

AI Control Towers: An Enterprise Operating Model for Autonomous Business Decision-Making for Saudi Vision 2030 Organizations

Choudhry Bilal Mazhar^{1*}, Tahani Muneer Algethami²

¹Digital Analytics & AI Lead, Arabian Agricultural Services Company (ARASCO), Riyadh, Saudi Arabia

²BI & Analytics Administrative, Arabian Agricultural Services Company (ARASCO), Riyadh, Saudi Arabia

*Corresponding Author

DOI: <https://doi.org/10.51244/IJRSI.2026.1313CS029>

Received: 26 September 2026; Accepted: 01 October 2026; Published: 13 October 2026

ABSTRACT

Enterprise artificial intelligence (AI) is moving from isolated predictive models and conversational assistants toward agentic systems that plan, use tools, collaborate, and act within delegated authority. This shift creates a governance gap because organizations can deploy AI agents faster than they establish oversight, decision authority, auditability, and value assurance. This study develops the Artificial Intelligence Control Tower (AICT), a governance-driven enterprise operating model for orchestrating and assuring autonomous business decision-making. Design Science Research is used to synthesize enterprise AI, multi-agent systems, decision intelligence, responsible AI, and major international governance frameworks. The resulting artifact comprises six integrated layers: enterprise data and knowledge foundation, AI workforce and agent registry, orchestration and workflow control, decision assurance, governance and compliance, and executive value oversight. Its control logic is formalized through nine equations covering aggregated assurance confidence, decision risk, autonomy assignment, graded policy admissibility, execution authority, expected net value, governance coverage, and maturity. The framework also specifies an eight-stage decision lifecycle, four risk-tiered autonomy levels, a five-level maturity model, and a balanced performance system. An ex ante, artificial, formative evaluation applies eight criteria and five illustrative enterprise scenarios. This conceptual assessment indicates that the architecture can represent ownership, evidence, policy checks, escalation, and outcome measurement; it does not establish causal performance improvement or validated deployment outcomes. The study contributes an integrated enterprise architecture, an explicit decision-governance model, seven testable propositions, an implementation roadmap, and a measurement agenda. Saudi Vision 2030 is treated as an application context, with the framework mapped to relevant national data, privacy, cybersecurity, and AI-governance expectations that would require sector-specific validation before implementation.

Keywords: Agentic Artificial Intelligence, AI Governance, Autonomous Decision-Making, Decision Assurance, Enterprise AI Operating Model

INTRODUCTION

Artificial intelligence (AI) has moved from a specialized analytical capability to a general-purpose enterprise technology. Machine learning, natural-language systems, generative AI, and intelligent automation are now embedded in customer operations, finance, procurement, supply chains, human resources, cybersecurity, and executive decision support. The economic rationale is substantial: generative AI alone has been estimated to create trillions of dollars in annual value across business functions, with much of the opportunity arising from productivity, faster knowledge work, and improved decision processes (Chui et al., 2023). Yet value realization remains uneven because many organizations scale access to AI faster than they redesign processes, governance, and accountability. Deloitte's enterprise study indicates that agentic adoption is expected to grow

sharply while mature governance for autonomous agents remains limited (Deloitte, 2026).

Agentic AI changes the nature of the enterprise problem. A conventional model generates a prediction, classification, or recommendation within a bounded application. An AI agent can interpret a goal, plan a sequence of tasks, use tools, retrieve data, communicate with other agents, and take actions within delegated authority. A multi-agent system may involve monitoring, diagnostic, forecasting, recommendation, compliance, and execution agents that jointly influence a business outcome. Consequently, risk no longer arises only from model error. It may emerge from agent interaction, task delegation, authority expansion, conflicting objectives, tool misuse, accumulated reasoning errors, or inadequate escalation.

Existing AI governance is essential but often fragmented. Model inventories, validation procedures, privacy controls, responsible AI principles, cybersecurity reviews, and data governance committees typically operate as separate mechanisms, and few organizations can state where AI governance sits within their wider corporate, IT, and data governance landscape (Mäntymäki et al., 2022). The NIST AI Risk Management Framework provides a practical structure through the functions Govern, Map, Measure, and Manage, and its generative AI profile extends that structure to confabulation, information integrity, data provenance, and human–AI configuration risks (National Institute of Standards and Technology, 2024; Tabassi, 2023). ISO/IEC 42001 establishes requirements for an organizational AI management system, while ISO/IEC 23894 supplies lifecycle guidance for AI-specific risk management (International Organization for Standardization, 2023a, 2023b). The OECD AI Principles and the Microsoft Responsible AI Standard emphasize human-centered values, transparency, robustness, safety, privacy, fairness, and accountability (Microsoft, 2022; Organisation for Economic Co-operation and Development, 2024), and the European Union Artificial Intelligence Act converts comparable expectations into binding, risk-tiered obligations for providers and deployers of high-risk systems, including human oversight, logging, and post-market monitoring (European Parliament and Council of the European Union, 2024). These foundations are necessary, but reviews of responsible AI consistently report that principles remain difficult to operationalize in day-to-day organizational practice (Papagiannidis et al., 2025). Enterprises therefore require an operational mechanism that applies these expectations continuously to the AI agents they run and to the business decisions those agents influence.

This paper proposes the Artificial Intelligence Control Tower (AICT) as such a mechanism. Borrowing the control-tower concept from aviation, logistics, cybersecurity, and supply-chain operations, the AICT provides centralized visibility and policy control without requiring every analytical or operational task to be centralized. It registers AI agents, coordinates workflows, monitors behavior, validates high-impact recommendations, enforces approval thresholds, maintains decision lineage, and measures business value. Because it spans architecture, governance, assurance, observability, and executive oversight, the AICT is best understood as an enterprise operating model and not as a single dashboard, platform, or model.

The Saudi Arabian context increases the relevance of this inquiry. The Saudi Data and AI Authority identifies data and AI as direct or indirect contributors to 66 of the 96 objectives of Saudi Vision 2030 and aims to position the Kingdom as a leading data-driven economy (Saudi Data and AI Authority, n.d.-a, n.d.-b). The National Strategy for Data and AI emphasizes skills, policy, investment, research, innovation, and ecosystem development. Organizations operating in agriculture, food, manufacturing, logistics, energy, finance, and public services therefore need architectures that convert AI investment into measurable national and enterprise value while maintaining trust, data governance, and regulatory readiness.

The challenge is not merely technical. Agentic systems alter decision rights, control structures, and the allocation of organizational responsibility. A procurement agent authorized to create a purchase order, a workforce agent proposing staffing changes, or a safety agent initiating containment may shape outcomes that were previously reserved for managers. The central governance question therefore becomes how organizations can delegate action without delegating accountability. This question is intensified when several agents collaborate, because the final recommendation may be an emergent product of distributed reasoning rather than the output of a single identifiable model.

The AICT also addresses an organizational scaling problem. Enterprises commonly create isolated governance

reviews for each AI use case, which produces long approval cycles and inconsistent evidence. A control-tower model enables reusable agent-registration standards, semantic controls, policy services, assurance tests, and approval patterns. It creates a common management language across business, technology, risk, legal, cybersecurity, privacy, audit, and finance. The objective is not to centralize all AI development, but to centralize the minimum controls and portfolio information required for coherent enterprise accountability.

The primary research question is:

How can enterprises establish an AI Control Tower that governs, orchestrates, monitors, and assures autonomous AI decision-making while maintaining accountability, compliance, and business value?

Three supporting questions guide the study:

1. Which architectural capabilities are required to coordinate enterprise AI agents?
2. Which controls should determine the permissible autonomy of an AI-enabled decision?
3. Which measures can demonstrate operational, economic, and governance outcomes?

The AICT is informed by prior work in this research program but is evaluated here as a distinct artifact. Earlier studies addressed multi-agent workforce architecture, pre-execution decision assurance, governance dashboards, cybersecurity observability, and smart-city governance separately (Mazhar, 2026a, 2026b, 2026c, 2026d, 2026e). The present manuscript does not claim novelty merely by combining those publications. Its independent contribution is the decision-centered coupling of agent identity, policy-aware orchestration, graded admissibility, human-owned execution authority, portfolio value measurement, and gated organizational maturity within one testable operating model. Table 1 assesses this contribution against international standards and external organizational, multi-agent, and decision-intelligence literature, while the earlier artifacts are retained only to document conceptual lineage.

The study makes five contributions. First, it defines the AICT as a distinct enterprise artifact rather than a dashboard, platform, or committee. Second, it integrates agent orchestration, decision assurance, and management-system governance in one six-layer architecture. Third, it formalizes confidence, risk, graded policy admissibility, execution authority, expected value, governance coverage, and gated maturity so that the framework is measurable and falsifiable rather than purely descriptive. Fourth, it proposes a maturity model, balanced performance framework, and seven testable propositions. Fifth, it positions Saudi Vision 2030 as an application context and maps the model to relevant national requirements without claiming national validation. The remainder of the paper reviews the literature, presents the DSR method and artifact, evaluates it conceptually, discusses implications, and identifies limitations and future validation needs.

LITERATURE REVIEW

Enterprise AI and the shift to agentic systems

Enterprise AI adoption is best understood as organizational transformation rather than software deployment. AI changes how information is created, evaluated, and embedded in processes; it can also distribute expertise and authority. Research on AI in organizations emphasizes tensions between automation and augmentation, autonomy and control, and local experimentation and enterprise coordination (Benbya et al., 2021; Berente et al., 2021). Raisch and Krakowski (2021) show that automation and augmentation are not sequential stages but a recursive paradox that organizations must manage continuously, since delegating a task to a machine reshapes the human work that surrounds it. Firm-level evidence further indicates that AI value derives from a bundled capability, in which tangible, human, and intangible resources combine with governance and coordination mechanisms, and not from model quality alone (Mikalef & Gupta, 2021). Digital-transformation research similarly shows that technology value depends on changes in strategy, structure, capabilities, and business models (Bharadwaj et al., 2013; Vial, 2019). The AICT responds to these tensions by combining federated innovation with enterprise-level guardrails.

Enterprise AI historically emphasized prediction, optimization, and automation within predefined processes. Generative AI broadened access by enabling natural-language interaction with organizational knowledge and software, and it also introduced new requirements for evaluation, input–output control, and oversight (Bandi et al., 2025; Dwivedi et al., 2023). Agentic AI extends this trajectory by combining reasoning, planning, memory, tool use, environmental feedback, and action. In enterprise settings, the unit of analysis consequently shifts from an isolated model to a socio-technical system of agents, data, tools, policies, people, and business processes. Recent reviews describe multi-agent systems as an emerging architecture for decomposing complex tasks among specialized agents, while also noting unresolved challenges involving orchestration, operational reliability, evaluation, and governance (Joshi, 2025; Rah, 2026). Those unresolved challenges constitute the design problem addressed in this paper.

Agentic systems can offer economic advantages when specialization and coordination reduce cycle time and improve the quality of decisions. A monitoring agent may detect a deviation; a diagnostic agent may identify drivers; a forecasting agent may estimate future impact; a recommendation agent may rank interventions; and an execution agent may create a transaction. However, increasing autonomy also creates a delegation problem. Enterprises must determine which agent may act, on which data, with which tools, under which financial and operational limits, and with what evidence. This requires identity, authorization, traceability, and run-time policy enforcement in addition to conventional model management.

