2, confirm policy acknowledgments including data-handling rules and public-service conduct; T3, provision IT accounts, devices, access rights, and digital-service tools; T4, assign role-specific training modules; T5, schedule and conduct manager check-ins and mentoring conversations; T6, confirm compliance attestations for data privacy, cybersecurity, and public-service obligations; T7, collect and summarize onboarding feedback. Each task is processed through the TC-MAGS layers. The application is summarized in Table 6.

Table 6: Illustrative TC-MAGS application to municipal HR onboarding (L = Low, M = Medium, H = High risk; decision-state abbreviations defined in the note below the table)

Task	**H/A/R**	**M-level**	**Risk**	**Decision**	**Justification**
T1 Documents	R	M3	M	AN	Deterministic intake with standardized validation; identity verification subject to data-protection oversight.
T2 Policy acknowledgments	A	M2	M	AWC	AI summarization aids comprehension; human confirmation retained for accountability.
T3 IT provisioning	R	M3	H	AWC	Mature rule-based provisioning, but access-rights exposure requires separation-of-duties and audit logging.
T4 Training assignment	A	M1	M	DPM	AI can recommend training paths, but learning-outcome metrics are weak; build measurement substrate first.
T5 Manager check-ins	H	M1	M	HR	Relational and developmental task; value lies in human attention and judgment.
T6 Compliance attestations	R	M3	H	AWC	Deterministic attestation, but regulatory exposure requires governance gate and trace.
T7 Feedback collection	A	M1	M	DPM	Sentiment summarization is feasible, but validated indicators are not established.

Note. AN = Adopt-Now; AWC = Adopt-with-Conditions; DPM = Defer-Pending-Measurement; DNA = Do-Not-Automate; HR = Human-Retain.

Within one onboarding workflow, the seven tasks receive five distinct decision states. A tool-centric question framed at the workflow level, should we use AI in onboarding, would necessarily produce a single answer that conceals this differentiation. The same question framed at the task level produces a structured pattern in which some tasks are adopted, some are adopted with conditions, some are deferred until measurement substrate is built, and one is retained as human work by design. This pattern is consistent with the within-workflow heterogeneity that Dell'Acqua and colleagues [3] document empirically; TC-MAGS provides the organizational decision substrate at which such heterogeneity can be acted upon rather than averaged away.

The governance implications are not trivial. Public-service onboarding involves access rights, personal data, cybersecurity exposure, and downstream accountability for decisions affecting citizens. Even tasks classified as rule-based automation, namely T1, T3, and T6, require governance checkpoints and a justification trace. T3 in particular is rule-based but high-risk because access provisioning

determines what the new staff member can subsequently see and do; the Adopt-with-Conditions decision reflects the governance overlay rather than automation difficulty. The framework's mapping to ISO/IEC 42001:2023 documented-information requirements [6], to NIST AI RMF governance and management functions [4], and where applicable to the risk-tier obligations of Regulation (EU) 2024/1689 [5], applies to these decisions in the same way it applies to AI-assisted ones. Governance is not reserved for AI; it is conditioned on risk.

The measurement implication is equally important. Tasks T4 and T7 receive Defer-Pending-Measurement not because they are unsuitable for AI assistance in principle, but because the organization lacks the measurement substrate to evaluate whether such assistance is achieving its intended outcome. The deferral is not a rejection. It is a recommendation that measurement work, meaning defining indicators, establishing baselines, and instrumenting collection, precede the adoption decision. This makes measurement readiness a gate in the adoption pathway rather than an afterthought. For public-sector organizations under heightened accountability obligations, this ordering matters: deploying AI without the measurement substrate to evaluate it leaves the organization unable to demonstrate either utility or safety after the fact.

The scenario illustrates the internal logic of the framework specified in Section 5 and prepares the ground for analytical evaluation against the design objectives of Section 3, which follows in Section 7.

