RAIS Journal for Social Sciences | Vol. 10, No. 2, 2026
ISSN 2574-0245 (Print) | ISSN 2574-1179 (Online) | DOI:10.5281/zenodo.24453941

BAR-Governed Human-Directed Agentic Enterprise Architecture: Formal Model, Toolkit, and Computational Evaluation

Bhagyeshkumar Chokhawala[1], Ritaben Tongadiya[2]

[1]Comcast, Philadelphia, Pennsylvania, USA
[2]Independent Researcher, West Chester, Pennsylvania, USA
Corresponding Author: bhagyeshkumar_chokhawala@comcast.com

Abstract: Artificial intelligence (AI) agents that plan, invoke tools, and change enterprise state pose linked problems of organizational adoption and runtime authorization. This study integrates the Business Value and Purpose, Architecture Readiness and Integration, and Responsible Adoption and Realization (BAR) framework with Human-Directed Agentic Enterprise Architecture (HDAE). Using conceptual theory building, design-science artifact specification, and a reproducible stratified-random conformance experiment, the study formalizes two interacting governance levels: BAR evaluates enterprise capability legitimacy, readiness, and realization; HDAE mediates individual agent actions through human intent, bounded delegation, policies, outcome contracts, domain integrity, evidence, and recovery. Three formal expressions specify action admissibility, lifecycle gating, and minimum-constraint execution quality. Eight testable theoretical propositions (P1-P8) and a seven-phase implementation method connect the model to reusable architecture and governance artifacts. In sixteen stratified-randomly sampled experimental scenarios, an API-connected baseline conformed to two of thirteen runtime cases and returned inappropriate EXECUTE/RETRY labels in eleven of eleven risk-bearing cases. HDAE conformed to all thirteen runtime cases without those inappropriate decisions; integrated BAR-HDAE also handled all three lifecycle-gate cases. These specification-derived findings establish internal implementation consistency, not causal effectiveness or field superiority. The contribution is a theory-guided architecture, executable conformance specification, and reproducible evaluation basis for subsequent empirical research.

Keywords: Agentic AI, Artificial Intelligence Governance, Computational Conformance, Design Science, Enterprise Architecture, Formal Methods, Human-Centered AI, Responsible Adoption

Introduction

Recent evidence distinguishes AI adoption from deployment of operational agents. In McKinsey's May–June 2026 survey (n=1,719), 44% reported enterprise AI scaling; among respondents from organizations with at least $1 billion in annual revenue (n=595), 40% reported scaling agents in at least one function (Tinkoff et al., 2026). Capgemini's 2025 executive survey (n=1,522) found 23% running agent pilots and 14% at partial/full implementation (Capgemini Research Institute, 2025). Definitions and samples differ; neither survey reports production agent-failure incidence. Agents that plan, select tools, and alter authoritative records differ from advisory AI. API reachability alone does not confer business permission, proportional autonomy, or recovery authority. Architecture must connect accountable human purpose with independent checks on consequential actions (ISO/IEC/IEEE, 2022; The Open Group, 2022).

Technical risk has a separate denominator: Zou et al. (2025) documented more than 60,000 policy violations among approximately 1.8 million adversarial attempts against agents

supported by 22 models; AgentDojo comprises 97 tasks and 629 security cases (Debenedetti et al., 2024). These are experimental attack results, not production incident rates.

A reported 2025 Replit agent deletion of a live database illustrates possible severity without establishing prevalence (Sullivan, 2025). For firms, state-changing failures threaten continuity and financial integrity; for employees, decision rights and intervention burdens; for clients, privacy and service reliability; and for regulators, attribution and auditability. These consequences were not measured here.

Problem Statement

The affected population is organizations adopting agents that may modify payments, access rights, entitlements, or systems of record. Capgemini Research Institute (2025) reported 23% pilots, 12% partial-scale, and 2% full-scale implementation among 1,522 executives. These are deployment categories, not incident statistics. Scale-up evidence remains distinct: a 2026 HFS Research/Netomi survey of 100 senior leaders reported that an average of 26% of their agent initiatives reached production and 43% of associated spending concerned stalled initiatives (HFS Research, 2026). Deloitte found 15% reporting scaled, cross-functional multi-agent adoption among 501 U.S. decision-makers (Widener et al., 2026). Neither study establishes that insufficient governance caused the shortfall.

The unresolved issue is the lack of an independently field-validated bridge from capability-level readiness, business authorization, and accountable human intent to action-level policy, delegated scope, evidence, recovery, and subsequent adoption decisions. BAR addresses capability legitimacy and realization; HDAE specifies governed outcome episodes. This study offers an internally testable model and 16 experimental conformance cases, not evidence of superior production security, financial returns, or employee acceptance.

Purpose of the Study

This conceptual and design-science study integrates BAR capability-lifecycle governance with HDAE outcome-episode execution. It articulates theoretical mechanisms, specifies technology-neutral artifacts and typed action/lifecycle predicates, and evaluates deterministic controller conformance on seeded, stratified-random experimental cases. It includes no real-world state mutations, organizational outcomes, or causal comparison.

The study yields a dual-loop architecture, propositions P1–P8, a seven-phase method, twelve toolkit artifacts, and experimental conformance results. The propositions remain research predictions rather than relationships verified with field data.

Research Questions and Evidence Sources

RQ1. How can strategic AI adoption governance and operational agent governance be integrated without collapsing their distinct responsibilities? Evidence source: the BAR–HDAE dual-loop architecture and responsibility/evidence map (Figure 1; Table 1).

RQ2. Which constructs connect value, readiness, bounded action, and realization? Evidence: theoretical constructs, P1–P8, and formal predicates (Table 2; Equations 1–3); these are not empirically validated organizational effects.

RQ3. What architecture method and implementation toolkit can instantiate the integrated framework in a technology-neutral manner? Evidence source: the seven-phase evidence-gated operationalization method, the 12 reusable toolkit artifacts, and the runtime mediation pattern (Tables 4–5).

RQ4. Does the reference controller conform to its prescribed runtime and BAR rules? Evidence: a disclosed 192-candidate frame, seeded selection of 16 cases, specification-derived oracle, and case-level counts (Table 3; Supplement S1). The negative-control API connector is not a representative governed alternative.

The contributions are (i) a capability/episode responsibility model, (ii) eight falsifiable propositions, (iii) a seven-phase BAR-gated method, (iv) reusable intent, authority, policy,

transaction, evidence, and adoption artifacts, and (v) transparent computational conformance and sensitivity checks.

Related-work disclosure: BAR's business-value, architecture-readiness, responsible-adoption dimensions and evidence gates originated in Chokhawala (2025). The present work adds HDAE outcome episodes, action-admissibility semantics, evidence linkage, and a runnable conformance reference; it does not claim to reinvent BAR.

Literature Review and Theoretical Framework

Comparative Scope of Architecture and AI Governance Frameworks

TOGAF structures the development of business, data, application, and technology architecture; ISO/IEC/IEEE 42010 organizes descriptions by stakeholder concerns and viewpoints (ISO/IEC/IEEE, 2022; The Open Group, 2022). NIST AI RMF frames Govern/Map/Measure/Manage, while ISO/IEC 42001 defines an AI management system (Tabassi, 2023; International Organization for Standardization, 2023). None is itself an action-authority protocol. BAR addresses enterprise adoption; HDAE specifies delegation and enforced tool-action decisions, with evidence returned to BAR.

AgentDojo and public red-teaming work identify prompt-injection and tool-misuse hazards (Debenedetti et al., 2024; Zou et al., 2025). Internal auditing and responsible-AI scholarship link assigned roles, records, and remediation (Jobin et al., 2019; Raji et al., 2020; Rakova et al., 2021). These motivate the proposed evidence bridge but do not validate its field effects.