The transition also changes the scale and granularity of automation. Research on enterprise AI has long distinguished between automating tasks, generating insights, and engaging stakeholders (Davenport & Ronanki, 2018). Agentic AI combines these categories within a single workflow: an agent may retrieve knowledge, analyze data, negotiate with another agent, and initiate a transaction. Jarrahi (2018) argues that the durable configuration is symbiotic rather than substitutive, with machines handling analytical complexity and humans supplying intuition, contextual judgment, and responsibility for equivocal decisions. This convergence increases potential value but weakens traditional boundaries between analytical systems and operational controls. The AICT therefore makes delegated authority, not model capability, the basis for determining permissible action.

Multi-Agent Orchestration and Observability

Multi-agent systems distribute problem solving among interacting autonomous components. Their value depends on coordination protocols, task decomposition, shared context, conflict resolution, state management, and quality control (Dorri et al., 2018; Jennings, 2000). Orchestration can be centralized, hierarchical, distributed, or hybrid. Enterprise environments generally favor hybrid arrangements: specialized agents retain local capabilities, while an orchestration layer controls routing, task dependencies, business rules, escalation, and final action authority. Jennings (2000) cautions that agent-based engineering fails when interaction patterns are left implicit. That condition arises in most enterprises, because agents are typically introduced one use case at a time.

Observability is equally important. An enterprise must reconstruct which agents participated, which sources were consulted, which tools were used, how confidence changed, who approved the result, and what action occurred. Technical logs alone are insufficient because executives, auditors, risk owners, and process owners require understandable decision lineage. The AICT therefore treats observability as a business-control capability rather than a purely technical monitoring function.

Classical multi-agent research emphasizes autonomy, social ability, reactivity, and proactiveness (Wooldridge & Jennings, 1995), and surveys of the field show that coordination, trust, and security remain the dominant open problems as agent populations scale (Dorri et al., 2018). Contemporary enterprise systems add large language models, retrieval, software tools, and natural-language coordination. This improves flexibility but creates new failure modes: agents may share incorrect assumptions, amplify an early error, overuse tools, or converge on a plausible but unsupported plan. Organizational control therefore requires local evaluation of each agent as well as system-level evaluation of the workflow. The control tower supplies the shared event history, identity model, policy context, and escalation mechanism that keep distributed agents governable.

AI Governance, Risk, and Responsible AI

AI governance establishes policies, roles, controls, and accountability for the organizational use of AI. Mäntymäki et al. (2022) define it as a system of rules, practices, and processes that keeps an organization's AI activities aligned with its strategy, obligations, and values, positioned alongside corporate, IT, and data governance rather than separate from them. NIST's AI RMF is voluntary, rights-preserving, sector-neutral, and designed to operationalize trustworthy AI through Govern, Map, Measure, and Manage (Tabassi, 2023). Its trustworthiness characteristics include validity, reliability, safety, security, resilience, accountability, transparency, explainability, privacy, and fairness, and its generative AI profile adds risk categories that are highly relevant to tool-using agents, including confabulation, information integrity, information security, and value-chain dependencies (National Institute of Standards and Technology, 2024). ISO/IEC 23894 translates ISO 31000 risk logic into AI-specific identification, analysis, treatment, and monitoring activities, while ISO/IEC 42001 specifies a certifiable AI management system covering organizational context, leadership, planning, support, operations, performance evaluation, and continual improvement (International Organization for Standardization, 2023a, 2023b). Schuett (2024) complements these process standards by mapping AI risk ownership onto the three-lines model, which distinguishes operational risk owners, independent risk and compliance functions, and internal audit. The AICT adopts that allocation for its agent, assurance, and audit responsibilities.

The OECD AI Principles promote innovative and trustworthy AI that respects human rights and democratic values, with emphasis on inclusive growth, transparency, robustness, security, safety, and accountability (Organisation for Economic Co-operation and Development, 2024). The European Commission's High-Level Expert Group on Artificial Intelligence (2019) earlier specified seven requirements for trustworthy AI, among them human agency and oversight, technical robustness, and accountability. These requirements were subsequently hardened into the risk-tiered obligations of the European Union Artificial Intelligence Act, which requires deployers of high-risk systems to maintain human oversight, logging, and traceability throughout the lifecycle (European Parliament and Council of the European Union, 2024). Floridi and Cowls (2019) consolidate the proliferation of ethical declarations into five principles: beneficence, non-maleficence, autonomy, justice, and explicability. Leslie (2019) converts comparable commitments into project-level practice through stakeholder impact assessment and process-based governance. Microsoft's Responsible AI Standard translates similar principles into product-development requirements and impact assessments (Microsoft, 2022), and Gasser and Almeida (2017) argue that governance must be layered across technical, ethical, and social levels rather than concentrated at any single one. Collectively, these sources establish a strong normative foundation. The unresolved issue, confirmed by systematic review evidence, is the operationalization of principles into structural, relational, and procedural practices at run time (Papagiannidis et al., 2025). The AICT addresses this gap by embedding policy checks, risk classification, approval logic, evidence capture, and performance monitoring directly in the agentic workflow.

Explainability is central to assurance because an accountable decision owner must be able to understand the basis, limits, and expected consequences of a recommendation. Explainable-AI research distinguishes transparent models from post hoc explanations and warns that explanation quality must be evaluated for the intended audience and task (Arrieta et al., 2020). Within the AICT, an explanation takes the form of an evidence package that records source provenance, assumptions, alternatives considered, uncertainty, policy checks, and the relationship between the recommendation and enterprise objectives. This evidence package is the unit on which the assurance functions defined in Section 4.4 operate, which allows explainability to be audited as a measurable property of a decision.

Decision Intelligence and Human Oversight

Decision intelligence connects data, analytical models, business context, human judgment, and outcomes. It extends the trajectory that business intelligence and analytics research has followed from describing performance to shaping decisions and learning from results (Chen et al., 2012). Shrestha et al. (2019) show that organizations can structure AI-enabled decisions in several distinct ways, including full delegation to the algorithm, hybrid sequential structures, and aggregated human–AI decision-making. In their account, the

appropriate structure depends on the specificity of the decision space, the interpretability of the method, the size of the alternative set, and the speed of feedback. Wilson and Daugherty (2018) similarly report that the largest performance gains occur where humans and machines are deliberately designed to complement one another. Agentic AI accelerates this cycle, but autonomous action should remain proportional to impact and uncertainty. Human oversight must therefore be designed rather than assumed. A person who receives hundreds of low-quality alerts does not provide effective oversight, and an approval that lacks sources, alternatives, confidence, and expected impact is ceremonial.

The AICT uses risk-tiered oversight. Low-impact, reversible, policy-bounded actions may be automated. Medium-impact decisions may require post-action review or single approval. High-impact decisions involving material financial exposure, safety, employment, legal rights, sensitive data, or strategic commitments require prior human authorization and, where appropriate, dual control. This approach preserves meaningful human accountability while allowing automation where the risk is acceptable.

Business value and AI Economics

AI governance should not be treated only as a compliance cost. It can improve value realization by reducing duplicated models, shortening approval cycles, increasing reuse of governed data and agents, and enabling controlled scaling. The business case should combine productivity, decision velocity, revenue or margin impact, risk avoidance, asset utilization, and workforce effects. The economic potential identified by Chui et al. (2023) will not be realized through technology access alone, since organizations must redesign workflows and measure whether AI changes outcomes. Capability-based evidence supports this conclusion, showing that firm performance follows from how AI resources are orchestrated and not from adoption itself (Mikalef & Gupta, 2021). Multidisciplinary assessments of generative AI reach a similar view, noting that opportunity and exposure rise together and that benefit claims require evidence (Dwivedi et al., 2023).

AICT creates an enterprise-level portfolio view. Executives can compare AI initiatives by value, risk, adoption, operational stability, and strategic alignment. Process owners can identify where human effort is reduced or redirected. Risk functions can observe incidents and authority violations. Technology leaders can monitor model and agent performance. This balanced perspective reduces the risk that organizations optimize a narrow technical metric while overlooking financial, operational, or human consequences.

Enterprise value also depends on data governance. Analytics and AI systems inherit the definitions, quality, lineage, and access decisions embedded in their data environments. Studies of data governance emphasize decision rights, accountabilities, stewardship, and coordination across organizational domains (Abraham et al., 2019; Khatri & Brown, 2010; Weber et al., 2009). The AICT extends these principles from data assets to AI agents and decisions: every agent requires ownership, every high-impact action requires authority, and every result requires traceable evidence.

Saudi Vision 2030 and National Data-and-AI Alignment

Saudi Vision 2030 and the National Strategy for Data and AI provide the application context for this study rather than an empirically validated national implementation environment. For an organization operating in the Kingdom, AICT controls should be configured against applicable requirements, including the Personal Data Protection Law and its implementing regulations, National Data Management Office policies and standards, SDAIA AI Ethics Principles, National Cybersecurity Authority controls, and sector-specific obligations (National Cybersecurity Authority, 2025; Saudi Data and AI Authority, 2025, n.d.-c). These requirements may include lawful and purpose-limited processing, data minimization, retention and destruction, data-subject rights, transfer and disclosure controls, classification, access management, cybersecurity, human accountability, fairness, transparency, robustness, and incident handling. Applicability and legal interpretation remain the responsibility of the competent organizational and regulatory authorities.

The AICT is relevant because it can translate national principles into organizational controls. Policies can encode national data-classification and privacy requirements; the agent registry can capture data residency and

vendor dependencies; assurance workflows can enforce human approval; and the executive layer can measure local capability development and value contribution. The framework is not jurisdiction-specific, but the Saudi context provides a strategic environment in which trustworthy scaling is especially consequential.

National alignment should include capability development as well as compliance. A mature control tower creates roles for AI product ownership, model and agent validation, semantic governance, policy engineering, assurance, incident response, and value management. It can also support local-language testing, documentation of model and data residency, and transparent reporting of economic contribution. These capabilities connect national AI ambition to repeatable enterprise practice and reduce dependence on informal experimentation.

Comparative synthesis of existing frameworks

The reviewed sources address different parts of the problem, and none was designed for a population of interacting enterprise agents that can act. Risk and management-system standards specify what must be governed, but they do not describe how tasks are routed between agents at run time. Multi-agent research specifies coordination protocols while assuming that questions of authority have already been settled elsewhere. Decision-intelligence and business-value research explains which outcomes matter without identifying the control points that make those outcomes attributable. Table 1 compares the principal reference frameworks against the capabilities that autonomous enterprise decision-making requires and shows the residual gap that the AICT is designed to close.