7. Analytical Evaluation

This section provides an ex-ante analytical evaluation of TC-MAGS. The purpose is to assess whether the artifact specified in Section 5 and illustrated in Section 6 exhibits internal coherence, aligns with the seven design objectives derived in Section 3, and satisfies the methodological expectations of the design-science tradition. The evaluation does not assess real-world utility, organizational outcomes, or comparative performance against alternative approaches. Such assessments require empirical instantiation and are deferred to the future research agenda in Section 9.

The evaluation orientation can be located precisely using the FEDS framework [21], which distinguishes evaluation strategies along two axes: formative versus summative, and artificial versus naturalistic. The present evaluation is formative, directed at improving and clarifying the artifact rather than rendering a final verdict, and it is artificial and ex-ante, conducted prior to organizational instantiation, on the basis of analytical reasoning about the artifact's specification. FEDS explicitly recognizes such evaluation as a legitimate stage in design-science research, particularly where the risks of premature naturalistic evaluation are high or where the artifact is not yet in a form suitable for deployment.

The artifact is evaluated against the seven design objectives in Table 7. Each row records the objective, the framework component responding to it, and an analytical note on coherence with the objective.

The analytical evaluation against the design objectives is supportive: each objective is responded to by a specified framework component, and the components combine coherently into the decision pipeline specified in Section 5. The illustrative scenario in Section 6 illustrates that the pipeline produces differentiated outputs across tasks within a single workflow, which is what DO1 through DO7 jointly require. The scenario is not evidence of utility in deployed practice; it demonstrates the internal logic of the artifact rather than its consequences in an organization.

The artifact may also be assessed briefly against the design-science research guidelines articulated by Hevner and colleagues [18]. TC-MAGS is presented as a designed artifact (Guideline 1); its problem relevance is established in Sections 1 and 2 (Guideline 2); its evaluation is conducted analytically in this section, with empirical evaluation deferred (Guideline 3); its research contribution lies in the integration of five reasoning dimensions into one decision-support artifact (Guideline 4); its research rigor is grounded in cited kernel theories (Guideline 5); its design has proceeded as a search process disciplined by the seven design objectives of Section 3 (Guideline 6); and its communication takes the form of the present manuscript (Guideline 7). The guidelines are addressed at the level appropriate for a conceptual artifact; complete satisfaction of Guideline 3 in particular awaits empirical evaluation.

The knowledge contribution of TC-MAGS is positioned using the framework of Gregor and Hevner

Table 7: Analytical evaluation of TC-MAGS against design objectives

Design objective	Framework response	Analytical evaluation note
DO1: Task as unit of analysis	L1 defines task inventory and decomposition with observable inputs, outputs, executor, and completion record	Satisfies the required unit-of-analysis shift; decisions are produced for tasks, not for systems or workflows
DO2: Explicit classification across work modes	L2 assigns each task to H, A, or R with a rubric-grounded rationale	Distinguishes the three work modes coherently; rationale requirement supports auditability
DO3: Measurement readiness as a gate	L3 applies the M0 to M3 ladder before adoption is authorized; M0/M1 tasks route to Defer-Pending-Measurement	Prevents adoption decisions where the evaluation substrate is absent; readiness functions as a gate
DO4: Confidence and risk at task level	L4 scores six dimensions, with regulatory exposure linked to applicable instruments	Moves risk assessment to the level at which task consequences materialize; aligned with risk-functional governance logic
DO5: CoI-A and CoI-E	L5 reasons separately about CoI-A and CoI-E for M2+ tasks; reasoning is qualitative or structured-comparative	Clarifies economic reasoning without claiming quantified ROI; asymmetric cases are visible rather than collapsed
DO6: Governance checkpoints plus justification trace	L6 applies G1/G2/G3 and produces the structured justification trace from Section 5.4	Converts governance from policy principle into documented decision practice
DO7: Mappability to external instruments	Section 5.5 maps TC-MAGS to NIST AI RMF, Regulation (EU) 2024/1689, ISO/IEC 42001	Complements external instruments without replacing them; supports proportionate compliance at task level

[20]. TC-MAGS occupies the Improvement quadrant: the problem of operationalizing task-level AI adoption and governance is known and substantial, and the solution is new in its structured integration of components previously scattered across multiple literatures. The contribution is integrative rather than empirical; ex-ante analytical assessment can support claims of conceptual coherence and theoretical grounding, but cannot support claims of practical utility in deployment, which are deferred to the empirical agenda outlined in Section 9.