Enterprise Architecture, Capabilities, and Responsibility Allocation

EA models enterprise structure and transformation (Zachman, 1987), while capability models preserve business meaning independently of changing platforms. State-changing agents add a responsibility question: which principal authorizes an outcome, which actor plans actions, who mediates controls, and which domain system preserves authoritative invariants?

Human-Centered Automation and Socio-Technical AI

Human control need not decrease as automation increases: Parasuraman et al. (2000) distinguish automation functions, while Shneiderman (2020) emphasizes reliable human authority. Delegation should therefore vary with reversibility, impact, and decision rights.

Hybrid-intelligence and organizational studies emphasize complementary roles, socialization, and work redesign (Berente et al., 2021; Chowdhury et al., 2022; Dellermann et al., 2019). Interviews on organizational AI readiness identify the importance of prerequisites (Jöhnk et al., 2021), and an AI-assistance field study demonstrates measurable worker outcomes in a different, non-agentic setting (Brynjolfsson et al., 2025). Neither supplies causal evidence for HDAE delegation. Workforce trust and uptake remain proposed BAR-R outcomes, not outputs of the experimental tests.

AI Engineering, Agentic Systems, and Operational Risk

AI engineering involves dependencies, feedback, monitoring, and technical debt (Amershi et al., 2019; Sculley et al., 2015). Planning agents add multi-step tool selection and uncertain environmental effects (Huang et al., 2024; Liu et al., 2024; Shinn et al., 2023; Yao et al., 2023). Agent architecture must therefore enforce authenticated identity, scoped authorization, independent policy, durable approvals, idempotency, transaction recovery, and provenance outside model prompts; OWASP GenAI Security Project (2025) identifies related agentic attack surfaces.

AI Governance and Responsible Adoption

NIST's AI RMF and generative-AI profile, ISO/IEC 42001, and the revised OECD AI principles emphasize risk ownership, monitoring, robustness, transparency, and human oversight (Autio et al., 2024; International Organization for Standardization, 2023; Organisation

for Economic Co-operation and Development, 2024; Tabassi, 2023). BAR frames these concerns at capability approval and realization; HDAE proposes their translation into action-level gates.

Four Complementary Perspectives on Organizational Adoption and Responsible Realization

Technology–Organization–Environment (TOE) distinguishes compatibility/complexity, leadership/resources, and regulatory/partner context (Tornatzky et al., 1990). BAR-A concerns technical foundations; BAR-B, legitimate investment; BAR-R, workforce and external readiness. Wang and Zhang's (2025) study of 261 SMEs concerns digital transformation, not agentic execution. A reachable payment API may coexist with untrained approvers or regulatory restrictions: TOE therefore informs adoption context, not Boolean admissibility.

Stakeholder theory identifies affected interests and influence (Freeman, 1984). BAR-R should record consultation by employees, sponsors, architects, risk teams, customers, and affected individuals; dissent; appeal routes; and changes made in response. Evidence from fairness co-design and sectoral risk studies supports caution, not a proven BAR–HDAE effect (Haley & Burrell, 2025; Madaio et al., 2020).

Resistance may signal uncertainty, threatened expertise, or legitimate design defects (Bareil, 2013; Oreg, 2003). Participatory change research and role-identity findings motivate joint work redesign, training, and human override (Fosfuri & Rønde, 2009; Schweiger et al., 2018; Strich et al., 2021). The effect on uptake remains a future empirical question.

Ethical deployment requires proportionality, fairness, privacy, contestability, redress, and independent review. BAR-B screens affected rights; HDAE maps material risks to accountable policy and intervention; BAR-R reviews harms and remedies. Law-enforcement and recruitment examples are context-specific rather than cross-sector proof (Burrell & Haley, 2026; Haley & Burrell, 2025).

Feedback should be accessible, triaged by a named owner, and linked to policy, authority, or training changes; Burrell (2026) motivates this in healthcare AI, without establishing a universal effect.

The BAR Framework

BAR, introduced by Chokhawala (2025), comprises Business Value and Purpose (B), Architecture Readiness and Integration (A), and Responsible Adoption and Realization (R). It demands accountable purpose, data/API/identity readiness, and monitored stakeholder and benefit outcomes, with experiment/pilot/scale/hold/remediate/stop decisions.

The HDAE Framework

HDAE's unit is the bounded outcome episode: a versioned human Intent Envelope (why, what, who, when, where, which) delegates the operational how to an agent. Technology mediation validates identity, policy, outcome contracts, approvals, evidence, and recovery; domain APIs retain invariant enforcement. Episodes terminate in verified completion, denial, escalation, or reconciliation.

Antecedent Contributions and HDAE Novelty

BAR's three adoption dimensions, reference architecture, and evidence-gated scale decisions are inherited from Chokhawala (2025); they are not claimed as new.

HDAE newly specifies responsibility allocation, the Intent Envelope, Authority Profile, action-admissibility predicate, outcome-contract mediation, idempotency/reconciliation semantics, and execution provenance fed back to BAR. P1–P8 theorize relationships not tested on employees or enterprises.

Theoretical Constructs, Design Requirements, and Implementation Practices

Theoretical constructs are observable-in-principle variables (alignment, boundedness, execution quality, realization); normative design requirements include independent policy, provenance, and recoverability; implementation practices such as API gateways, OPA,

idempotency keys, and telemetry are replaceable. Building an artifact tests implementability, not the organizational propositions (Gregor, 2006; Hevner et al., 2004).

Research Gap and Integrated Theoretical Position

The narrow gap is an operationally specified connection between BAR's adoption authorization and HDAE's per-action authority and evidence feedback. The present study tests only the specification, while the sociotechnical effectiveness of the connection remains unverified.

The resulting integration is proposed as a design-theory contribution rather than a field-tested intervention. BAR's evaluation unit is the capability and its adoption decisions; HDAE's unit is the bounded, authorized action episode. A BAR approval is neither a universal permission token nor a substitute for independent downstream checks; conversely, execution that respects every technical rule may lack social legitimacy or sufficient realized benefit. This distinction creates an empirically falsifiable research program: assess how capability readiness affects episode-level mediation, how episode evidence changes later BAR decisions, and when stakeholder and organizational conditions moderate sustainable adoption. The present model makes these connections explicit but cannot yet establish their causal direction or magnitude.

Methodology

Research Design

The study uses targeted narrative synthesis, theory building, design-science artifact specification, formal operational semantics, and stratified-random selection of experimental conformance inputs (Bacharach, 1989; Gregor, 2006; Hevner et al., 2004; Whetten, 1989). Randomness affects case instantiations, not enterprise sampling, assignment to treatment, or runtime actions.

The evaluator returns symbolic decisions, not real payments, incidents, user behavior, or business value. P1–P8 are independently testable but are not measured as causal or organizational effects here.

Source Base and Synthesis Procedure

The targeted narrative corpus spans EA, human-centered automation, risk and agent security, design science, and organizational TOE/stakeholder/change/ethics research. Sources were selected for conceptual relevance to the capability lifecycle and outcome episode, technical control design, or bounded empirical context. BAR (Chokhawala, 2025) is a conceptual antecedent; standards are not empirical observations.

The research record lacks retrievable database logs, exact search strings, run dates, date filters, exclusion totals, and quality-screening sheets. Terms such as 'agentic AI', 'delegated authority', 'TOE', and 'responsible adoption' describe covered topics, not confirmed historical search queries. No systematic-review or PRISMA claim is made.