Table 1. Comparative coverage of reference frameworks against the requirements of autonomous enterprise decision-making

Framework or research stream	Primary focus	Agent identity and delegated authority	Run-time orchestration and pre-execution assurance	Enterprise value measurement
NIST AI RMF 1.0 and Generative AI Profile (National Institute of Standards and Technology, 2024; Tabassi, 2023)	Trustworthiness characteristics and risk functions	Partial: actor roles defined, agent-level authority not specified	Partial: measurement and management functions, no workflow control point	Not addressed
ISO/IEC 42001 and ISO/IEC 23894 (International Organization for Standardization, 2023a, 2023b)	Management system and lifecycle risk process	Partial: system-level ownership and documentation	Not addressed: periodic rather than transactional controls	Partial: performance evaluation clause only
OECD Principles, EU HLEG requirements, and EU AI Act (European Commission High-Level Expert Group on Artificial Intelligence, 2019; European Parliament and Council of the	Normative and legal obligations, risk tiers, human oversight	Partial: provider and deployer duties, not intra-organizational agent authority	Partial: oversight, logging, and traceability duties without architectural mechanism	Not addressed

European Union, 2024; Organisation for Economic Co-operation and Development, 2024)				
Organizational AI governance research (Mäntymäki et al., 2022; Papagiannidis et al., 2025; Schuett, 2024)	Governance definition, practices, and risk ownership	Addressed at organizational level, not at agent level	Not addressed	Partial: effects theorized, rarely instrumented
Multi-agent systems research (Dorri et al., 2018; Jennings, 2000; Joshi, 2025; Rah, 2026)	Coordination, task decomposition, tool use, benchmarks	Partial: agent roles defined technically, business authority assumed	Addressed for orchestration, weak on pre-execution business assurance	Not addressed
Decision intelligence and AI business value (Chen et al., 2012; Mikalef & Gupta, 2021; Shrestha et al., 2019)	Decision structures, capability, and firm performance	Not addressed	Partial: delegation structures described conceptually	Addressed
Prior artifacts of the present research programme (Mazhar, 2026a, 2026b, 2026c, 2026d, 2026e)	Multi-agent operating model, decision assurance, governance dashboards	Addressed for single artifacts, not consolidated	Addressed separately for orchestration and for assurance	Partial: executive reporting without formal value model
Proposed AI Control Tower	Integrated enterprise operating model for governed autonomy	Addressed through the agent registry and delegated authority limits	Addressed through policy-aware routing and the assurance gate in Eq. (5)–(6)	Addressed through the balanced performance framework and Eq. (7)

Conceptual lineage from the Authors' Existing Work

The earlier artifacts supply partial design ingredients rather than evidence for the present model's validity. SEMAI informs agent specialization and orchestration; the Autonomous Decision Assurance Layer informs pre-execution controls; and the governance-dashboard studies inform lineage, observability, and executive evidence (Mazhar, 2026a–e). AICT advances beyond those components by making the governed decision—not an individual model, dashboard, or agent—the unit that binds authority, admissibility, execution, economic value, and maturity. Its novelty is therefore judged against the external gaps summarized in Table 1 and remains subject to empirical validation.

Research Gap and Propositions

Table 1 isolates the gap addressed by this study. No current framework specifies agent-level accountability, run-time orchestration with pre-execution assurance, and enterprise value measurement together for populations of interacting agents. The propositions below translate that gap into testable relationships, and Table 8 links each proposition to indicative measures.

- P1: A governed enterprise data and knowledge foundation positively influences the reliability and explainability of agentic decisions.
- P2: An agent registry with explicit ownership, purpose, tools, and authority limits reduces accountability gaps and authority drift.
- P3: Orchestration with policy-aware task routing improves decision consistency and cycle time compared with uncoordinated agent deployments.
- P4: Risk-tiered decision assurance increases trust and control without eliminating the productivity benefits of bounded autonomy.
- P5: Integrated observability and decision lineage improve auditability, incident response, and organizational learning.
- P6: Executive portfolio oversight strengthens alignment between AI investment, enterprise strategy, and measurable economic value.
- P7: AICT maturity mediates the relationship between enterprise AI capability and sustainable operational performance.

METHODOLOGY

Research Design

This study uses Design Science Research (DSR) because its objective is to create and evaluate an artifact that addresses a relevant organizational problem. DSR extends organizational capability through purposeful artifacts and evaluates them for utility, quality, and efficacy (Hevner et al., 2004), building on the distinction between build and evaluate activities and between constructs, models, methods, and instantiations (March & Smith, 1995). The AICT is accordingly positioned as a model-and-method artifact: it specifies constructs (layers, agent roles, risk tiers, decision lineage), a model of their relationships, and a method for applying them to enterprise decisions. The process follows the six activities proposed by Peffers et al. (2007): problem identification and motivation, definition of solution objectives, design and development, demonstration, evaluation, and communication. Evaluation is ex ante, artificial, and formative in the terms of the FEDS framework, which is the appropriate strategy when the cost of premature naturalistic evaluation is high and the artifact is still being specified (Venable et al., 2016). Table 2 maps each activity to the corresponding section of this paper and to its output.

DSR activity (Peffers et al., 2007)	Method applied in this study	Output	Reported in
Problem identification and motivation	Analysis of the governance gap between agent deployment and enterprise oversight	Six control needs and one primary research question	Introduction and Section 3.2
Definition of solution objectives	Purposive synthesis of literature, standards, and national strategy sources	Comparative coverage analysis and artifact objectives	Sections 2.7 and 3.3
Design and development	Mapping of control needs to architectural layers and formalization of control logic	Six-layer architecture, decision lifecycle, Eq. (1)–(9), maturity model	Sections 4.1–4.4 and 4.9
Demonstration	Application to five representative enterprise decision scenarios	Scenario walk-throughs with agents, risks, and controls	Sections 3.5 and 4.10
Evaluation	Ex ante artificial formative	Criterion-by-criterion	Sections 3.6 and

	evaluation against eight design criteria (Venable et al., 2016)	assessment and identified residual risks	5.7
Communication	Specification of reusable components and an empirical validation protocol	Implementation roadmap, constructs, and research agenda	Sections 3.7, 4.11, and 5.8

Problem Identification and Objectives

The focal problem is the absence of an integrated enterprise mechanism for governing multiple AI agents that influence business decisions. The problem was decomposed into six control needs: trusted context, agent identity and accountability, workflow coordination, decision assurance, policy enforcement, and business-value measurement. The artifact objectives were therefore to be platform-neutral, risk-based, auditable, scalable across functions, compatible with existing governance systems, and understandable to both executives and technical teams.

Knowledge-Base Synthesis

The framework was informed by purposive synthesis of four evidence categories. The first comprised peer-reviewed design-science and multi-agent literature. The second comprised international governance sources, including NIST AI RMF, OECD AI Principles, ISO/IEC 42001, and the Microsoft Responsible AI Standard. The third comprised enterprise and economic research on generative and agentic AI. The fourth comprised Saudi national strategy sources describing Vision 2030 and the National Strategy for Data and AI. Sources were selected for relevance to architecture, governance, assurance, economic value, or national alignment. As a conceptual DSR study, the method does not claim systematic review or statistical meta-analysis.

Artifact Design

The design process mapped each identified control need to an architectural layer and management capability. Trusted context became the enterprise data and knowledge foundation. Accountability became the AI workforce and agent registry. Coordination became the orchestration and workflow-control layer. Safe autonomy became the decision-assurance layer. Regulatory and ethical requirements became the governance and compliance layer. Strategic alignment and value measurement became the executive control-tower layer. Five cross-cutting requirements were then applied across all layers: security, privacy, observability, change management, and human accountability.

Conceptual Demonstration Scenarios

The artifact was demonstrated through five representative enterprise scenarios. Procurement optimization tests spend monitoring, supplier analysis, recommendation, financial thresholds, and approval. Inventory forecasting tests time-series data, uncertainty, service-level trade-offs, and operational execution. Workforce planning tests sensitive data, fairness, privacy, and senior approval. Operational risk monitoring tests real-time detection, safety escalation, and incident response. Executive KPI management tests semantic consistency, narrative generation, strategic prioritization, and decision lineage.

Evaluation Criteria

The study uses an ex ante, artificial, formative evaluation against eight criteria derived from the research problem and from external governance expectations: completeness, traceability, accountability, controllability, interoperability, scalability, economic measurability, and responsible-AI alignment. To reduce circularity, the evaluation distinguishes designed mechanisms from independent acceptance evidence. Architectural presence alone is treated as necessary but insufficient. A criterion is considered conceptually supported only when the scenario identifies a responsible owner, specifies the evidence that an independent reviewer would inspect, defines a failure or escalation condition, and identifies a measurable outcome or residual risk. The assessment therefore evaluates internal coherence and evaluability rather than observed effectiveness. No criterion score is interpreted as proof of operational performance, and no scenario outcome is treated as measured evidence.

Field implementation, external stakeholder review, quantitative comparison with a baseline, and longitudinal confirmation remain necessary for summative validation.

Table 3. Ex ante evaluation criteria and accepted evidence

Criterion	Evaluation question	Evidence accepted in this study
Completeness	Does the artifact cover every identified control need?	One-to-one mapping of the six control needs to architectural layers
Traceability	Can a completed decision be reconstructed end to end?	Decision lineage record and the coverage measure in Eq. (8)
Accountability	Is an identifiable human owner responsible for each action?	Registry ownership fields, the accountable-owner delegation condition in Eq. (5), and the execution-authority rule in Eq. (6).
Controllability	Can autonomy be constrained, paused, or reversed?	Execution rule in Eq. (6), tiered approval, and rollback controls
Interoperability	Can the artifact coexist with existing governance systems?	Explicit alignment with NIST AI RMF, ISO/IEC 42001 and 23894, and the EU AI Act
Scalability	Does control effort grow sub-linearly with the number of agents?	Reusable policy, registry, and assurance services shared across use cases
Economic measurability	Can benefits and costs be attributed to governed decisions?	Balanced performance framework and the net value expression in Eq. (7)
Alignment with responsible AI	Are fairness, privacy, safety, and oversight enforced before action?	Policy admissibility test and tier 4 prohibition on autonomous execution

Communication and Reproducibility

To support reuse and independent replication, the artifact is communicated through layer definitions, agent roles, risk tiers, the formal control expressions in Section 4.4, performance metrics, propositions, maturity model, and an implementation roadmap. These components allow researchers and practitioners to instantiate the framework on different platforms and in different sectors while preserving its control logic. A future empirical protocol should document, at minimum: baseline process performance for the target decision; the agent configuration and version history; the policy set and approval thresholds in force; the weighting vectors used in Eq. (3) and Eq. (9); every escalation, override, and incident; and post-deployment outcomes over a defined measurement window. Reporting these parameters allows results to be compared across organizations and allows differences in performance to be attributed to configuration rather than to unobserved context.

Ethics and Data Statement

No human participants, personal data, confidential organizational data, animals, or experimental interventions were used. The paper develops and evaluates a conceptual framework using public literature and illustrative business scenarios. Formal ethical approval was therefore not required. Future empirical research involving employees, operational systems, or organizational records should obtain appropriate ethical and institutional permissions.

Results: The Ai Control Tower Framework

Definition and design principles

The AI Control Tower is defined as a federated enterprise capability that registers, orchestrates, observes, assures, and governs AI agents and AI-enabled decisions across business functions. It combines technical services, management processes, decision rights, and executive oversight. This definition is narrower than AI governance and broader than model monitoring: the object of control is the decision rather than the model, and

the scope covers the portfolio of agents that influence that decision. Five principles guide the design: bounded autonomy, evidence before action, ownership by design, proportional control, and continuous learning.

Bounded autonomy requires every agent to operate within explicit data, tool, financial, temporal, and process limits. **Evidence before action** requires material recommendations to include sources, assumptions, alternatives, confidence, expected impact, and relevant policy checks. **Ownership by design** assigns a business owner, technical owner, risk owner, and escalation path. **Proportional control** aligns oversight with impact, reversibility, uncertainty, and regulatory sensitivity. **Continuous learning** links outcomes and incidents back to agents, policies, models, and processes. Figure 1 summarizes the six-layer architecture and the cross-cutting controls that connect the technical and managerial components of the AICT.