8. Discussion

What TC-MAGS contributes is integration, not new claims. That the task is the appropriate unit at which to analyze technology fit and adoption was made decades ago [9]. That AI's effects vary substantially across tasks within a single workflow has been established more recently, by field-experimental evidence [3] and by the task-exposure literature reviewed in Section 2. That AI requires governance underlies the NIST AI RMF [4], Regulation (EU) 2024/1689 [5], and ISO/IEC 42001:2023 [6]. What is new here is the combination: task classification, measurement readiness, confidence and risk assessment, cost-of-inaction and cost-of-inefficiency reasoning, and governance checkpoints brought into one per-task decision-support artifact. In the design-science taxonomy of Gregor and Hevner [20], this is an Improvement contribution.

Task-technology fit [9] explains why the alignment between technology functionality and task characteristics conditions performance; the jagged technological frontier [3] shows that within a single workflow this alignment varies sharply. Neither, however, prescribes the organizational decision substrate through which an organization recognizes and acts upon this heterogeneity. TC-MAGS provides that substrate. It does not displace the task-technology fit perspective; it operationalizes it for the AI adoption decision under contemporary governance constraints.

The practical implications for operational AI governance follow from this positioning. Organizations

applying TC-MAGS conceptually would shift adoption decisions from implicit, tool-centric framings toward explicit, traceable per-task framings. Measurement readiness would function as a precondition for adoption rather than as a retrofit applied after deployment; tasks at M0 or M1 would be routed to deferral and measurement-building rather than opportunistic experimentation. Governance would move from policy documents to operational gates that produce documented decisions at G1, G2, and G3. Management-system thinking of the kind required by ISO/IEC 42001:2023 [6] would be supported by structured task-level records that populate AI system inventories, impact assessments, and operational control documentation. These are conceptual implications; whether they translate into improved organizational outcomes is an empirical question this paper does not address.

The relevance of these implications is elevated in public-sector and digital-service-delivery contexts. Municipal organizations participating in smart-city ecosystems make decisions about AI adoption that affect citizen access to services, data rights, fairness of administrative outcomes, and public trust. Empirical evidence from public administration indicates that AI capability development is uneven and that public value outcomes are mediated by organizational and task-level factors [12]. For such organizations, the per-task decision substrate, the explicit confidence and risk scoring, and the justification trace are not procedural overhead; they are the operational means by which accountability obligations are discharged. In this sense, the framework treats smart governance not as a city-wide abstraction, but as a set of task-level organizational routines through which digital public services become measurable, accountable, and governable.

A further implication concerns technology culture. The framework's structural features, namely task-level justification traces, measurement readiness as a gate, and governance checkpoints, embed three cultural orientations into AI adoption practice rather than asserting them at the level of policy. The justification trace embeds a culture of accountability by making the reasoning behind each adoption decision attributable to named roles. The M0 to M3 ladder embeds a culture of measurement by treating the absence of measurement substrate as a reason to wait rather than a permission to proceed. The H/A/R classification and the Human-Retain decision state embed a culture of human oversight by making the choice to retain human work a positive design decision rather than a residual category. Operational AI governance, on this view, ultimately serves human and social ends rather than measurable outputs alone; the framework's role is to make those ends executable through structured organizational practice.

The discussion supports conceptual value and plausible usefulness; it does not establish empirical effectiveness. TC-MAGS complements the NIST AI RMF [4], Regulation (EU) 2024/1689 [5], and ISO/IEC 42001:2023 [6]; it does not replace them, and it does not relieve organizations of obligations imposed by applicable law or adopted management standards. The limitations of the present work and the empirical agenda required to test the conceptual claims made here are addressed in Section 9.

