WISDOMTWIN RESEARCH

Judgment Latency in Regulated Enterprises

A Conceptual Framework and Research Agenda for Role-Specific AI Decision Support

Roman Bodnarchuk

Co-Founder and Chief Executive Officer, WisdomTwin, Inc.
Toronto, Ontario, Canada
roman@wisdomtwin.ai
ORCID: 0009-0004-3113-2118

Series record	WisdomTwin Working Paper WT-100-001
Version	2.0, fact-checked revision
Date	September 3, 2026
Status	Preprint; not peer reviewed; DOI pending deposit; CC BY 4.0

Scope note. This paper proposes constructs, design requirements, testable propositions, and an evaluation protocol. It reports no live-enterprise deployment and no causal product-performance result.

Suggested citation: Bodnarchuk, R. (2026). Judgment latency in regulated enterprises: A conceptual framework and research agenda for role-specific AI decision support (WisdomTwin Working Paper WT-100-001, Version 2.0). WisdomTwin, Inc.

Abstract

Organizations can possess the information required for a decision while still waiting for a particular role to attend to it. This paper defines judgment latency as the elapsed time between decision readiness and attention by the person or forum whose judgment is required. It develops a conceptual framework for studying that interval and proposes the AI Judgment Twin as a role-specific, evidence-grounded decision-support architecture. The proposed system is not a personal replica and does not independently hold organizational authority. It is intended to retrieve authorized precedents, expose the principles, thresholds, exceptions, and escalation patterns reflected in prior decisions, and produce cited recommendations within an explicitly validated boundary. The argument integrates research on work interruption, meeting science, decision rights, organizational memory, digital and human twins, and human reliance on automation. It also specifies a non-overlapping model of potentially recoverable time, distinguishes externally reported estimates from author-derived arithmetic, and states six propositions suitable for empirical testing. For regulated settings, the paper treats provenance, least-privilege access, human confirmation, safe decline, monitoring, and contestability as design requirements rather than demonstrated product capabilities. A staged evaluation design is proposed, beginning in shadow mode and advancing only if decision quality and permission integrity meet pre-specified thresholds. The paper is conceptual and theory-building. WisdomTwin is pre-revenue, has no production users, and has not validated the proposed effects in a live enterprise.

Keywords: judgment latency; AI-assisted decision making; organizational memory; decision rights; regulated enterprises; human oversight; provenance; asynchronous management

Claim-status convention

To keep evidence, calculation, and hypothesis distinct, this paper uses the following status categories.

Status	Meaning in this paper
Established evidence	A bounded statement supported by the cited empirical, scholarly, or official source.
Author-derived	Arithmetic or synthesis performed here from cited inputs; not reported by the underlying source.
Conceptual proposition	A testable theoretical claim that has not been established for the proposed system.
Illustrative	A synthetic example or scenario; not an observed organization, user, outcome, or forecast.
Design requirement	A capability or control that a conforming implementation should satisfy; not a claim that the capability has been validated.

1. Introduction

Enterprise decisions often wait even after the relevant facts are available. A contract exception may wait for legal review, a variance for a controller, a protocol change for clinical leadership, or a vendor approval for compliance. Existing work describes several adjacent mechanisms, including decision rights, meeting processes, interruptions, organizational memory, and information retrieval. It does not, however, provide a settled construct for the time spent waiting specifically for role-bound judgment. This paper calls that interval judgment latency.

The practical setting is important but the popular statistics used to describe it require care. Microsoft reported in 2023 that, within measured intentional activity across selected Microsoft 365 applications, 57% of time was spent in meetings, email, and chat and 43% in documents, spreadsheets, and presentations. This is

product telemetry, not a complete time-use census or a universal estimate for all knowledge workers (Microsoft, 2023). A 2025 Microsoft report stated that employees were interrupted every two minutes during core hours, but its methodology specifies that this statistic was calculated for the top 20% of users by ping volume and excluded education and European Union tenants (Microsoft, 2025). It should not be generalized to the average worker.

Peer-reviewed interruption studies support a more defensible claim: work is fragmented, task switching can leave attention on the prior task, and interruption effects vary by task, person, timing, and context (Leroy, 2009; Mark et al., 2005; Puranik et al., 2020). Mark et al. (2008) found that participants completed interrupted tasks faster but experienced greater stress, frustration, time pressure, and effort. That experiment did not establish a universal 20-minute recovery time, so this paper does not use that figure.