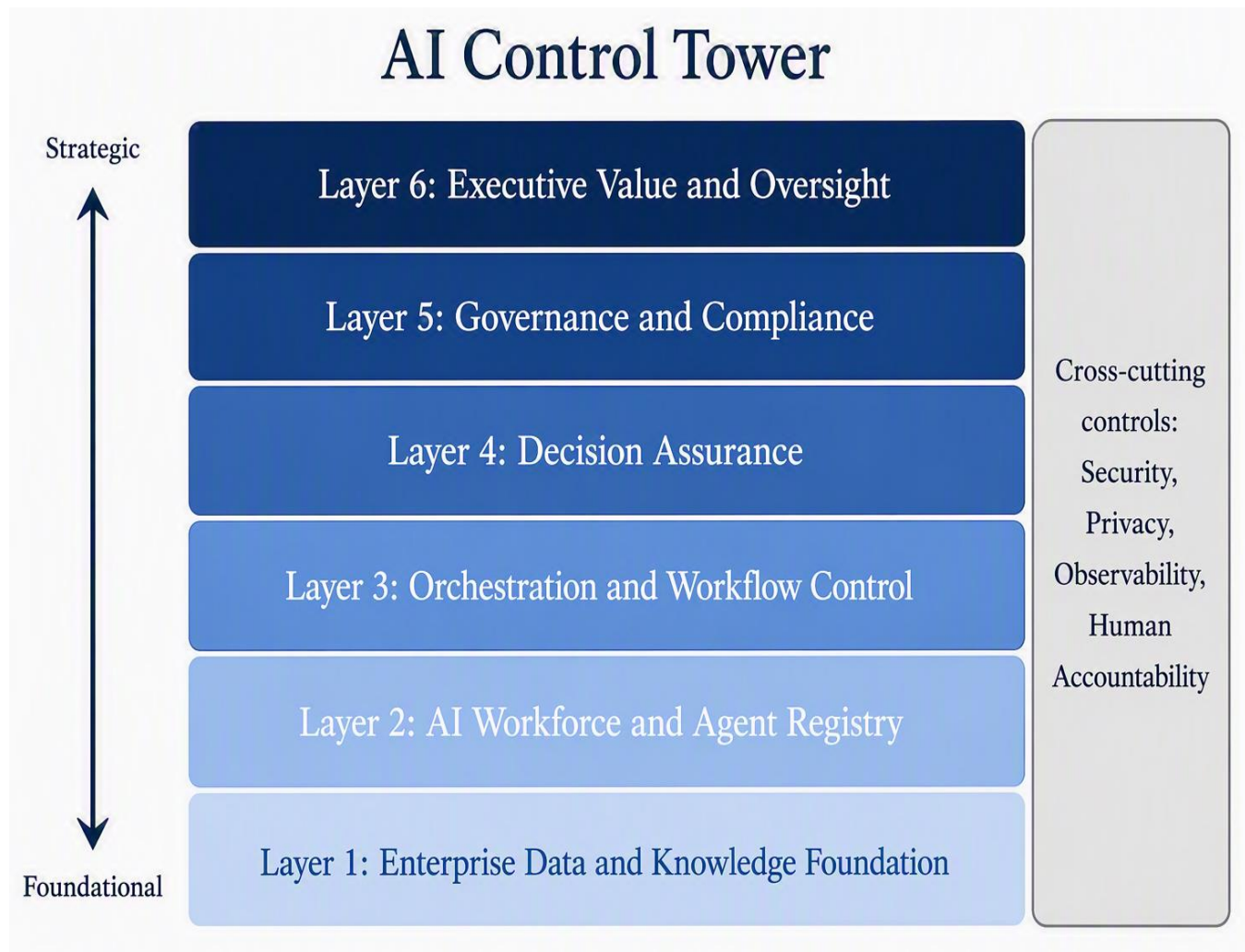


Figure 1. Six-layer Artificial Intelligence Control Tower architecture.

Six-Layer Architecture

Layer 1: Enterprise Data and Knowledge Foundation

The foundation integrates structured, semi-structured, and unstructured sources through governed data products, semantic models, metadata, lineage, master data, and retrieval services. It supplies agents with authorized, current, and contextually meaningful information. Core controls include classification, access management, data quality, retention, residency, provenance, and semantic ownership. The layer may be implemented through a lakehouse, warehouse, data fabric, knowledge graph, vector index, or hybrid architecture.

Layer 2: AI Workforce and Agent Registry

The registry is the system of records for every production agent. It records purpose, owner, model, version, prompts or policies, tools, data domains, risk tier, permitted actions, business process, dependencies, evaluation results, and retirement status. Typical roles include monitoring, diagnostic, forecasting, recommendation, compliance, risk, financial-impact, execution, and executive-narrative agents. Registration prevents unowned or undocumented agents from entering production and enables portfolio control.

Layer 3: Orchestration and Workflow Control

This layer decomposes goals, routes tasks, manages shared state, coordinates agent communication, enforces sequencing, resolves conflicts, manages retries, and applies time and cost limits. It separates business-process logic from model reasoning. Policy-aware routing ensures that sensitive data or high-impact tools are available only to authorized agents. The orchestrator also supports graceful degradation: when an agent fails or confidence falls below a threshold, the workflow pauses, selects an alternative, or escalates to a human.

Layer 4: Decision Assurance

The assurance layer evaluates whether an AI-generated recommendation or action is sufficiently trustworthy for its intended use. It assesses source grounding, confidence, uncertainty, consistency, policy compliance, financial exposure, safety impact, fairness, privacy, reversibility, and required approval. It may compare outputs from multiple agents, run deterministic business rules, perform simulation, or require supporting documents. Decisions receive one of four dispositions: approve, approve with conditions, escalate, or reject.

Layer 5: Governance and Compliance

This layer connects enterprise policies, NIST AI RMF processes, an ISO/IEC 42001-aligned AI management system, data governance, cybersecurity, privacy, internal audit, legal requirements, and responsible AI. It maintains policies as executable controls where feasible and as documented procedures where judgment is required. It also manages incidents, exceptions, change approvals, third-party AI risk, periodic reviews, and evidence for assurance or certification.

Layer 6: Executive Value and Oversight

The executive layer presents an enterprise portfolio view of AI adoption, performance, risk, compliance, and value. It shows which agents are active, which decisions are automated, where human approvals are delayed, which incidents recur, and which use cases produce measurable benefits. The control tower supports prioritization and investment governance rather than replacing operational dashboards. Its principal outputs are decisions about scaling, remediation, funding, ownership, and risk acceptance.

AI Decision Lifecycle

The AICT governs an eight-stage lifecycle: **detect** a business event; **diagnose** likely causes; **predict** future impact; **generate** alternatives; **assure** evidence, risk, and compliance; **approve** according to decision rights; **execute** the authorized action; and **learn** from outcomes. The lifecycle creates a control point before execution and feedback loops after it. For low-risk actions, assurance and approval may be automated; for higher-risk actions, evidence is presented to accountable humans. Figure 2 represents this lifecycle as a governance loop. The control tower sits at the center because evidence, risk, policy, and accountability apply at every stage, not only immediately before execution.

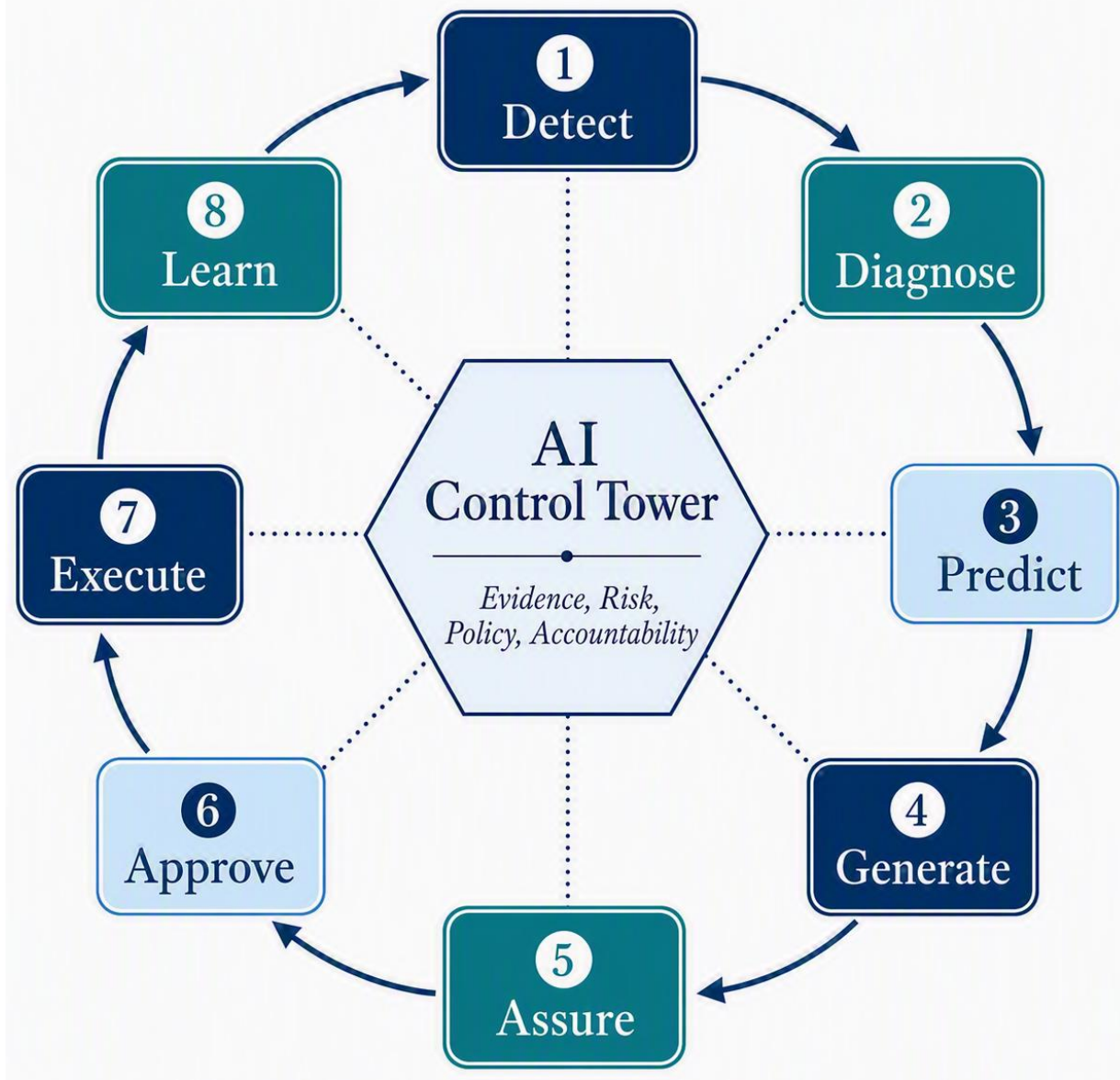


Figure 2. Governed AI decision lifecycle.

Formal Model of Governed Autonomous Decision-Making

Sections 4.1–4.3 describe the architecture narratively. This section formalizes the control logic so that autonomy becomes a set of explicit, reviewable parameters rather than a hidden system property. The formulation is intentionally interpretable, but its simplifying assumptions are stated and sensitivity testing is required. Table 4 defines the notation. Equations (1)–(9) define the governed decision, assurance confidence, decision risk, autonomy tier, graded policy admissibility, execution authority, expected net value, traceability, and non-compensatory maturity.