Construct mapping proceeded from source to level and artifact: architecture/capabilities informed BAR-A and outcome contracts; risk standards informed BAR-R and policy/evidence gates; human-centered automation informed Intent Envelopes and bounded delegation; agent-security research informed fail-closed tool mediation and reconciliation; TOE/stakeholder/change research informed BAR-R contextual moderators and proposed field indicators. Tables 1–2 define constructs and Tables 4–5 operational artifacts; Equations 1–2 specify only inspectable technical conditions.

The synthesis was deliberately interpretive rather than a database-exhaustive literature audit. For each retained concept, the authors assigned a unit of analysis, the normative decision it could inform, a candidate observability indicator, and an artifact that could instantiate it. The separation prevents a common category error: an ISO management requirement is not a Boolean variable tested in an agent runtime; a policy-engine response is not a validated measure of workforce trust; and an attack benchmark does not measure production incident prevalence. The mapping can be audited through the propositions, formal definitions, responsibility map, and toolkit, although a future systematic review should test whether omitted literature changes the proposed theoretical boundaries.

Units of Analysis and Boundary Conditions

The two nested units are the capability lifecycle (BAR: experiment, pilot, scale, remediate, retire) and the outcome episode (HDAE: intent to final or reconciled state). Episode observations could inform later capability decisions; this study does not observe such organizational feedback.

Scope is agents performing consequential enterprise workflows, not prediction-only systems. The design assumes identifiable capability owners, systems of record, identity/policy infrastructure, and auditable state; absent foundations are readiness deficiencies, not agent problems automatically solved by HDAE.

Formal Model and Operational Semantics

Let c be an AI-enabled capability, I a versioned human Intent Envelope, D a delegated Authority Profile, a a proposed action, and s an authoritative state (not simulated). V(I), V(D), Perm(a), P(a,s,I,D), E(a), S(a), R(a), and H(a) respectively denote valid intent, valid delegation, permission and nonprohibition, affirmative independent policy, available evidence, state mutation, available recovery, and satisfied approval requirements.

$$\mathrm{Adm}(a,s) = V(I) \wedge V(D) \wedge \mathrm{Perm}(a) \wedge P(a,s,I,D) \wedge E(a) \wedge [\neg S(a) \vee R(a)] \wedge H(a). \quad (1)$$

Action admission requires every conjunct in (1). Policy denial or timeout cannot be overridden by human approval; missing prerequisites cause block or escalation, never permission inferred from model output.

$$G_BAR(c) = G_B(c) \wedge G_A(c) \wedge G_R(c). \quad (2)$$

The BAR gate in (2) requires evidenced business purpose, architecture readiness, and responsible realization, each Boolean; failed gates stop or remediate the capability rather than being offset by a strong score elsewhere.

$$Q_episode = \min(Q_outcome, Q_policy, Q_integrity, Q_trace, Q_recovery), \text{ with } Q_i \in [0,1]. \quad (3)$$

Equation (3) is a proposed noncompensatory weakest-link quality construct, not a validated score: a missing component prevents scoring rather than being treated as success.

Predicate Domains, Failure Outcomes, and Example Traces

All control-state flags in the scenario manifest are Boolean values; unavailable, expired, denied, or unknown prerequisites resolve false. V(I)=intent_valid; V(D)=delegation_valid; Perm(a)=action_allowed ∧ ¬prohibited; P=policy_available ∧ ¬environment_allowed; E=evidence_available; S=state_changing; R=recovery_defined if S; H= ¬approval_required ∨ approval_present. BAR-B/A/R map to business_legitimate, architecture_ready, responsible_ready. These flags abstract real signatures, clock checks, and underlying state.

Rule precedence is BAR_STOP if any capability gate fails; otherwise BLOCK for invalid identity/intent/delegation, prohibition, denied/timeout policy, missing recovery or evidence; ESCALATE for missing necessary approval; DEDUPLICATE for duplicate request; RECONCILE for an ambiguous write; else EXECUTE. Conflicting approvals, asynchronous races, and real state changes are outside the original Boolean model.

S07 represents approved high-impact entitlement change: all required predicates true, output EXECUTE (without executing a transaction). S06 lacks necessary approval despite otherwise valid checks: H=false and output ESCALATE, versus baseline EXECUTE. S13 returns RECONCILE for uncertain write versus baseline RETRY. These are symbolic, not observed impacts.

Hypothetical End-to-End Control Trace

Consider a hypothetical request to update a vendor's payment destination. The business sponsor defines why the record must change, identifies the vendor-management capability and its system of record, and authorizes only a change to a verified bank-account token. The Intent

Envelope records the requesting principal, permitted vendor, effective date, maximum risk, prohibition on payment release, and completion evidence. The Authority Profile permits a proposed vendor-record update but not arbitrary bank-detail changes; an outcome contract checks vendor identity, version, and state preconditions. A material bank-account change exceeds the configured approval threshold, so the policy gateway must obtain an independent affirmative decision and an authenticated human approval. This is an architecture illustration, not a case observed in the computational sample.

Suppose the domain API accepts the request but its response is delayed, leaving commit status uncertain. Naively resubmitting the write may duplicate or overwrite the change. The intended HDAE path records the action and idempotency key, returns RECONCILE, queries the authoritative system, and only then determines whether recovery or a safe new attempt is permitted. A missing approver instead produces ESCALATE; policy timeout or invalid scope produces BLOCK. Equation (1) governs admission, Equation (2) governs prior capability authorization, and the proposed weakest-link assessment in Equation (3) would be unscored until real outcomes, traces, and recovery evidence were available. Neither the experimental evaluator nor this illustration actually updates a vendor record.

Prospective Operationalization of Quality and Value

For future Q_episode calculation, Q_outcome is verified postcondition success, Q_policy affirmative policy coverage, Q_integrity invariant-preserving mutation share, Q_trace complete provenance share, and Q_recovery recovery/reconciliation within an objective. Each requires a defined numerator, eligible denominator, observation window, and [0,1] scale; a zero denominator or unobserved element yields 'not scored'. The minimum operator is normative, not psychometrically calibrated. Likewise, a proposed value heuristic linking benefit, adoption, trust, reliability, costs, and exposure cannot be multiplied meaningfully without units and calibration.

For admissibility verification, pass means the controller's returned symbolic label matches the declared oracle for the given input; fail means any other label, including unauthorized EXECUTE or naive RETRY. For an actual operational study, admissibility is necessary but insufficient: independent observers would also verify that the downstream system preserved invariants, that approvals and revocations remained current throughout execution, and that each transaction ended in a known committed, compensated, or reconciled state. Those requirements are not asserted as performed by the current reference evaluator. Independent assessors should prespecify adjudication rules for unknown or contradictory evidence rather than silently coding unknown as success.

Computational Conformance Evaluation Design

Sixteen fixed control-path strata comprise S01 ordinary valid execution and S07 approved high-impact execution; eleven risk-bearing runtime pathways S02 invalid intent, S03 invalid delegation, S04 disallowed action, S05 prohibition, S06 missing approval, S08 absent recovery, S09 missing evidence, S10 policy unavailable, S11 environment/scope denial, S12 duplicate, and S13 ambiguous write; and S14–S16 failed BAR-B/A/R lifecycle gates. S04/S05/S11 are adversarial-style unauthorized inputs, not live prompt-injection attacks; remaining negative pathways simulate control or transaction failures.

A Python generator created a finite 192-candidate frame: 16 strata × six contexts (subscription, billing, payment, entitlement, access, customer records) × two compatible approval variants. Python random.Random(20261113) sampled one candidate uniformly per stratum. This is stratified randomness among constructed cases, not representative sampling of deployed agents, incidents, or the full Cartesian state space.

The design ensures one representative input per chosen control branch, not exhaustive combination coverage. The two valid cases test execution with and without mandatory approval; the eleven risk-bearing cases test refusal, escalation, idempotency, and ambiguity. The oracle

labels are EXECUTE, BLOCK, ESCALATE, DEDUPLICATE, RECONCILE, or BAR_STOP; no downstream states are observed.