A related decision-making survey by McKinsey & Company reported that 61% of more than 1,200 surveyed managers believed at least half of decision-making time was ineffective. Under assumptions for a typical Fortune 500 company, the authors estimated about 530,000 manager-days and approximately USD 250 million in wages annually (De Smet et al., 2019). Section 8 reproduces the calculation and clearly separates McKinsey's estimate from this paper's author-derived weekly transformation. Neither number is evidence of savings available from an AI system.

The paper makes four contributions. First, it defines judgment latency as one component of broader decision latency and specifies its unit of analysis. Second, it distinguishes role-specific decision support from personal simulation. Third, it translates governance principles into testable architecture requirements. Fourth, it offers propositions, a non-overlapping economic model, and an evaluation protocol that can falsify the core claims. The proposed AI Judgment Twin remains an unvalidated concept at the pre-pilot stage.

2. Method and scope

This is an integrative conceptual paper, not a systematic review, meta-analysis, field study, or product evaluation. The source set began with the literature cited in version 1.0 and was expanded purposively where a stronger peer-reviewed review, original empirical study, or current official source was needed. The selection emphasizes work interruption, meeting science, organizational memory, decision rights, digital and human twins, human-automation reliance, and AI governance. Because the search was targeted rather than protocol-driven, the review should not be treated as exhaustive.

Bibliographic metadata and the factual claims carried forward were checked against publisher pages, article versions of record, or official government and institutional sources on September 3, 2026. A companion source-verification log records the verification result and material caveat for each source. One draft citation, a 2026 newsletter item attributed to Bodnar, could not be located in an independently accessible original and was removed rather than reconstructed from memory.

The proposed architecture is analyzed as a design object. Statements about what an implementation should do are normative design requirements. Synthetic scenarios clarify mechanisms but provide no empirical support. The economic examples are sensitivity analyses, not forecasts. Legal and regulatory discussion identifies issues for local review and is not legal advice.

3. Related literature

3.1 Work fragmentation and attention

In a field study of 24 information workers in one organization, Mark et al. (2005) observed highly fragmented activity and reported an average duration of roughly 11 minutes for a working-sphere segment. The narrow sample and setting limit generalization. Experimental work by Mark et al. (2008) further showed that interruption can change work strategy and subjective workload even when elapsed task time does not increase. Leroy (2009) demonstrated attention residue across task transitions, while Puranik et al. (2020), reviewing 247

publications, emphasized that interruptions have heterogeneous causes and outcomes. The literature therefore supports measuring fragmentation in the target workflow rather than importing a single recovery-time constant.

Perlow's (1999) field intervention with software engineers showed that interdependent requests and individual concentration can form a self-reinforcing time pattern. Scheduled quiet-time phases were associated with higher self-reported productivity for many participants, but the single-site design does not establish a general effect size. Together, these studies motivate a workflow-level question: can some coordination requests be answered without imposing a synchronous attention switch on the role holder?

3.2 Meetings, decision rights, and speed

Meetings are not inherently wasteful. A review of 253 publications describes workplace meetings as important sites of leading, interaction, time management, engagement, and relationship work (Allen & Lehmann-Willenbrock, 2023). Meeting satisfaction is also a distinct facet of job satisfaction, with its relationship to overall job satisfaction becoming stronger as meeting demands rise (Rogelberg et al., 2010). The appropriate target is therefore not meetings as a category, but avoidable attendance, repeated context reconstruction, and waiting caused by unclear authority.

Consulting evidence complements the scholarly literature but should be interpreted as observational and practice-oriented. Mankins et al. (2014) analyzed time budgets in 17 large corporations and argued that organizational time is rarely governed with the discipline applied to capital. Rogers and Blenko (2006) focused on decision-role clarity, while De Smet et al. (2019) distinguished big-bet, cross-cutting, delegated, and ad hoc decisions and reported a strong survey association between faster and higher-quality decisions. Correlation does not show that making a decision faster causes it to become better.

3.3 Organizational memory and role knowledge

Walsh and Ungson (1991) framed organizational memory as stored information from organizational history that can influence present decisions through acquisition, retention, and retrieval. Nonaka (1994) explained knowledge creation as an interaction between tacit and explicit knowledge. These traditions support the importance of historical action and retrieval, but they do not imply that tacit judgment is fully captured in digital records. Silence, undocumented conversations, political constraints, and unrecorded context can all make the record incomplete.