Table 4. Notation and governance assumptions

Symbol	Definition	Domain and governance source
d, D	Governed decision and decision population	Decision inventory
$A(d), E(d)$	Participating agents and evidence package	Registry and orchestration log
$c(a), q(a), w(a)$	Calibrated confidence, groundedness, and contribution weight	[0,1]; comparable validation window and approved calibration
$C(d)$	Aggregated assurance confidence	[0,1]; Equation 2
$I(d), V(d), S(d)$	Normalized impact, irreversibility, and regulatory or ethical sensitivity	[0,1]; anchored risk rubric

$k_1 \dots k_4$	Risk weights expressing risk appetite	Non-negative and summing to one; sensitivity-tested
$R(d), \tau(d), \theta(t)$	Risk index, assigned tier, and tier ceiling	[0,1]; Table 5
$\Pi(d)$	Aggregate policy admissibility after resolving satisfied, conditional, not-applicable, unresolved, conflicted, and failed rules	{0,1}; versioned policy service plus authorized human interpretation
$x(d), x^*(\tau, o), c^*(\tau)$	Exposure, accountable-owner limit, and minimum confidence	Delegation matrix and assurance policy
$G(d), \tau^*_{\text{exec}}(a_e)$	Assurance gate and designated execution agent's maximum authorized tier	{0,1} and {1...3}; registry and least-privilege authorization
$p(d), B(d), L(d), K(d)$	Correctness probability, benefit, loss, and governance/operating cost	Outcome tracking and benefit case
$\lambda(d), \Gamma$	Lineage completeness and portfolio coverage	{0,1} and [0,1]; observability record
$m(j), \beta(j), M, g_{j,r}$	Layer score, weight, composite maturity, and mandatory floor for layer j at level r	[1,5]; maturity assessment and progression policy

The Governed Decision Object

The controlled object of the AICT is the decision, not the model. A governed decision is therefore defined as an ordered tuple that binds a business goal to the agents that acted on it, the evidence they produced, the action taken, the human owner accountable for it, and the time of execution:

$$d = \langle g, A(d), E(d), a(d), o(d), t(d) \rangle \quad (1)$$

Here g is the business goal that triggered the workflow. Because every element of the tuple is persisted, the record supports both regulatory traceability obligations (European Parliament and Council of the European Union, 2024) and organizational learning: an outcome observed months later can still be attributed to the agents, evidence, and owner that produced it.

Assurance Confidence and Decision Risk

In a multi-agent workflow, confidence scores are comparable only after calibration against task-specific holdout outcomes. For each registered agent, $c(a)$ is therefore interpreted as an empirically calibrated probability or reliability score for the defined task and review window, using an approved method such as reliability curves, Brier scores, isotonic regression, or Platt scaling. The groundedness term $q(a)$ is assessed separately against authorized-source coverage and citation support. Scores from different tasks or materially different validation windows must not be aggregated until they have been normalized to the same [0, 1] interpretation and accepted by the assurance owner. The orchestration weight $w(a)$ reflects the agent's documented contribution to the final recommendation and is not a self-reported model confidence. Where calibration evidence is absent, stale, or non-comparable, the workflow cannot claim a valid aggregate confidence and must be escalated. Subject to these conditions, the assurance layer computes:

$$C(d) = \sum a \in A(d) w(a) \cdot c(a) \cdot q(a), \text{ with } \sum w(a) = 1 \quad (2)$$

Confidence alone is an insufficient basis for autonomy, because a highly confident recommendation may still be consequential, irreversible, or legally sensitive. The assurance layer therefore computes a decision risk index that combines business impact, residual uncertainty, irreversibility, and regulatory or ethical sensitivity. Residual uncertainty is taken as the complement of aggregated confidence, so that weak evidence increases measured risk rather than being absorbed silently:

$$R(d) = k_1 I(d) + k_2 [1 - C(d)] + k_3 V(d) + k_4 S(d) \quad (3)$$

Equation (3) is a transparent screening model, not a claim that the four dimensions are inherently linear or independent. Impact, uncertainty, irreversibility, and sensitivity must first be normalized through organization-

approved rubrics with documented anchors. The weights express risk appetite and require periodic governance approval. Before production use, the organization should test alternative weight sets and interaction terms, especially where high impact and irreversibility jointly create disproportionate loss. A conservative override applies when any dimension reaches a prohibited threshold, regardless of the weighted total. This prevents a low score in one dimension from masking an unacceptable value in another.

Autonomy Assignment, Admissibility, and Execution

Autonomy is then assigned rather than assumed. The decision is placed in the lowest tier whose risk ceiling it satisfies, which makes the risk appetite of the organization an explicit and reviewable set of thresholds:

$$\tau(d) = \min \{ t \in \{1, 2, 3, 4\} : R(d) \leq \theta(t) \}, \quad \theta(1) < \theta(2) < \theta(3) < \theta(4) = 1 \quad (4)$$

Tier assignment determines oversight but does not authorize execution. Policy evaluation is represented as a set of rule outcomes rather than a single binary judgment. Each applicable rule is classified as satisfied, conditionally satisfied, not applicable, unresolved, conflicted, or failed. Not-applicable decisions require a recorded rationale; conditional results require specified compensating controls; conflicting or unresolved rules require interpretation by the designated policy owner; and any failed mandatory rule blocks execution. The aggregate indicator $\Pi(d)$ equals 1 only when all mandatory rules are satisfied or conditionally satisfied with approved controls and no conflict remains. The assurance gate additionally requires calibrated confidence and financial exposure within the accountable owner's delegated authority:

$$G(d) = 1 \Leftrightarrow [\Pi(d) = 1] \wedge [C(d) \geq c^*(\tau(d))] \wedge [x(d) \leq x^*(\tau(d), o(d))] \quad (5)$$

Execution authority is determined by the specific execution agent and the accountable human owner, not by the most privileged agent participating elsewhere in the workflow. Let $\tau^*_{\text{exec}}(a_e)$ denote the highest tier for which the designated execution agent a_e is registered, and let $x^*(\tau, o)$ denote the owner's delegated financial or operational limit. Advisory agents may contribute evidence at higher capability levels but cannot expand execution authority. If several execution agents act, the workflow adopts the most restrictive applicable authorization and tool scope. Any substitution of an execution agent, tool, owner, or workflow purpose triggers re-evaluation of the gate. The disposition is therefore:

$$\text{Act}(d) = \text{execute if } G(d) = 1 \wedge \tau(d) \leq \tau^*_{\text{exec}}(a_e); \text{ refer for approval if } G(d) = 1 \wedge \tau(d) > \tau^*_{\text{exec}}(a_e); \text{ escalate or reject if } G(d) = 0 \quad (6)$$

The rule preserves two governance constraints. Tier 4 decisions remain ineligible for autonomous execution. In addition, authority cannot be inherited from a monitoring, diagnostic, or recommendation agent with broader registration; only the designated execution agent may act, within the accountable owner's approved limits and the workflow's least-privilege scope. This makes delegated authority explicit, revocable, and attributable.

Worked Example and Sensitivity Analysis

Consider a procurement recommendation produced by three agents with normalized contribution weights of 0.30, 0.40, and 0.30. Their calibrated confidence scores are 0.90, 0.80, and 0.70, and their groundedness scores are 0.95, 0.90, and 0.85. Equation (2) gives $C(d) = (0.30 \times 0.90 \times 0.95) + (0.40 \times 0.80 \times 0.90) + (0.30 \times 0.70 \times 0.85) = 0.723$. Assume normalized impact $I = 0.60$, irreversibility $V = 0.30$, sensitivity $S = 0.20$, and equal risk weights of 0.25. Equation (3) then gives $R(d) = 0.25(0.60) + 0.25(1 - 0.723) + 0.25(0.30) + 0.25(0.20) = 0.344$. With illustrative ceilings $\theta(1) = 0.20$, $\theta(2) = 0.40$, $\theta(3) = 0.70$, and $\theta(4) = 1.00$, the decision falls in Tier 2. It may execute only if all mandatory policies are admissible, the Tier 2 confidence minimum is no greater than 0.723, the proposed exposure is within the accountable owner's delegation, and the designated execution agent is registered for Tier 2.

A one-way illustrative sensitivity analysis shows that the result depends on governance choices. Increasing the uncertainty weight from 0.25 to 0.40 and reducing the remaining three weights proportionally to 0.20 each gives $R(d) = 0.20(0.60) + 0.40(1 - 0.723) + 0.20(0.30) + 0.20(0.20) = 0.331$. In this example, risk decreases

slightly because residual uncertainty (0.277) is lower than impact (0.60); the decision remains in Tier 2. If calibrated confidence falls to 0.60 under the same weights, risk becomes 0.38 and remains within Tier 2. Risk exceeds the Tier 2 ceiling of 0.40 only when calibrated confidence falls below 0.55; at exactly 0.55, risk equals 0.40 and remains within Tier 2 under Equation (4). A high impact or irreversibility score may also activate the conservative override even when the weighted total remains moderate. These values are illustrative and do not constitute empirical validation; production thresholds require back-testing, loss analysis, and approval by the relevant governance body.

Economic Justification, Traceability, and Maturity

Bounded autonomy must also be economically justified. The expected net value of executing a decision under a given oversight configuration is the probability-weighted benefit less the probability-weighted loss and the cost of governing, assuring, and operating the workflow:

$$E[V(d)] = p(d) B(d) - [1 - p(d)] L(d) - K(d) \quad (7)$$

Delegation is justified only where the expected net value under automation is at least equal to the expected net value under human execution for the same decision. Equation (7) makes three managerial insights explicit. First, increasing oversight raises $K(d)$ but may also raise $p(d)$, so governance can be evaluated as an investment rather than treated only as overhead. Second, decisions carrying large asymmetric losses cannot be justified by high average accuracy. Third, reusable control services lower $K(d)$ across the portfolio, which provides the economic rationale for a control tower rather than case-by-case review.

Governance coverage expresses the proportion of decisions in a period for which a complete lineage record exists, where completeness requires the agent set, the evidence package, the policy results, the approval, and the executed action all to be recoverable:

$$\Gamma = (1 / |D|) \cdot \sum_{d \in D} \lambda(d), \quad \lambda(d) \in \{0, 1\} \quad (8)$$

Enterprise capability is summarized as a weighted index across the six architectural layers, assessed on the five-level scale in Section 4.9. The index is descriptive and cannot by itself authorize progression. To prevent compensatory scoring, each maturity level r also has mandatory floor values $g_{j,r}$ for critical layers. An organization may report attainment of level r only when M meets the composite threshold and every mandatory layer satisfies $m(j) \geq g_{j,r}$. In particular, progression to AI Control Tower or Autonomous Enterprise maturity requires minimum levels for data governance, agent accountability, decision assurance, governance and compliance, and observability; high performance in orchestration or value management cannot compensate for a failure in those controls.

$$M = \sum_{j=1}^6 \beta(j) \cdot m(j), \quad \text{with } \sum \beta(j) = 1 \quad \text{and } m(j) \in [1, 5] \quad (9)$$

Equations (1)–(9) convert the propositions of Section 2.9 into observable quantities. P1 and P4 concern $C(d)$ and $R(d)$; P2 and P3 concern $\tau^*_{\text{exec}}(a_e)$ and the routing that produces $A(d)$; P5 concerns Γ ; P6 concerns the distribution of $E[V(d)]$ across the portfolio; and P7 concerns M together with the mandatory layer floors. The parameters are governance choices that must be documented, sensitivity-tested, approved, and revised with observed evidence.