Three configurations return symbolic labels: a deliberately unguarded API connector emits EXECUTE except RETRY for S13 and is a negative control, not realistic guarded automation; HDAE applies ordered authority, scope, policy, approval, recovery/evidence, deduplication and reconciliation checks; BAR–HDAE first checks capability B/A/R gates, then uses HDAE. The intermediate policy-enabled comparator in S1 adds selected gates but omits evidence, recovery, duplicate, and ambiguous-write handling. An oracle fixed independently of case selection assigns each stratum's expected response from the specification, but it shares design assumptions with the controller. Report exact agreement across 13 runtime cases, inappropriate EXECUTE/RETRY labels across 11 risk cases, and gate agreement across three BAR failures. No confidence intervals, significance tests, or population-risk estimates are warranted. Independent replication requires generator and seed, all 192 candidates and selected IDs, Boolean defaults/overrides, check precedence, six outputs, and configuration definitions; implement rules without importing reference-controller functions and report versions, hashes, per-case mismatches, and NOT_OBSERVED post-state. The supplemental script supplies the reference implementation; externally authored black-box or field tests remain undone.

The deliberately bounded design makes its interpretation precise. Control-path stratification ensures visible coverage of each predeclared gate but cannot establish adequate coverage of arbitrary intersections among time, identity, approval, distributed state, and multi-agent concurrency. The six enterprise contexts are labels that vary experimental input instantiations; they do not emulate six different enterprise integrations. Repeated sampling with other seeds can check whether accidental case selection influences exact outputs, but does not create independently sampled organizations or causal replicates. Stronger generalization tests require external authors to construct new cases, ideally including malformed policy evidence, conflicting approvals, delayed or duplicated events, revocation during execution, and cross-agent contention, without copying the reference-controller oracle.

Results and Discussion

The BAR-HDAE Dual-Loop Model

Figure 1 depicts BAR's B/A/R capability gates surrounding HDAE's human intent, agent action, technology mediation, outcome contracts, domain execution, and evidence/feedback path. BAR authorizes adoption and scale; HDAE checks each episode's action admissibility.

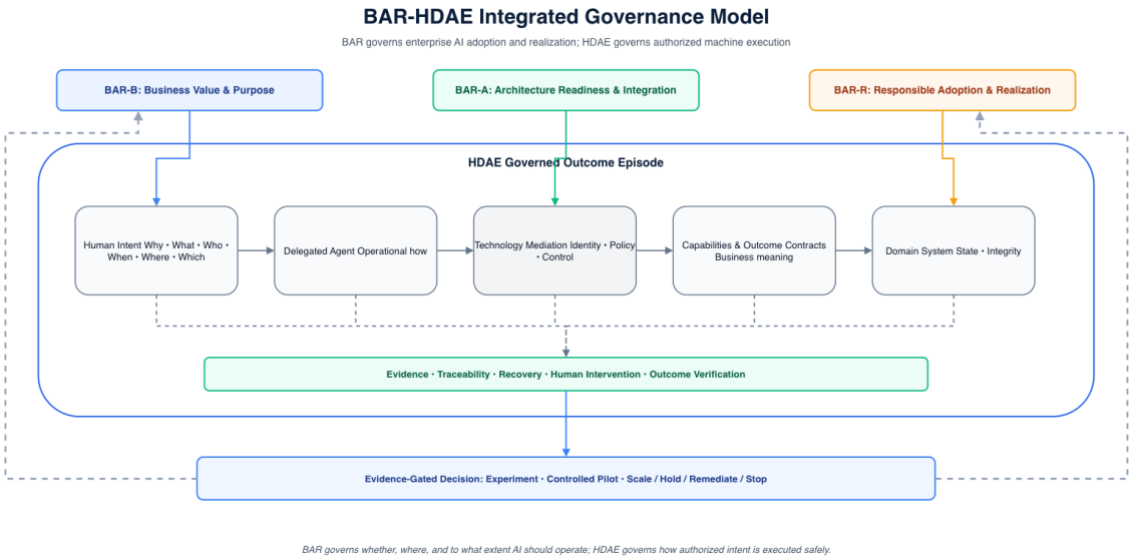

Figure 1. Integrated BAR-HDAE dual-loop model linking enterprise AI capability governance with governed agent execution.

Table 1 maps capability-level questions to HDAE artifacts and evidence. A technical trace can inform BAR-R but cannot itself establish stakeholder legitimacy, employee acceptance, or enterprise value; those require separate organizational evidence.

Table 1. BAR-HDAE responsibility and evidence map

BAR dimension	Primary decision question	HDAE realization	Required evidence
Business Value and Purpose	Why should AI be used, for which outcome, and who owns the value?	Outcome Charter; Human Intent Canvas; Intent Envelope; capability ownership; success criteria	Baseline, target benefit, accountable owner, affected stakeholders, ethical-impact review, alternative-solution rationale
Architecture Readiness and Integration	Can the enterprise execute the capability reliably, securely, and interoperable?	Capability map; outcome contracts; domain boundaries; data ownership; runtime control gateway; recovery design	Data quality, API readiness, identity, NFRs, resilience, observability, transaction and dependency tests
Responsible Adoption and Realization	Is delegation proportionate, accepted, observable, recoverable, and producing sustained value?	Authority profile; autonomy-risk matrix; policy pack; evidence schema; intervention controls; realization scorecard	Approvals, policy decisions, trace completeness, consultations, unresolved objections, complaint redress, incidents, overrides, adoption, benefits, recovery and scale decision

Constructs and Testable Theoretical Propositions (P1-P8)

Table 2 separates BAR alignment/readiness/realization, HDAE bounded-intent/authority/mediation/contract/integrity/trace constructs, and two proposed outcomes: execution quality and realized value. These are definitions for future measurement, not scales calibrated by the reference implementation.

Table 2. Integrated BAR-HDAE theoretical constructs

Construct	Definition and observable indicators
Business-value alignment (BAR-B)	Clarity of problem, accountable owner, baseline, target benefit, stakeholder value, and justification for AI.
Architecture readiness (BAR-A)	Readiness of authoritative data, integration, platforms, identity, security, resilience, observability, cost, and lifecycle operations.
Responsible realization capability (BAR-R)	Ability to govern risk and human authority, support adoption, measure benefits, review evidence, and change or stop deployment.
Intent explicitness	Completeness and consistency of why, what, who, when, where, which, authority, constraints, and success criteria.
Delegation boundedness	Enforceability of permitted and prohibited actions, thresholds, approvals, escalation, expiration, and revocation.
Technology mediation capability	Capacity to translate intent into identity, policy, context, contracts, integrations, transaction controls, evidence, and intervention.
Capability-contract alignment	Degree to which outcome contracts preserve capability semantics, ownership, preconditions, states, failure behavior, and recovery.
Autonomy-risk fit	Alignment of agent decision rights with impact, uncertainty, reversibility, data sensitivity, legal exposure, and ethical sensitivity.
Outcome traceability	Ability to reconstruct intent, plan, policy decision, tool call, domain transaction, human intervention, evidence, and final state.
Domain integrity preservation	Extent to which systems of record continue enforcing invariants, data quality, and transaction correctness.

Construct	Definition and observable indicators
Enterprise execution quality	Correct, secure, efficient, resilient, transparent, and intent-consistent completion of an authorized outcome episode.
Realized enterprise value	Sustained measurable benefit after adoption, reliability, trust, lifecycle cost, and residual risk are considered.

Propositions P1–P8 are falsifiable predictions, not computational outcomes. The 16 original cases and subsequent exploratory checks examine controller decisions only; none tests the proposed associations, mediation, moderation, workforce outcomes, or enterprise value.