A decision-support system built on historical records therefore models an evidentiary trace of past practice, not the complete mind of an incumbent and not an objective account of the best decision. The distinction is central: retrieval can preserve precedent while also preserving historical bias, obsolete policy, and prior error.

3.4 From digital twins to role-specific agents

Digital twins emerged as virtual representations connected to physical entities and processes (Grieves & Vickers, 2017). A systematic review found substantial definitional variation and identified fidelity, data ownership, and integration as open questions (Jones et al., 2020). Human digital twin research extends the label to representations of people, often using physiological, behavioral, or contextual data (Lin et al., 2024). Generative-agent research demonstrates that language-model agents can simulate patterns of human behavior in bounded environments (Park et al., 2023). None of these literatures establishes that a language model can faithfully exercise the judgment of an organizational role in regulated practice.

The term AI Judgment Twin is therefore used here as a proposed design label, not as a claim of equivalence to an engineering digital twin. Its intended object is narrower than a person: a validated representation of decision precedent within a defined role, purpose, data boundary, and authority level.

3.5 Human reliance and governance

Automation can be used, misused, disused, or abused when trust and system capability are misaligned (Parasuraman & Riley, 1997). In AI-assisted decision making, explanations alone may not prevent overreliance;

interaction designs that require active evaluation can sometimes reduce it (Buçinca et al., 2021). These findings argue against treating provenance or a confidence score as sufficient safety mechanisms. Review must be meaningful, contestable, and proportionate to consequence.

Current governance sources reinforce this position. The NIST AI Risk Management Framework calls for governed, mapped, measured, and managed risk processes (Tabassi, 2023), and its generative-AI profile identifies risks including confabulation, data privacy, information security, and human-AI configuration (National Institute of Standards and Technology, 2024). In July 2026, Canada's Office of the Superintendent of Financial Institutions advised federally regulated financial institutions to treat AI outputs as inputs rather than definitive outcomes, control data provenance and permissions, log activity, limit autonomy, and retain accountable human oversight for material decisions (OSFI, 2026).

4. The judgment-latency construct

4.1 Position within the decision lifecycle

A decision episode can be decomposed into at least four intervals: recognition, information assembly, judgment, and execution. Recent engineering work has used decision latency as a broad metric covering recognition, coordination, approval, and implementation delay (Gopalsamy, 2026). Judgment latency is narrower. For a defined decision episode k, let t(ready,k) be the earliest time at which the minimum pre-specified decision packet is complete, and let t(attention,k) be the first substantive attention by the authorized role or forum.

$$\mathrm{JL_k} = \mathrm{t(attention,\ k)} - \mathrm{t(ready,\ k)}$$

The readiness threshold must be defined before measurement. Otherwise a system can appear to reduce latency by declaring a request ready too early, or a team can hide waiting time by delaying the timestamp. Judgment latency ends at substantive attention, not necessarily at final approval. Time spent resolving ambiguity after attention belongs to deliberation or information assembly unless the protocol specifies otherwise.

4.2 Boundary conditions

The construct is most useful for recurring decisions with identifiable decision rights and observable readiness. It is less informative where the problem is not yet recognized, the evidence packet remains incomplete, authority is genuinely collective, or the decision is novel and requires new value formation. Low latency is not intrinsically good. A mandatory cooling-off period, independent review, or structured dissent can legitimately increase latency while improving process integrity.

Four variables should therefore accompany any latency estimate: decision class, consequence level, authority structure, and decision quality. A reduction in elapsed time is beneficial only if quality, compliance, fairness, and implementation outcomes remain within pre-specified bounds.

4.3 Role judgment rather than personal likeness

A personal clone aims to reproduce an individual's language or behavior across contexts. A role-specific judgment system should instead be restricted to authorized organizational decisions. Its representation may include recurring principles, quantitative or categorical thresholds, documented exceptions, escalation conditions, and comparable precedents. It should exclude personality imitation, private opinions, and domains outside the role's approved purpose.

This restriction does not eliminate personhood concerns. A role record may contain personal data, confidential communications, protected activity, privilege, or contributions owned or controlled by others. The phrase role-specific describes the design objective, not a legal conclusion about ownership or permissible use.