Autonomy and Approval Model

Table 5 operationalizes the tier assignment produced by Eq. (3) and Eq. (4). The tiers are cumulative, so each level inherits the controls of the levels below it and adds a further constraint on who may act, on what evidence, and with what record. The thresholds $\theta(1)$ to $\theta(3)$ are governance parameters and not technical defaults. Calibrating them is the main instrument through which an organization expresses its risk appetite for autonomous action.

Table 5. Risk-tiered autonomy and approval model

Tier	Decision characteristics	Permitted authority	Required control
1	Low impact, reversible, policy-bounded	Automatic execution	Continuous logging and periodic review
2	Moderate impact with established procedure	Conditional automation	Threshold checks and post-action review
3	Material financial, operational, or customer impact	Recommendation only	Prior approval by accountable owner
4	Safety, employment, legal rights, strategic commitments, or sensitive data	No autonomous execution	Dual control, documented rationale, and formal audit trail

AI workforce roles

The AI workforce is specified as a set of registered roles and not as a collection of tools. Specialization serves a control purpose: narrow, well-evaluated agents are easier to validate, monitor, and retire than general-purpose agents with diffuse responsibilities, and separating roles allows the orchestration layer to withhold sensitive data and high-impact tools from agents that do not need them. Table 6 lists representative roles together with the control that makes each role governable.

Table 6. Representative AI workforce roles

Agent	Primary responsibility	Key control
Monitoring agent	Observes KPIs, events, and exceptions	Approved indicators and alert thresholds
Diagnostic agent	Identifies drivers and causal hypotheses	Source traceability and alternative explanations
Forecasting agent	Estimates future outcomes and uncertainty	Back-testing, drift monitoring, and confidence intervals
Recommendation agent	Ranks response alternatives	Policy, cost, risk, and constraint checks
Compliance agent	Evaluates rules, obligations, and evidence	Versioned policies and human legal review where required
Execution agent	Creates or completes authorized transactions	Least privilege, transaction limits, and rollback
Executive agent	Produces strategic narratives and escalation summaries	Grounded evidence, neutrality, and owner approval

Performance and value framework

Performance measurement in the AICT is balanced across several dimensions because relying on a single dimension can distort AI governance and investment decisions. An automation rate on its own rewards unchecked delegation, accuracy on its own ignores financial asymmetry, and compliance on its own ignores value. Table 7 therefore pairs six measurement dimensions with the management question each is intended to answer, and every dimension maps to a quantity defined in Section 4.4.

Table 7. AI Control Tower performance indicators

Dimension	Illustrative indicators	Management question
Operational	Decision-cycle time, resolution time, automation rate, service level	Is the process faster and more reliable?
Financial	Verified savings, avoided cost, revenue impact, working-capital impact, ROI	Is AI creating measurable economic value?
Quality	Accuracy, groundedness, forecast error, rework, recommendation acceptance	Are outputs fit for the intended decision?
Governance	Registered-agent coverage, approval compliance, lineage completeness, review closure	Is accountability operating as designed?

Risk	Incidents, authority violations, high-risk escalations, rollback success, residual risk	Is autonomy within risk appetite?
Workforce	Adoption, AI literacy, time redirected to higher-value work, user trust	Are people able and willing to use the system effectively?

Research constructs and proposition mapping

Table 8. Construct operationalization and proposition mapping

Construct	Type	Operational indicators	Data source and test	Proposition
Governed data foundation	Formative	Quality pass rate; semantic conformance; lineage coverage; authorized-access rate	Metadata and access logs; composite index; regression or PLS-SEM	P1
Agent accountability	Formative	Owner coverage; documented limits; current evaluation; lifecycle status	Agent registry; coverage index; case comparison	P2
Orchestration capability	Formative	Routing accuracy; conflict resolution; workflow completion; recovery time	Workflow logs; before-and-after or longitudinal analysis	P3
Decision assurance	Formative	Calibrated confidence; policy admissibility; approval compliance; reversible-action coverage	Assurance records; incident and approval data; multilevel analysis	P4
Observability	Formative	Lineage completeness; detection time; reconstruction time; audit-evidence retrieval	Telemetry and audit logs; event-study analysis	P5
Executive portfolio governance	Formative	Strategic alignment; verified benefits; risk-adjusted funding; remediation closure	Portfolio and finance records; case or panel analysis	P6
AICT maturity	Second-order formative	Six layer scores subject to mandatory governance and assurance floors	Assessment instrument; PLS-SEM or longitudinal maturity study	P7
Sustainable operational performance	Reflective outcome	Decision speed; reliability; risk-adjusted value; user trust; control stability	Process, financial, incident, and survey data; confirmatory factor analysis	P3–P7

Maturity model

Level 1 - **AI awareness**: isolated experiments, limited ownership, and no consolidated inventory.

Level 2 - **AI assistants**: sanctioned copilots and task-level automation with basic governance.

Level 3 - **AI workforce**: multiple specialized agents coordinated within selected processes.

Level 4 - **AI Control Tower**: centralized registration, assurance, observability, risk-tiered autonomy, and portfolio management.

Level 5 - **Autonomous enterprise**: cross-functional, self-optimizing processes operating within explicit strategy, policy, and risk boundaries. Progression requires stronger governance and organizational capability, not merely more capable models.

Figure 3 illustrates progression in technical capability, governance, integration, and verified value. The weighted index in Equation (9) supports comparison, but advancement is non-compensatory. Level 3 requires at least Level 2 in data governance, registry and accountability, decision assurance, and governance and compliance. Level 4 requires at least Level 3 in those layers plus observability. Level 5 requires at least Level 4 in all six layers, demonstrated control effectiveness, and sustained outcome evidence. A deficient mandatory layer caps the reported maturity level regardless of the composite score.

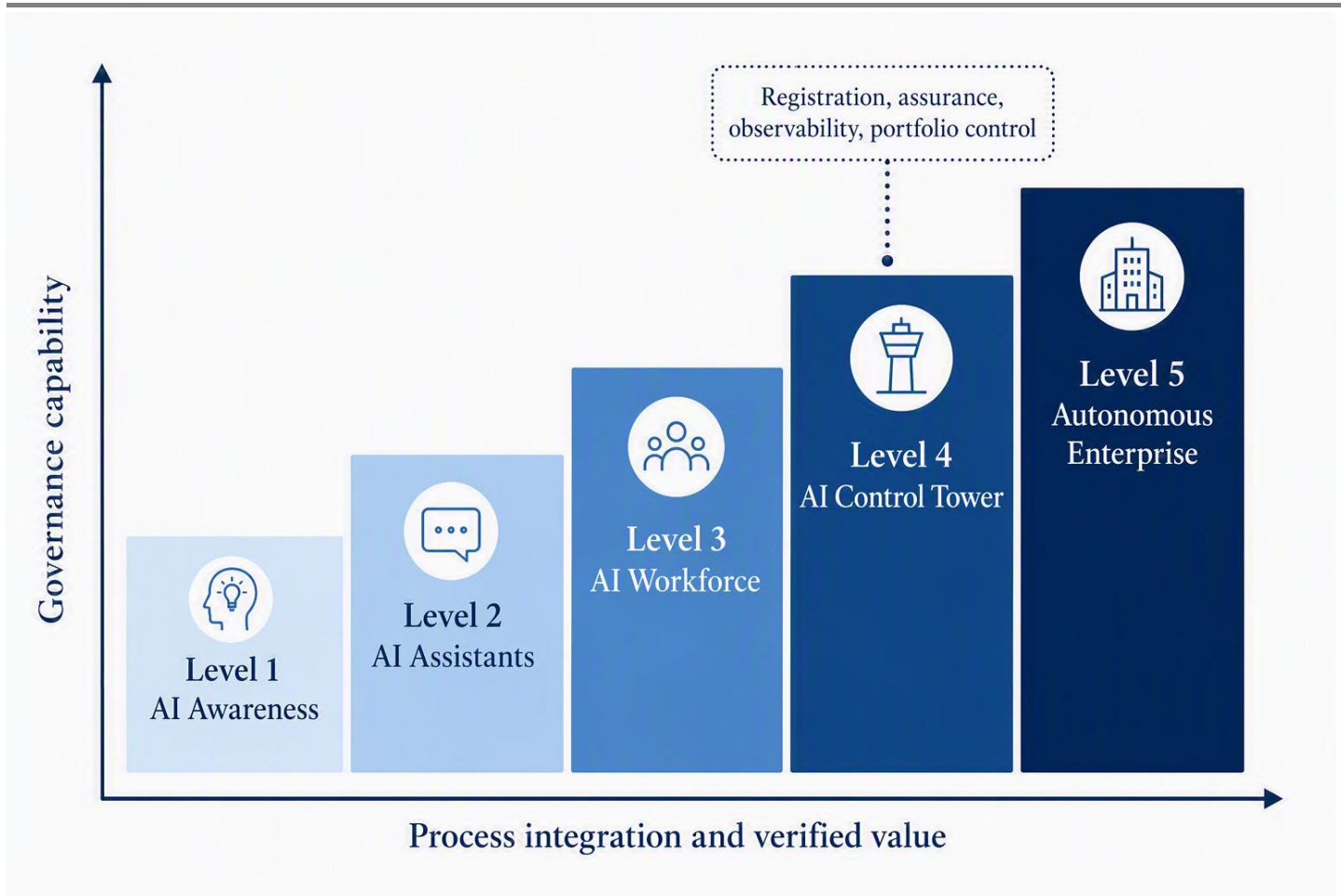


Figure 3. AI Control Tower maturity model.

Conceptual scenario evaluation

The five scenarios are illustrative walkthroughs used to test whether the artifact can represent the required actors, evidence, policy decisions, authority boundaries, and measurable outcomes. In procurement, the model specifies spend detection, supplier analysis, financial thresholds, and accountable approval. Inventory planning illustrates uncertainty, service-level trade-offs, and reversible parameter changes. Workforce planning illustrates restricted data, fairness and privacy review, and prohibition of autonomous employment decisions. Operational risk illustrates pre-approved containment under human command. Executive KPI management illustrates certified semantics and source-grounded narratives. These walkthroughs support internal coherence and identify how a future implementation could be evaluated; they do not show that spend leakage, inventory levels, response times, trust, or performance improved.