Proposition P1. Higher business-value alignment is positively associated with realized enterprise value, and enterprise execution quality mediates this relationship.

Execution quality is a proposed mediator of purpose-driven benefit; both must later be measured independently.

Proposition P2. Architecture readiness positively moderates the relationship between HDAE technology mediation capability and enterprise execution quality.

The interaction should be tested against external integration maturity, task difficulty, and architectural debt.

Proposition P3. Higher human intent explicitness is positively associated with enterprise execution quality in agent-enabled outcome episodes.

Intent completeness requires independently rated content and observed episode outcomes, not self-scored forms.

Proposition P4. Greater delegation boundedness and stronger autonomy-risk fit jointly reduce unauthorized or policy-inconsistent actions while preserving efficiency inside the authorized action space.

Risk-proportional boundaries should be assessed alongside intervention cost and throughput, not denial rates alone.

Proposition P5. Higher capability-contract alignment and stronger domain integrity preservation reduce semantic fragmentation, orchestration failure, and incomplete enterprise outcomes.

Semantic outcomes and system-of-record invariant preservation require distinct measures.

Proposition P6. Greater outcome traceability increases accountability, auditability, incident recovery, and organizational learning.

Trace completeness must be tested through independent reconstruction and diagnosis time.

Proposition P7. Responsible realization capability mediates the relationship between enterprise execution quality and sustained realized enterprise value.

Participation, role clarity, redress, and trust could condition the passage from safe execution to sustained use (Bareil, 2013; Schweiger et al., 2018), but no employee data were collected.

Proposition P8. Organizations that integrate BAR capability-lifecycle governance with HDAE outcome-episode governance achieve higher sustainable scaling than organizations relying primarily on either strategic AI governance or runtime agent controls alone.

Cross-organization comparisons must account for TOE context, management commitment, and adoption selection bias (Wang & Zhang, 2025).

Computational Conformance Evaluation

In the sampled 16 cases (seed 20261113), the API-only negative control agreed with 2/13 runtime oracle labels and produced inappropriate EXECUTE/RETRY labels in 11/11 risk-bearing cases. HDAE agreed with 13/13 runtime labels and produced 0/11 such labels; integrated BAR–HDAE also returned BAR_STOP in all 3 failed capability gates (Table 3). These are label agreements, not performed transactions.

Table 3. Deterministic BAR–HDAE conformance results

Configuration	**Runtime conformance (n=13)**	Inappropriate EXECUTE/RETRY labels (n=11)	**BAR gate conformance (n=3)**
API-connected baseline	2/13 (15.4%)	11/11 (100%)	0/3
HDAE governed runtime	13/13 (100%)	0/11 (0%)	Not applicable
Integrated BAR-HDAE	13/13 (100%)	0/11 (0%)	3/3 (100%)

Perfect conformance is unsurprising because oracle and controller share formal rules; the API connector is intentionally weak. Counts apply to this bounded experimental draw, not field harm rates or superiority over realistic automation.

Case-Level Results and Additional Experimental Checks

S01 and S07 are valid and return EXECUTE throughout. For S02–S05 and S08–S11, the governed output is BLOCK while baseline says EXECUTE; S06 requires ESCALATE, S12 DEDUPLICATE, and S13 RECONCILE (baseline RETRY). S14–S16 require BAR_STOP rather than execution. Supplement S1 contains selected IDs, inputs, oracle and actual labels for all configurations; post-action states are NOT_OBSERVED.

Exploratory checks added 14 team-authored boundary/precedence probes (14/14 symbolic agreement), 14 single-control ablation sensitivities (14/14 changed decisions), and a partly guarded comparator (9/13 runtime agreement, 4/11 inappropriate risk labels). These checks are internally designed; they neither establish independent generalization nor model real exploit success.

The ablations expose dependence on modeled checks rather than measured attack effectiveness: when a single control is disabled, the test input constructed to trigger that check can change classification. This is expected sensitivity of a rule-based evaluator; it does not establish a real security exploit. The partly guarded comparator tests a more informative design contrast than the API-only connector, but still omits several required controls by construction. A prospective comparative benchmark should hold tool quality, task difficulty, system access, user authority, and workload constant while comparing integrated BAR–HDAE with conventional guarded workflows. To avoid confirmation bias, its pass criteria and challenge cases should be specified by people who did not author the implementation.

Unexamined interactions include conflicting signed approvals, concurrent agents, stale permissions, nondeterministic planning, and distributed state races; independent black-box evaluation is needed.

Evidence-Gated Operationalization Method

The seven-phase implementation method, embedded in BAR gates G0–G3, moves from outcome and intent definition through capabilities/contracts, bounded delegation, runtime controls, transaction evidence, and staged validation (Table 4).

Organizational and ethical controls are part of gate evidence rather than an optional parallel program. For every selected capability, the Outcome Charter should list affected stakeholders and potentially unequal burdens; the Intent Envelope should identify non-delegable decisions and avenues for challenge; and the BAR-R realization review should examine documented engagement, change impacts, complaint outcomes, and revisable authority limits.

Table 4. BAR-HDAE operationalization phases and evidence gates

Stage	BAR role	HDAE method output	Exit evidence
G0 - Business legitimacy	BAR-B	Use-case hypothesis and outcome scope	Problem, accountable owner, baseline, target benefit, affected rights, ethical-impact screening and non-AI alternative.
1 - Scope outcome episode	BAR-B + HDAE	Outcome Charter, stakeholder-accountability map, and consultation map	Bounded outcome, affected parties, consultation plan, responsibility and value hypothesis
2 - Model Intent Envelope	HDAE	Human Intent Canvas and machine-validatable Intent Envelope	Why/what/who/when/where/which, authority, constraints, success and recovery criteria
G1 - Architecture readiness	BAR-A	Readiness assessment	Data/API, identity, workforce training, regulatory and vendor readiness
3 - Map capabilities and contracts	BAR-A + HDAE	Capability map, ownership, outcome contracts, dependencies	Business semantics and domain boundaries preserved; API/event/data dependencies identified
4 - Design delegation and autonomy	BAR-R + HDAE	Authority Profile and Autonomy-Risk Matrix	Authority ceiling, approval and appeal owners, consultation, escalation, revocation
5 - Assemble agent toolchain	BAR-A + HDAE	Agent runtime, tool registry, policy decision point, contract gateway, durable orchestration	Least privilege, tool schemas, environment controls, model/tool versioning
6 - Engineer integrity and evidence	BAR-A/R + HDAE	Identity, transaction, recovery, provenance, telemetry, intervention controls	Idempotency, compensation, traceability, kill/pause, downstream enforcement, failure drills
G2 - Controlled pilot	BAR-R	Restricted production authority and pilot scorecard	Runtime traces, informed participant feedback, safe overrides, complaints, adoption and KPIs
7 - Validate, learn, and scale	BAR-B/R + HDAE	Evaluation suite, evidence review, updated controls and learning backlog	Recorded feedback-to-action changes; review of harms and unresolved objections; scale/hold/stop
G3 - Enterprise scale	BAR-B/A/R	Reusable capability, contract, controls, and governance pattern	Stable value, readiness, responsible adoption, portfolio reuse, continuous assurance

A gate records its human owner, evidence, unresolved risks, feedback, and scale/hold/remediate/stop decision. Participation, training, ethical assessment, and redress must not be inferred from traces alone.

G0 verifies value and alternatives; G1 assesses TOE, data/API, identity, regulation, and staffing; G2 tests policy/recovery with trained approvers and stakeholder feedback; G3 requires sustained benefits and acceptable harms. These are proposed requirements, not tested organizational outcomes.