9. Limitations and Future Research

TC-MAGS is a conceptual artifact. It is grounded in cited kernel theories and has been analytically evaluated against its design objectives in Section 7, but it has not been implemented in any organization, tested against alternatives, or shown to alter adoption outcomes or governance compliance. Whether its conceptual coherence translates into utility in deployed practice is an open empirical question. The paper is positioned as the communication step of a design-science process [19], not as the demonstration of realized organizational impact.

A second limitation concerns task granularity. The framework's operation depends on how an organization decomposes the tasks within its scope. Too broad a definition collapses decisions back to the workflow level and hides within-workflow heterogeneity; too narrow a definition generates an inventory whose administrative cost exceeds its decision value. TC-MAGS does not prescribe a universal decomposition rule, and applicable standards likewise defer to organizational context [4, 6]. Future work should examine appropriate levels of granularity across different organizational settings.

A third limitation concerns the scope of demonstration. The municipal HR onboarding scenario in Section 6 illustrates the framework's internal logic; it is not evidence of effectiveness, and it covers

a single domain. Other relevant domains within smart-city ecosystems, such as procurement, citizen-service triage, urban-data monitoring, permitting, and emergency-response support, would require separate study.

A fourth limitation concerns the framework components themselves. The M0 to M3 ladder, the six confidence and risk dimensions, and the G1/G2/G3 checkpoints are conceptually specified but have not been subjected to the construct-measurement procedures described by MacKenzie and colleagues [13]. The anchors are intended to be sufficient to support pilot application; they are not yet calibrated instruments.

The empirical agenda required to address these limitations is identifiable. It includes expert review or a Delphi study to refine the constructs; inter-rater reliability testing for H/A/R and M0 to M3 classifications; multiple case studies in public-sector or digital-service-delivery organizations under formative-naturalistic strategies [21]; action research or design-science instantiation in collaboration with a participating organization; longitudinal comparison of task-level decisions before and after framework use; and adaptation studies that map TC-MAGS to national or sector-specific AI governance instruments where applicable.

Read in this light, these limitations are not weaknesses that undermine the contribution but properties of the kind of contribution this paper makes: a model-oriented conceptual article [22] prepared for empirical instantiation rather than presenting it.

10. Conclusion

This paper began from a problem of unit-of-analysis mismatch. Artificial intelligence is being adopted at a pace few organizational practices have caught up with, yet adoption decisions are typically made at the wrong granularity. Strategy is formulated at the firm level, governance is regulated at the system level, but the value of AI, the risk it carries, the measurement substrate required to evaluate it, and the accountability attached to its outputs all materialize at the task level. The literatures reviewed in Section 2 each illuminate part of this mismatch; none, on its own, supplies the organizational decision substrate at which the mismatch can be addressed.

In response, the paper proposed TC-MAGS, a Task-Centric Framework for Measurable AI Adoption and Smart Governance, as a conceptual design-science artifact. The framework integrates a task inventory, the H/A/R classification of work modes, the M0 to M3 measurement readiness ladder, six-dimensional confidence and risk assessment, cost-of-inaction and cost-of-inefficiency reasoning, three governance checkpoints, and five decision-support output states into one structured per-task decision pipeline. Each decision carries a justification trace and is mappable upward to leading governance instruments, and is positioned as a design-science Improvement contribution [20].

For organizations participating in smart-city ecosystems and delivering digital public services, the question of how to adopt AI in a measurable and accountable way is not solely technical. It is also part of building a technology culture that takes accountability, measurement, and human oversight as constitutive rather than incidental. The framework presented here is offered as a conceptual contribution to that work; the empirical research required to instantiate and test it in real organizational settings remains to be done.