Table 9. Conceptual scenario evaluation and evidence required for validation

Scenario	Principal risk	AICT control represented	Potential outcome	Evidence required
Procurement optimization	Financial authority and supplier bias	Delegation limits, evidence checks, approval	Lower spend leakage and shorter sourcing cycle	Baseline comparison, approved savings, cycle-time and bias audit
Inventory planning	Forecast error and excess working capital	Calibration, simulation, reversible automation	Improved service and inventory efficiency	Forecast error, stockout, inventory, and working-capital measures
Workforce planning	Privacy, fairness, and employment	Restricted data, fairness	More transparent and accountable	PDPL review, fairness tests, approval records,

	impact	assessment, dual approval	recommendations	and stakeholder assessment
Operational risk	Safety and operational continuity	Pre-approved containment, logging, human command	Faster bounded response	Response-time, false-alarm, rollback, and incident-severity data
Executive KPI management	Inconsistent definitions and unsupported narratives	Certified measures, grounding, executive approval	More consistent decision intelligence	Semantic-conformance, citation-support, correction, and user-trust measures

Implementation roadmap

Table 10. AI Control Tower implementation roadmap

Phase	Focus	Key deliverables	Exit criteria
1. Discover	AI portfolio and process inventory	Agent register, decision inventory, risk taxonomy	Ownership and scope approved
2. Govern	Policies and decision rights	Autonomy tiers, approval matrix, control library	Governance board approval
3. Pilot	Controlled use cases	Orchestration workflow, assurance checks, logs	Target quality and risk thresholds met
4. Scale	Reusable enterprise services	Shared agents, semantic models, policy services	Multi-function adoption and stable operations
5. Optimize	Portfolio and economic management	ROI dashboard, incident learning, maturity assessment	Demonstrated value and continual improvement

DISCUSSION

Theoretical contribution

The AICT integrates three domains that are usually treated independently: agent architecture, AI governance, and business-value management. Multi-agent research explains how autonomous components collaborate, governance frameworks define trustworthy outcomes and management practices, and decision intelligence links evidence to action. The proposed artifact unifies these domains around the enterprise decision as the controlled object, which moves governance from periodic model review toward continuous system-level assurance. In doing so it answers the layered-governance argument of Gasser and Almeida (2017) by showing where the technical, ethical, and social layers meet in a single operational control point. It extends the organizational AI governance construct proposed by Mäntymäki et al. (2022) downward into run-time execution, and it responds to the operationalization gap identified in systematic review evidence (Papagiannidis et al., 2025) by specifying computable controls in place of principles.

The framework also clarifies that autonomy is not a binary property. As formalized in Eq. (4) to Eq. (6), enterprise autonomy is a delegated and revocable authority determined by impact, uncertainty, reversibility, and policy. A highly capable agent may receive little authority in a sensitive process, while a simpler rules-based agent may execute low-risk actions automatically. This distinction separates technical capability from organizational permission and provides a basis for empirical research on autonomy design. It also extends the decision-structure typology of Shrestha et al. (2019) from a choice made once at design time to a property computed for each decision. On the same logic, the automation–augmentation paradox (Raisch & Krakowski, 2021) becomes a governable parameter: organizations do not resolve it once, but settle it repeatedly through the threshold vector they publish.

Managerial Implications

Executives should treat AICT as an operating-model initiative rather than a dashboard project. Leadership must establish decision rights, risk appetite, funding, and accountability. Technology teams should implement

the registry, orchestration, observability, and assurance services. Business owners must define outcomes, rules, exceptions, and acceptance criteria. Risk, legal, privacy, cybersecurity, internal audit, and data governance functions should contribute to reusable controls instead of reviewing each use case independently.

A practical implementation sequence is recommended. First, establish an inventory of AI systems and agents. Second, classify use cases by impact and define decision authority. Third, standardize data, semantic models, logs, and evidence. Fourth, pilot a limited number of high-value, medium-risk processes. Fifth, measure operational and financial outcomes. Sixth, scale through reusable agent, policy, and assurance services. This sequence supports governance before broad autonomy and avoids building a centralized structure without practical use cases.

Business and Economic Implications

The AICT is designed to improve the economics of AI through four mechanisms: reuse of governed agents, tools, and data products; lower control friction through reusable policies; reduced incident and rework exposure through observability and rollback; and more comparable value-and-risk information for investment decisions. These mechanisms are propositions rather than measured findings in this study. Economic evaluation should therefore use verified baselines and distinguish gross benefits from implementation, infrastructure, change-management, assurance, and operating costs.

For industrial organizations, potential benefits include improved forecast accuracy, lower inventory, reduced procurement leakage, faster exception resolution, fewer compliance failures, and better utilization of assets and specialist time. These benefits should not be assumed. Each use case requires an accountable benefit owner, baseline, target, measurement period, and attribution method. The control tower should make unsuccessful use cases visible so resources can be redirected.

Workforce and Organizational Implications

Agentic AI changes roles rather than simply reducing tasks. Employees become supervisors, exception handlers, policy designers, domain evaluators, and decision owners, which is consistent with evidence that the strongest performance gains arise where human and machine responsibilities are complementary by design (Jarrahi, 2018; Wilson & Daugherty, 2018). Managers need literacy in uncertainty, automation risk, data quality, and human–AI collaboration. Organizations should redesign jobs, incentives, and training alongside technology. Trust will depend on whether users can understand outputs, challenge recommendations, correct errors, and see that escalation has consequences. Where oversight is nominal, the confidence measured in Eq. (2) becomes a ritual number instead of a control.

The framework also supports separation of duties. The team that develops an agent should not be solely responsible for approving its high-risk deployment. Following the three-lines logic applied to AI by Schuett (2024), business ownership forms the first line, independent assurance, risk, privacy, and security review the second, and internal audit the third, with the agent registry and decision lineage supplying the evidence each line requires. Smaller organizations may combine roles, but decision rights should remain explicit and documented.

Saudi Vision 2030 Implications

In the Saudi context, the AICT offers a candidate mechanism for translating national data-and-AI ambitions and applicable obligations into enterprise controls. Its registry can record processing purpose, data classification, residency, vendor dependency, and accountable ownership; orchestration can enforce least privilege and approved data routes; assurance can require privacy, fairness, safety, and human-review evidence; observability can retain decision lineage and incident records; and executive oversight can monitor local capability and verified value. These mappings are conceptual. The study includes no Saudi organizational deployment, regulator assessment, sector implementation, or stakeholder validation.

Saudi organizations should contextualize the model for Arabic-language performance, national data policies, sector requirements, local risk appetite, and organizational maturity. The framework is intended to complement, and not to replace, formal obligations and guidance issued by the relevant authorities. Cross-sector collaboration and anonymized use-case evidence could accelerate national learning while protecting confidential information.

Vision 2030 alignment matrix

Table 11. Saudi-context regulatory and strategic mapping for AICT implementation

Requirement family	AICT mechanism	Evidence retained	Status in this study
PDPL and implementing regulations	Purpose and legal-basis fields; minimization; access, transfer, retention, deletion, breach, and data-subject-request controls	Processing record, consent or legal basis, access log, transfer approval, retention action, incident record	Conceptual mapping; legal and sector validation required
NDMO data governance policies and standards	Classification, ownership, quality, metadata, lineage, sharing, and lifecycle controls	Data catalog, quality results, lineage graph, sharing approval, ownership record	Conceptual mapping; organization-specific control design required
SDAIA AI Ethics Principles	Human oversight, fairness, privacy, transparency, robustness, accountability, and risk assessment	Impact assessment, fairness test, explanation package, approval, monitoring outcome	Conceptual mapping; stakeholder assessment required
NCA cybersecurity controls	Identity, least privilege, secure tool use, logging, resilience, third-party risk, incident response	Access and configuration logs, security test, vendor assessment, incident and recovery evidence	Conceptual mapping; applicable NCA scope and sector profile must be confirmed
Vision 2030 and national strategy	Measured productivity, local capability, responsible innovation, and portfolio value oversight	Verified benefit case, adoption and skills measures, risk-adjusted investment decisions	Strategic context only; no national outcome attribution claimed

LIMITATIONS

This research is conceptual and relies on ex ante, artificial, formative evaluation. It provides no causal evidence that AICT improves speed, quality, financial performance, trust, or compliance. The scenarios are illustrative and the criteria test internal coherence and evaluability; despite the independent-evidence requirements added in Section 3.6, some self-reference remains because the artifact is judged partly against the objectives used to design it. The formal model also simplifies reality. Confidence aggregation assumes task-specific calibration and comparability, the risk index uses normalized additive dimensions unless an override applies, policy admissibility depends on correct rule interpretation, and the worked example is illustrative rather than validated. The maturity index requires proposed minimum floors, but those thresholds have not been empirically calibrated. Saudi Vision 2030 and the national regulatory landscape are application context only: no Saudi organization, regulator, sector implementation, or stakeholder sample was evaluated. The framework is platform-neutral and does not estimate organization-specific integration or operating cost. Finally, Rah (2026) is a preprint, and the status and bibliographic details of all recent 2025–2026 sources should be rechecked immediately before submission.

Future research agenda

Future research should proceed through four complementary streams. The first is **measurement and validation**. Case-based and longitudinal studies should compare decision-cycle time, errors, incidents, trust, adoption, and risk-adjusted return before and after implementation. The first seven constructs in Table 8 are modeled formatively because their indicators jointly create the capability and need not covary; sustainable

operational performance is modeled reflectively because its indicators are expected manifestations of the outcome. Researchers should validate content through expert review, test collinearity and indicator relevance for formative blocks, assess reliability and discriminant validity for reflective measures, and use partial least squares structural-equation modeling, multilevel analysis, interrupted time series, or matched-case comparisons as sample size and context permit. Equations (2), (3), (7), and (8) provide operational measures, while Equation (9) should be tested both as a weighted index and as a gated maturity classification.

The second stream is **human–AI governance**. Experiments should compare approval designs, explanations, confidence displays, escalation thresholds, and workload to identify when human oversight is substantive rather than ceremonial. Researchers should examine automation bias, accountability perceptions, professional identity, and the distribution of decision authority across managers, experts, and agents.

The third stream is **technical resilience**. Studies should evaluate multi-agent failure propagation, secure tool use, prompt and policy injection, identity and authorization, memory contamination, rollback, and system-level observability. Benchmarking should include not only task success but also evidence quality, policy compliance, reversibility, and graceful degradation.

The fourth stream is **institutional and national context**. Cross-sector and cross-country research could examine how regulation, language, data sovereignty, organizational culture, and risk appetite affect AICT design. Saudi studies could compare maturity and value across agriculture, food, logistics, energy, finance, healthcare, public administration, and smart-city environments. A validated maturity instrument, an anonymized repository of controls and incidents, and sector-specific reference architectures would be valuable outcomes.

Figure 4 consolidates these streams into a testable research model in which the six architectural capabilities act as antecedents of AICT maturity, and maturity in turn mediates the relationship between enterprise AI capability and sustainable operational and economic performance.