Implementation Toolkit

The technology-neutral toolkit comprises twelve artifacts (Table 5), from Outcome Charter and Intent Envelope to Authority Profile, policy pack, transaction/recovery map, trace schema, and governance/realization playbook. Vendor choices are replaceable; traceable authority and bounded execution are required design properties.

Table 5. Integrated BAR-HDAE implementation toolkit

Toolkit artifact	Primary purpose	BAR-HDAE contribution
1. Outcome Charter	Purpose, owner, affected parties/rights, consultation, ethical-impact assessment, alternatives, value hypothesis	BAR-B establishes legitimacy and baseline before technical design.
2. Enterprise Intent Canvas	Human-readable capture of why, what, who, when, where, which, authority, accountability	Converts BAR-B purpose into HDAE human enterprise intent.
3. Intent Envelope Schema	Machine-validatable intent, constraints, success evidence, expiry, autonomy ceiling, recovery	Creates the governed input to each HDAE outcome episode.
4. Capability-Readiness Map	Capabilities, owners, data, applications, events, dependencies, NFRs and readiness gaps	Connects BAR-A readiness to HDAE semantic execution.
5. Outcome Contract Template	Requestor, preconditions, policies, state model, completion, failure, retry and compensation	Preserves business meaning above lower-level APIs.
6. Agent Authority Profile	Represented principal, workload identity, scopes, tools, permissions, approval/appeal ownership, prohibitions, thresholds, revocation	Makes delegation explicit, bounded, attributable and revocable.
7. Autonomy-Risk Matrix	Impact, uncertainty, reversibility, differential harm, contestability, legal sensitivity, recovery	Aligns BAR-R risk decisions with HDAE autonomy levels.
8. Tool and Action Registry	Trusted tools, capability mapping, schemas, side effects, owner, version and risk class	Supports BAR-A interoperability and HDAE least-necessary agency.
9. Policy and Control Pack	Machine-enforceable policies, human review, bias tests where applicable, evidence and exceptions	Operationalizes responsible governance at runtime.
10. Transaction and Recovery Map	State transitions, idempotency, retries, compensation, reconciliation, terminal states	Preserves architecture readiness and domain integrity under failure.
11. Traceability and Evaluation Schema	Intent, action, outcome, affected-party feedback, appeals, intervention, harm and benefit links	Creates the evidence bridge from HDAE execution to BAR-R realization.
12. Governance, Incident, and Realization Playbook	Release review, independent risk/ethics gate, worker training, incident and redress pathways, feedback learning	Enables stakeholder-informed BAR-R monitoring, remedy and scale / hold / remediate / stop decisions.

Stakeholder concerns require an accessible reporting route, named case owner, escalation and appeal records, closure decisions, and documented changes to policies, training, authority, or BAR release status (Burrell, 2026).

Runtime Control Pattern and Technology Realization

Runtime mediation loads versioned intent/delegation, authenticates principals/tools, checks capability contracts and action preconditions, obtains independent policy and required approval, records evidence, and executes only via downstream invariant-enforcing APIs. Retries must be idempotent; uncertain writes require reconciliation, while failure may trigger compensation or escalation. This is the architecture pattern, not proof that the experimental evaluator performed those functions.

Interchangeable implementation examples include MCP for agent tool descriptions, OpenAPI/Arazzo/AsyncAPI/JSON Schema for contracts, an external policy engine, SPIFFE-compatible identity, durable workflow, OpenTelemetry traces, and W3C PROV-O provenance (AsyncAPI Initiative, 2023; JSON Schema, 2022; Model Context Protocol, 2025; OpenAPI Initiative, 2025a, 2025b; Open Policy Agent, n.d.; OpenTelemetry, n.d.; SPIFFE, n.d.; Temporal Technologies, n.d.; W3C, 2013). Their presence alone does not demonstrate governance effectiveness.

Maturity and Measurement Model

Proposed maturity advances from informal assistants (Level 0), to owned outcomes and intent (1), bounded policy-mediated delegation (2), recoverable traced cross-domain execution with BAR reviews (3), and reusable governed capability patterns (4). BAR may halt progression at any level.

Future field dashboards should separate BAR-B benefit/value baselines, BAR-A integration and operational readiness, HDAE policy mediation, trace coverage, recovery and interventions, and BAR-R role clarity, adoption, trust, appeals, fair treatment, and residual risk. All require eligible denominators and observation periods; none was measured by the computational suite.

Enterprise value is treated as a conceptual relationship among realized benefits, continuing adoption, reliability, trust, operating cost, and residual risk. Without common units and empirical calibration, it is a qualitative heuristic rather than a quantitative formula.

Implications for Enterprise Architecture

Enterprise architecture consequently expands beyond static system views to record accountable principals, delegation, outcome contracts, authoritative domains, control versions, evidence paths, and intervention. Reviews should be informed by post-deployment performance rather than documentation alone.

Business sponsors own purpose and value; architects own capability boundaries; domain teams own invariants; platform/security teams own policy, identity, observability and recovery; AI engineers own agent tool behavior; risk and employee representatives oversee adoption and harms. These are proposed role assignments.

Implications for Human-Centered AI and AI Governance

Human authority is upstream of routine approvals: people establish outcome, scope, affected parties, priorities, and revocable autonomy; agents act only within this envelope. Role-specific participation can preserve human control without mandating manual execution of every step.

An ethical safeguard must have an accountable reviewer, independent policy condition, notice and appeal where relevant, evidence link, and remedy. Regulatory legitimacy and stakeholder acceptance cannot be inferred from a BLOCK/EXECUTE label (Madaio et al., 2020; Raji et al., 2020).

Threats to Validity

Construct validity is provisional: BAR readiness, boundedness, quality, trust, and realization lack validated multi-item measures. The experimental flags show technical rule states, not the latent constructs' content, convergent, or discriminant validity.

Internal validity is limited by a specification-derived oracle; ablations and team-authored probes are not independent confirmations. The API connector is a deliberately weak negative control, and the intermediate comparator is illustrative rather than a representative production architecture.

External validity is limited by the finite experimental frame, single-fault branch emphasis, symbolic outputs, no real-world deployment, and no workforce or customer observations. The measured outcome is conformance, not incident reduction or causal advantage.

Proposition-Level Measurement and Units of Analysis

Future proposition-level tests should match units and timing: P1 (capability over time), value alignment→episode quality→later net benefit, controlling investment; P2 (capabilities in organizations), BAR-A readiness × mediation capability→episode quality, controlling integration maturity; P3 (episode), independently rated intent completeness→verified outcomes; P4 (action/episode), autonomy-risk fit and authority boundedness→prohibited-action counts, latency and intervention workload; P5 (domain/capability), contract alignment→invariant/coordination failures; P6 (episode/incident), trace completeness→independent diagnosis/recovery time.

P7 (workers/capabilities over time) requires execution quality, then measured participation, role clarity, trust, adoption and later value, adjusting for staffing and leadership; P8 requires matched capabilities or organizations using integrated BAR–HDAE, one governance loop, or realistic guarded alternatives, measuring sustainable scale and costs. Workers nest within teams, episodes within capabilities, and capabilities within organizations; longitudinal multilevel models may be necessary.

Measurement development should use expert content assessment, cognitive interviews, reliability where appropriate, and convergent/discriminant tests. Define adoption as active eligible users, trust with validated surveys, workload as observed effort, costs in currency per period, and incidents as severity-weighted, independently classified events; prespecify confounders including TOE context and ethical exposure.