References

[1] Eloundou T, Manning S, Mishkin P, Rock D. GPTs are GPTs: Labor market impact potential of LLMs. Science; 384(6702): 1306-1308, 2024. https://doi.org/10.1126/science.adj0998

[2] Felten EW, Raj M, Seamans R. Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. Strategic Management Journal; 42(12): 2195-2217, 2021. https://doi.org/10.1002/smj.3286

[3] Dell'Acqua F, McFowland E III, Mollick E, Lifshitz-Assaf H, Kellogg KC, Fruzman S, Harrison T, Lakhani KR. Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. Organization Science; 37(2), 2026. https://doi.org/10.1287/orsc.2025.21838

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

[5] European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union; L 2024/1689, 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

[6] International Organization for Standardization, International Electrotechnical Commission. Information technology, Artificial intelligence, Management system. ISO/IEC 42001:2023, 2023. https://www.iso.org/standard/81230.html

[7] Autor DH, Levy F, Murnane RJ. The skill content of recent technological change: An empirical exploration. The Quarterly Journal of Economics; 118(4): 1279-1333, 2003. https://doi.org/10.1162/003355303322552801

[8] Acemoglu D, Restrepo P. Automation and new tasks: How technology displaces and reinstates labor. Journal of Economic Perspectives; 33(2): 3-30, 2019. https://doi.org/10.1257/jep.33.2.3

[9] Goodhue DL, Thompson RL. Task-technology fit and individual performance. MIS Quarterly; 19(2): 213-236, 1995. https://doi.org/10.2307/249689

[10] Jöhnk J, Weißert M, Wyrtki K. Ready or not, AI comes: An interview study of organizational AI readiness factors. Business & Information Systems Engineering; 63(1): 5-20, 2021. https://doi.org/10.1007/s12599-020-00676-7

[11] Uren V, Edwards JS. Technology readiness and the organizational journey towards AI adoption: An empirical study. International Journal of Information Management; 68: 102588, 2023. https://doi.org/10.1016/j.ijinfomgt.2022.102588

[12] van Noordt C, Tangi L. The dynamics of AI capability and its influence on public value creation of AI within public administration. Government Information Quarterly; 40(4): 101860, 2023. https://doi.org/10.1016/j.giq.2023.101860

[13] MacKenzie SB, Podsakoff PM, Podsakoff NP. Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. MIS Quarterly; 35(2): 293-334, 2011. https://doi.org/10.2307/23044045

[14] Mäntymäki M, Minkkinen M, Birkstedt T, Viljanen M. Defining organizational AI governance. AI and Ethics; 2(4): 603-609, 2022. https://doi.org/10.1007/s43681-022-00143-x

[15] Birkstedt T, Minkkinen M, Tandon A, Mäntymäki M. AI governance: Themes, knowledge gaps and future agendas. Internet Research; 33(7): 133-167, 2023. https://doi.org/10.1108/INTR-01-2022-0042

[16] Papagiannidis E, Mikalef P, Conboy K. Responsible artificial intelligence governance: A review and research framework. Journal of Strategic Information Systems; 34(2): 101885, 2025. https://doi.org/10.1016/j.jsis.2024.101885

[17] Shneiderman B. Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. ACM Transactions on Interactive Intelligent Systems; 10(4): 1-31, 2020. https://doi.org/10.1145/3419764

[18] Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Quarterly; 28(1): 75-106, 2004. https://doi.org/10.2307/25148625

[19] Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. Journal of Management Information Systems; 24(3): 45-77, 2007. https://doi.org/10.2753/MIS0742-1222240302

[20] Gregor S, Hevner AR. Positioning and presenting design science research for maximum impact. MIS Quarterly; 37(2): 337-355, 2013. https://doi.org/10.25300/MISQ/2013/37.2.01

[21] Venable J, Pries-Heje J, Baskerville R. FEDS: A framework for evaluation in design science research. European Journal of Information Systems; 25(1): 77-89, 2016. https://doi.org/10.1057/ejis.2014.36

[22] Jaakkola E. Designing conceptual articles: Four approaches. AMS Review; 10(1): 18-26, 2020. https://doi.org/10.1007/s13162-020-00161-0