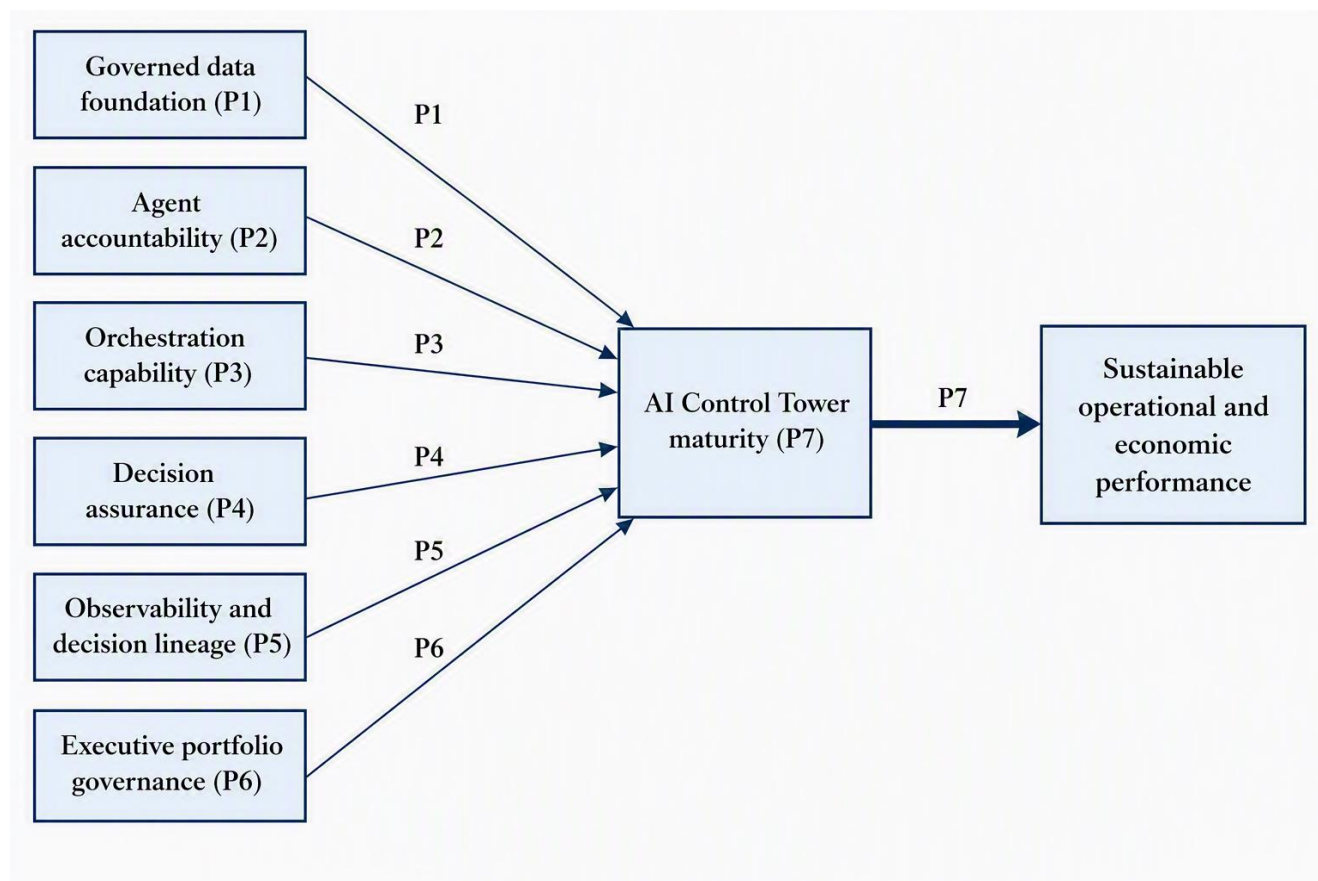


Figure 4. Proposed nomological network for empirical validation of the AI Control Tower.

CONCLUSION

This study developed the Artificial Intelligence Control Tower as an enterprise operating model for governing, orchestrating, assuring, and measuring autonomous AI-enabled decisions. The artifact addresses a growing gap between the pace of agent deployment and the maturity of organizational oversight. Its six layers integrate trusted data and knowledge, an accountable AI workforce, orchestration, decision assurance, governance and compliance, and executive value management.

The framework contributes a risk-tiered autonomy model, an eight-stage decision lifecycle, a balanced performance system, and a five-level maturity path. The conceptual walkthroughs illustrate how the artifact represents ownership, evidence, policy control, escalation, and outcome measurement across five scenarios. They do not demonstrate performance improvement. The AICT is therefore proposed as a testable model for bounded autonomy proportional to risk and subject to meaningful human accountability.

For organizations operating in Saudi Arabia, the framework offers a structured application context for aligning agentic AI with Vision 2030 and relevant data, privacy, cybersecurity, and AI-governance expectations. Its practical value remains contingent on sector-specific legal interpretation, field implementation, calibrated thresholds, trusted data, organizational capability, and disciplined measurement. Future studies should test the propositions, compare outcomes with baselines, validate maturity gates, and refine the controls with Saudi stakeholders and competent authorities. AICT is therefore presented as a foundation for empirical testing and responsible implementation, not as a validated national operating model.

ACKNOWLEDGEMENTS

The authors acknowledge the contributions of researchers, standards bodies, and public institutions advancing enterprise AI, responsible AI, decision intelligence, and data governance. No external grant funding was received for this conceptual study.

Conflict of Interest

The authors declare no conflict of interest.

Ethical Approval

Not applicable. This conceptual study did not involve human participants, animals, personal data, or confidential organizational data.

Funding

This research received no external funding.

Data Availability

No primary dataset was generated or analyzed. The study is based on public literature and conceptual enterprise scenarios.

Author Contributions

- **Choudhry Bilal Mazhar:** conceptualization, methodology, framework and formal model design, formal analysis, investigation, visualization, writing the original draft, and project administration.
- **Tahani Muneer Algethami:** literature curation, validation of the evaluation criteria and enterprise scenarios, data and governance context analysis, writing (review and editing), and verification of references and citations.

Both authors read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Generative Ai Disclosure

Generative AI tools were used for language refinement, structural drafting support, and the creation of conceptual figures. The authors independently reviewed and revised all generated material, verified the manuscript's arguments and citations against the cited sources, and take full responsibility for its content, interpretations, and conclusions.

REFERENCES

1. Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49, 424–438. <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
2. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bannetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
3. Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2025). The power of generative AI: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet*, 17(1), 23. <https://doi.org/10.3390/fi17010023>
4. Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. *Journal of the Association for Information Systems*, 22(2), 281–303. <https://doi.org/10.17705/1jais.00662>
5. Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3), 1433–1450. <https://doi.org/10.25300/MISQ/2021/16274>
6. Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. *MIS Quarterly*, 37(2), 471–482. <https://doi.org/10.25300/MISQ/2013/37:2.3>
7. Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188. <https://doi.org/10.2307/41703503>
8. Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zimmel, R. (2023). The economic potential of generative AI: The next productivity frontier. McKinsey & Company. <https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>
9. Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116.
10. Deloitte. (2026). The state of AI in the enterprise: The untapped edge. <https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html>
11. Dorri, A., Kanhere, S. S., & Jurdak, R. (2018). Multi-agent systems: A survey. *IEEE Access*, 6, 28573–28593. <https://doi.org/10.1109/ACCESS.2018.2831228>
12. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., et al. (2023). “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
13. European Commission High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. European Commission.
14. European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L 2024/1689. <http://data.europa.eu/eli/reg/2024/1689/oj>

15. Floridi, L., & Cows, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
16. Gasser, U., & Almeida, V. A. F. (2017). A layered model for AI governance. *IEEE Internet Computing*, 21(6), 58–62. <https://doi.org/10.1109/MIC.2017.4180835>
17. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
18. International Organization for Standardization. (2023a). ISO/IEC 23894:2023 Information technology—Artificial intelligence—Guidance on risk management. <https://www.iso.org/standard/77304.html>
19. International Organization for Standardization. (2023b). ISO/IEC 42001:2023 Information technology—Artificial intelligence—Management system. <https://www.iso.org/standard/42001>
20. Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586. <https://doi.org/10.1016/j.bushor.2018.03.007>
21. Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence*, 117(2), 277–296. [https://doi.org/10.1016/S0004-3702\(99\)00107-1](https://doi.org/10.1016/S0004-3702(99)00107-1)
- Joshi, S. (2025). Review of autonomous and collaborative agentic AI and multi-agent systems for enterprise applications. *International Journal of Innovative Research in Engineering and Management*, 12(3), 65–76. <https://doi.org/10.55524/ijirem.2025.12.3.9>
22. Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148–152. <https://doi.org/10.1145/1629175.1629210>
23. Leslie, D. (2019). Understanding artificial intelligence ethics and safety. The Alan Turing Institute. <https://doi.org/10.5281/zenodo.3240529>
24. Mäntymäki, M., Minkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*, 2(4), 603–609. <https://doi.org/10.1007/s43681-022-00143-x>
25. March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15(4), 251–266. [https://doi.org/10.1016/0167-9236\(94\)00041-2](https://doi.org/10.1016/0167-9236(94)00041-2)
26. Mazhar, C. B. (2026a). From AI assistants to AI workforces: A multi-agent enterprise operating model for Saudi Vision 2030 organizations. *Nexus Science Review*, 1(2). <https://doi.org/10.66104/9zjxnd21>
27. Mazhar, C. B. (2026b). Autonomous Decision Assurance Layer for AI-driven enterprise analytics: Bridging governance, multi-agent AI, and executive decision intelligence in Saudi Vision 2030 organizations. *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2026.84573>
28. Mazhar, C. B. (2026c). From smart services to autonomous governance: An agentic AI framework for Saudi smart cities. *European Journal of Computer Science and Information Technology*. <https://doi.org/10.37745/ejcsit.2013/vol13n121202>
29. Mazhar, C. B. (2026d). National-scale data governance dashboards: Real-time quality, lineage, access, and compliance for Saudi Arabia (SDAIA-aligned). *Veredas do Direito*, 23(5), e235431. <https://doi.org/10.18623/rvd.v23.5431>
- Mazhar, C. B. (2026e). Real-time AI-driven cybersecurity analytics dashboards for critical-infrastructure protection under Saudi Vision 2030: An NCA- and SDAIA-aligned governance-to-execution framework. *Nexus Science Review*, 1(2). <https://doi.org/10.66104/berfcb18>
30. Microsoft. (2022). Microsoft Responsible AI Standard, version 2: General requirements. <https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft-Responsible-AI-Standard-General-Requirements.pdf>
31. Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3), 103434. <https://doi.org/10.1016/j.im.2021.103434>
- National Cybersecurity Authority. (2025). Essential Cybersecurity Controls (ECC 2-2024). <https://nca.gov.sa/en/regulatory-documents/controls-list/ecc/>
32. National Institute of Standards and Technology. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.600-1>

33. Organisation for Economic Co-operation and Development. (2024). OECD AI Principles. <https://oecd.ai/en/ai-principles>
34. Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, 34(2), 101885. <https://doi.org/10.1016/j.jsis.2024.101885>
35. Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Rah, A. (2026). Agentic and multi-agent systems: A systematic review of tool use, benchmarks, and governance [Preprint]. Research Square. <https://doi.org/10.21203/rs.3.rs-9839978/v1>
36. Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1), 192–210. <https://doi.org/10.5465/amr.2018.0072>
37. Saudi Data and AI Authority. (n.d.-a). National Strategy for Data and AI. <https://sdaia.gov.sa/en/SDAIA/SdaiaStrategies/Pages/NationalStrategyForDataAndAI.aspx>
38. Saudi Data and AI Authority. (n.d.-b). SDAIA and Vision 2030. <https://sdaia.gov.sa/en/SDAIA/SdaiaStrategies/Pages/sdaiaAnd2030Vision.aspx>
- Saudi Data and AI Authority. (2025). AI Ethics Principles. <https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf>
- Saudi Data and AI Authority. (n.d.-c). Personal Data Protection Law. <https://sdaia.gov.sa/en/SDAIA/about/Documents/PersonalDataProtectionLaw.pdf>
39. Schuett, J. (2024). Three lines of defense against risks from AI. *AI & Society*. <https://doi.org/10.1007/s00146-023-01811-0>
40. Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66–83. <https://doi.org/10.1177/0008125619862257>
41. Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
42. Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, 25(1), 77–89. <https://doi.org/10.1057/ejis.2014.36>
43. Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), 118–144. <https://doi.org/10.1016/j.jsis.2019.01.003>
44. Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all—A contingency approach to data governance. *Journal of Data and Information Quality*, 1(1), Article 4. <https://doi.org/10.1145/1515693.1515696>
45. Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, 96(4), 114–123.
46. Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2), 115–152. <https://doi.org/10.1017/S0269888900008122>