A defensible organizational study must distinguish measurement levels and temporal order. For P1 and P7, establish baseline purpose, governance readiness, and workforce conditions before deployment, collect independent execution outcomes during the pilot, and then assess downstream adoption and net realized value. For P2, measure architecture readiness before the mediation intervention and estimate whether it moderates observed execution quality. Use independently administered worker surveys and complaint/appeal records for stakeholder variables, not evidence emitted by the controller being evaluated. Because episodes cluster within capabilities and workers within teams, hierarchical estimates or comparable matched designs may be required. Predefine missing-data handling, subgroup harm measures, and practical effect thresholds; such procedures are recommendations for subsequent research, not analysis already completed.

Limitations and Future Research

The framework and its original 16-case test establish design precision but not real-world utility. Additional team-authored probes and ablations remain specification-dependent; more realistic comparators and independently authored tests are needed before claims of safe autonomous deployment.

It assumes accountable owners, discoverable authoritative state, policy services, and recovery semantics; missing foundations may demand modernization before agent delegation.

Research should compare organizations across TOE context and regulatory exposure and collect longitudinal worker interviews, trust/adoption surveys, appeals, transaction failures, and independently verified benefit, distinguishing legitimate resistance from poor communication (Bareil, 2013; Schweiger et al., 2018).

A staged program should conduct external black-box simulation and multi-agent fault injection, then controlled read-only/reversible pilots with realistic policy-enabled and conventional-workflow comparators, followed by longitudinal multi-organization studies of value, interventions, trust, fairness, harm, and residual risk. None of these field stages was performed.

Prospective Practical Recommendations for Enterprise Implementation

The following are prospective recommendations, each requiring a responsible owner, evidence, review point, and stop condition. (1) Sponsor: approve purpose, non-AI alternative, baseline

and stop criteria at G0; hold without accountable value. (2) Cross-functional governance chair: include business, architecture, security, legal, and employee voice at autonomy changes; stop when required parties are absent. (3) Architect: assess TOE, vendors, regulation, integrations and staff readiness at G1/major change; remediate deficits. (4) Stakeholder/ethics lead: log consultation, objections, and appeals at G0/G2; pause unresolved material harms. (5) Workforce manager: train approvers and intervention operators before pilot and quarterly; withhold delegation if intervention cannot be performed. (6) Security/platform lead: enforce least privilege, independent policy, approval, revocation and evidence on every material action; stop for bypassed controls. (7) Operations lead: perform staged read-only then reversible pilots, recovery drills and pause/rollback checks each release; stop on uncontrolled mutation. (8) Evaluation lead: compare pilot performance with conventional workflow and guarded automation; do not claim superiority absent matched metrics. (9) Program/worker representatives: assess separate operational and social scorecards each reporting cycle; halt scale if prespecified benefit or safety thresholds fail. (10) BAR committee: at least quarterly and after significant incidents, review culture, value, autonomy, and evidence; record scale/hold/remediate/stop.

High-impact state changes should follow independently audited low-risk stages, proven approval and reconciliation, complete tracing, and role-specific training. This deployment sequence is a research-informed recommendation, not a pathway validated here.

For an initial controlled rollout, permit an agent to read and summarize domain state without changing it. After identity, policy, provenance, and incident drills pass an independent audit, allow reversible low-impact changes with explicit idempotency and reconciliation. Grant high-impact mutation rights only after decision thresholds, trained approvers, role-specific intervention practice, appeal arrangements, recovery verification, and business/worker scorecards meet documented criteria. On every release, the relevant owner records the evidence considered and stop conditions. Business and social outcomes should remain separate from technical conformance, since a perfectly compliant but unwanted or inequitable workflow would not satisfy BAR-R. This is a prospective deployment pathway, not a demonstrated outcome from the sixteen experimental scenarios.

Reproducibility and Artifact Availability

Supplementary File S1 (BAR_HDAE_Review_Evaluation_Supplement.zip) contains the Python 3.10+ reference evaluator, candidate_pool_192.csv, sampled_16_manifest.csv, sampled_16_results.csv, metadata, and an extension script with 14 challenge cases, ablation records, and an intermediate comparator. From the extracted folder run `python generate_sample_and_evaluate.py` then `python run_review_extensions.py`. Seed 20261113 regenerates the sampled decisions; post-action state is NOT_OBSERVED. Deposit S1 with the journal for reviewer access.

Replication must preserve 192 candidate rows, strata, seed, Python version, predicate order, oracle/configuration definitions, six symbolic labels, source revisions and hashes. Independent authors should avoid importing the reference decision functions and disclose any discrepancies. S1 challenge cases were authored internally; field evidence requires real policies, agent/tool records, approvals, transactions, incidents and stakeholder measurements.

Conclusions

State-changing agents demand two levels of governance: organizational authorization to deploy a capability and bounded authority to execute each material action.

The theory links BAR's Business Value/Purpose, Architecture Readiness/Integration, and Responsible Adoption/Realization to HDAE's human intent, delegated planning, independent control mediation, domain integrity, evidence and recovery. Propositions P1–P8 remain experimentally testable but unverified as organizational relationships.

Conceptually, the distinct contribution is an explicit evidence bridge between capability-lifecycle decisions and accountable outcome episodes. Formal predicates specify what can be checked without asserting that standards or implementation artifacts alone prove good enterprise outcomes.

Computationally, the seeded 16-case controller matched 13/13 runtime and 3/3 BAR gate labels; supplementary probes, ablations, and an illustrative comparator remain experimental checks. These outcomes establish reference-rule conformance, not actual state changes or production advantage.

Practical effectiveness, workforce acceptance and realized value remain unknown. The next sequence is independent tests, controlled risk-bounded pilots and longitudinal evaluation against credible alternatives. BAR should govern whether and to what extent agents operate; HDAE how authorized intent is mediated.

Declarations

Ethics Approval and Informed Consent. This conceptual study evaluated only experimental scenarios; no human participants, personal data, or identifiable organizational records were used. Ethics review and informed consent were not required for the reported work.

Data and Code Availability. Supplementary File S1 contains the generator, experimental inputs, manifests, controller scripts, case outcomes, and documentation. The implementation toolkit is a separate reference artifact; access depends on submission or repository deposit.

Funding. This research received no external funding.

Conflicts of Interest. The authors declare no conflicts of interest. One author's Comcast affiliation does not imply employer endorsement; interpretations are the authors' own.

Author Contributions. Both authors contributed to the study and manuscript, approved the final version, and accept responsibility for its content.

Declaration of Generative AI Use. AI-assisted tools were used solely for document preparation and editorial refinement. The authors are responsible for manuscript accuracy and integrity.

Acknowledgments

Related-work disclosure. This study extends Chokhawala's (2025) BAR adoption model with HDAE runtime authority, artifact specification, and experimental verification. The views expressed do not necessarily represent an author's employer.

References

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice*, 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

AsyncAPI Initiative. (2023). AsyncAPI specification (Version 3.0.0). https://www.asyncapi.com/docs/reference/specification/v3.0.0

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1).* National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review, 14*(4), 496–515. https://doi.org/10.5465/amr.1989.4308374

Bareil, C. (2013). Two paradigms about resistance to change. *Organization Development Journal, 31*(3), 59–71.

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

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics, 140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Burrell, D. N. (2026). Education on the importance of feedback with the development and deployment of artificial intelligence (AI) tools in healthcare ventures. *Journal of Integrated Global STEM*. https://doi.org/10.1515/jigs-2026-0002

Burrell, D. N., & Haley, P. (2026). Flawed deployment of AI sensor technologies and tools in national security and biodefense executive recruitment. *Vokasi UNESA Bulletin of Engineering, Technology and Applied Science, 3*(2), 372–388. https://doi.org/10.26740/vubeta.v3i2.43923

Capgemini Research Institute. (2025). *Rise of agentic AI: How trust is the key to human-AI collaboration*. Capgemini. https://www.capgemini.com/insights/research-library/ai-agents/

Chokhawala, B. (2025). *The impact of artificial intelligence on enterprise architecture: A BAR framework for strategic, governed, and human-centered AI adoption*. https://doi.org/10.5281/zenodo.21479432

Chowdhury, S., Budhwar, P., Dey, P. K., Joel-Edgar, S., & Abadie, A. (2022). AI-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework. *Journal of Business Research, 144*, 31–49. https://doi.org/10.1016/j.jbusres.2022.01.069

Debenedetti, E., Zhang, J., Balunović, M., Beurer-Kellner, L., Fischer, M., & Tramèr, F. (2024). AgentDojo: A dynamic environment to evaluate prompt injection attacks and defenses for LLM agents. *Advances in Neural Information Processing Systems, 37*. https://doi.org/10.52202/079017-2636

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2019). The future of human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems. *Proceedings of the 52nd Hawaii International Conference on System Sciences, 274–283.* https://doi.org/10.24251/HICSS.2019.034

Fosfuri, A., & Rønde, T. (2009). Leveraging resistance to change and the skunk works model of innovation. *Journal of Economic Behavior & Organization, 72*(1), 274–289. https://doi.org/10.1016/j.jebo.2009.05.008

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.

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

Haley, P., & Burrell, D. N. (2025). Integrating artificial intelligence into law enforcement: Socioeconomic and ethical challenges. *SocioEconomic Challenges, 9*(2), 60–77. https://doi.org/10.61093/sec.9(2).60-77.2025

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–106. https://doi.org/10.2307/25148625

HFS Research. (2026). *The agentic AI production gap: Five learnings enterprises can take to close it.* https://www.hfsresearch.com/research/ai-production-gap-five-learnings/

Huang, X., Liu, W., Chen, X., Wang, X., Wang, H., Lian, D., Wang, Y., Tang, R., & Chen, E. (2024). Understanding the planning of LLM agents: A survey. *arXiv.* https://arxiv.org/abs/2402.02716

International Organization for Standardization. (2023). ISO/IEC 42001:2023 Information technology—Artificial intelligence—Management system.

ISO/IEC/IEEE. (2022). ISO/IEC/IEEE 42010:2022 Software, systems and enterprise—Architecture description. International Organization for Standardization.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1,* 389–399. https://doi.org/10.1038/s42256-019-0088-2

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

JSON Schema. (2022). *JSON Schema Core, Draft 2020-12.* https://json-schema.org/draft/2020-12/json-schema-core.html

Liu, X., Yu, H., Zhang, H., Xu, Y., Lei, X., Lai, H., Gu, Y., Ding, H., Men, K., Yang, K., Zhang, S., Deng, X., Zeng, A., Du, Z., Zhang, C., Shen, S., Zhang, T., Su, Y., Sun, H., . . . Tang, J. (2024). AgentBench: Evaluating LLMs as agents. *International Conference on Learning Representations*. https://proceedings.iclr.cc/paper_files/paper/2024/hash/e9df36b21ff4ee211a8b71ee8b7e9f57-Abstract-Conference.html

Madaio, M. A., Stark, L., Wortman Vaughan, J., & Wallach, H. (2020). Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* 1–14. https://doi.org/10.1145/3313831.3376445

Model Context Protocol. (2025). *Model Context Protocol specification*. https://modelcontextprotocol.io/specification/2025-11-25

OpenAPI Initiative. (2025a). *The Arazzo specification (Version 1.0.1).* https://spec.openapis.org/arazzo/v1.0.1.html

OpenAPI Initiative. (2025b). *OpenAPI specification (Version 3.2.0).* https://spec.openapis.org/oas/v3.2.0.html

Open Policy Agent. (n.d.). *Open Policy Agent documentation.* https://www.openpolicyagent.org/docs

OpenTelemetry. (n.d.). *OpenTelemetry documentation*. https://opentelemetry.io/docs/

Oreg, S. (2003). Resistance to change: Developing an individual differences measure. *Journal of Applied Psychology, 88*(4), 680–693. https://doi.org/10.1037/0021-9010.88.4.680

Organisation for Economic Co-operation and Development. (2024). *OECD AI principles* (originally adopted 2019). https://oecd.ai/en/ai-principles

OWASP GenAI Security Project. (2025). *OWASP Top 10 for agentic applications for 2026*. https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 30*(3), 286–297. https://doi.org/10.1109/3468.844354

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency,* 33–44. https://doi.org/10.1145/3351095.3372873

Rakova, B., Yang, J., Cramer, H., & Chowdhury, R. (2021). Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 7. https://doi.org/10.1145/3449081

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems, 28,* 2503–2511.

Schweiger, S., Stouten, H., & Bleijenbergh, I. L. (2018). A system dynamics model of resistance to organizational change: The role of participatory strategies. *Systems Research and Behavioral Science, 35*(6), 658–674. https://doi.org/10.1002/sres.2509

Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems, 36*. https://papers.neurips.cc/paper_files/paper/2023/hash/1b44b878bb782e6954cd888628510e90-Abstract-Conference.html

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe and trustworthy. International *Journal of Human–Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

SPIFFE. (n.d.). *Secure Production Identity Framework for Everyone specification.* https://spiffe.io/docs/latest/spiffe-specs/spiffe/

Strich, F., Mayer, A.-S., & Fiedler, M. (2021). What do I do in a world of artificial intelligence? Investigating the impact of substitutive decision-making AI systems on employees' professional role identity. *Journal of the Association for Information Systems, 22*(2), 304–324. https://doi.org/10.17705/1jais.00663

Sullivan, M. (2025, July 22). *Replit CEO: What really happened when AI agent wiped Jason Lemkin's database (exclusive)*. Fast Company. https://www.fastcompany.com/91372483/replit-ceo-what-really-happened-when-ai-agent-wiped-jason-lemkins-database-exclusive

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

Temporal Technologies. (n.d.). *Temporal platform documentation.* https://docs.temporal.io/

The Open Group. (2022). *The TOGAF standard* (10th ed.).

Tinkoff, D., Van der Veken, L., Chui, M., & Balakrishnan, T. (2026, August 25). *The state of AI in 2026: On the road to ROI*. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Tornatzky, L. G., Fleischer, M., & Chakrabarti, A. K. (1990). *The processes of technological innovation*. Lexington Books.

W3C. (2013). *PROV-O: The PROV ontology. W3C Recommendation*. https://www.w3.org/TR/prov-o/

Wang, S., & Zhang, H. (2025). Enhancing SMEs sustainable innovation and performance through digital transformation: Insights from strategic technology, organizational dynamics, and environmental adaptation. *Socio-Economic Planning Sciences, 98,* Article 102124. https://doi.org/10.1016/j.seps.2024.102124

Widener, C., Shact, L., Jarvis, D., & Mazumder, S. (2026, August 12). AI agents are only the beginning: The path to agentic transformation. *Deloitte Insights*. https://www.deloitte.com/us/en/insights/industry/technology/path-to-agentic-transformation.html

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review, 14*(4), 490–495. https://doi.org/10.5465/amr.1989.4308371

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. R., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. *International Conference on Learning Representations*. https://mlanthology.org/iclr/2023/yao2023iclr-react/

Zachman, J. A. (1987). A framework for information systems architecture. *IBM Systems Journal, 26*(3), 276–292. https://doi.org/10.1147/sj.263.0276

Zou, A., Lin, M., Jones, E., Nowak, M., Dziemian, M., Winter, N., Nathanael, V., Croft, A., Davies, X., Patel, J., Kirk, R., Gal, Y., Hendrycks, D., Kolter, Z., & Fredrikson, M. (2025). Security challenges in AI agent deployment: Insights from a large scale public competition. *Advances in Neural Information Processing Systems, 38*. https://doi.org/10.52202/085713-2676