5. From decision record to governed representation

5.1 The admissible decision record

Enterprise systems may contain emails, approvals, tickets, policy documents, meeting records, and version histories that bear on prior decisions. Coverage varies substantially across organizations and periods. A defensible project begins with an admissibility manifest that lists included repositories, purposes, legal bases, retention rules, exclusions, custodians, privilege handling, and access-control semantics. Availability in a corporate system does not by itself authorize model use.

A decision episode should be represented as a linked set of evidence: the question, decision maker, decision class, relevant inputs, action or recommendation, reasons recorded at the time, later outcome if known, and superseding policy. Missing fields must remain missing. A model should not convert absence of documentation into evidence that no exception or concern existed.

5.2 Proposed judgment structure

A candidate representation can organize evidence into five components. Each component remains traceable to source records and subject to human correction.

- **Principles:** recurring reasons explicitly stated in, or cautiously inferred from, multiple decision episodes.
- **Thresholds:** quantitative or categorical boundaries associated with a change in disposition.
- **Exceptions:** departures from an apparent rule, including the contemporaneous rationale and later outcome when available.
- **Escalation conditions:** circumstances in which the role referred the matter to another person or forum.
- **Precedents:** individual prior decisions retained for direct comparison rather than absorbed into an uncited summary.

Inference should be conservative. A principle inferred from several records is not equivalent to an approved policy. Conflicts among records should be surfaced, not silently averaged. Where intuition is involved, reliable extraction is most plausible when the domain contains recurring patterns, timely feedback, and sufficient opportunity to learn, which are also conditions associated with skilled intuitive judgment (Kahneman & Klein, 2009).

5.3 Role lineage, drift, and contestability

Records from multiple incumbents may reveal durable practice, disagreement, or policy change. They should not be collapsed into a single timeless voice. A lineage-aware representation records which incumbent, time period, policy version, and evidence support each element. Current authorized policy outranks historical pattern, and the current role holder's validation does not retroactively legitimize unlawful or biased precedent.

Drift can arise from new law, policy, personnel, risk appetite, products, or operating conditions. A conforming system should support expiration dates, revalidation triggers, challenge channels, correction history, and removal where retention or legal requirements demand it. Affected workers and subject-matter experts need a meaningful way to contest both source inclusion and inferred rules.

6. Proposed architecture and design requirements

The architecture is presented as a set of requirements, not as a verified description of a deployed product. It comprises an authorized evidence layer, a role-model layer, an interaction layer, and controls that operate independently of free-form model instructions.

6.1 Authorized evidence layer

Every indexed item should retain source identity, version, custodianship, retention status, and access-control metadata. Retrieval should enforce the effective permissions of the requesting user and the narrower purpose limitation of the application. The system should deny retrieval when source permissions cannot be evaluated reliably. High-risk deployments may require customer-controlled encryption, private-cloud or on-premises processing, and separate indexes for privileged or specially protected data. These are deployment options to be validated, not assumed capabilities.

6.2 Role-model layer

Extraction produces candidate principles, thresholds, exceptions, escalation conditions, and precedent clusters. Named reviewers accept, revise, reject, or time-limit each candidate. Validation sets an operating boundary by decision class and consequence level. Performance must be tested on held-out historical cases and prospectively in shadow mode. Source citation is necessary for review but does not prove the recommendation is correct.

6.3 Interaction and authority levels

Authority should increase only after evidence. Four levels are proposed: Level 0 retrieves and summarizes approved sources; Level 1 offers a cited recommendation; Level 2 drafts a response or decision packet that requires human approval; Level 3 performs a bounded, low-consequence action under a deterministic allow-list with monitoring and rollback. Material, novel, rights-affecting, or legally reserved decisions remain with authorized humans or forums. No level makes the system an organizational officeholder or transfers accountability by itself.

Table 1. Governance controls as falsifiable design requirements

Control	Purpose	Minimum test criterion
Admissibility manifest	Define purpose, data scope, exclusions, custodians, and legal constraints.	Every source class has documented authority and an accountable owner before indexing.
Source-level authorization	Prevent retrieval through permissions the user does not hold.	Red-team tests produce zero unauthorized evidence disclosures; failures stop deployment.
Provenance	Make the basis of each output reviewable.	Every material factual assertion maps to accessible source passages and versions.
Boundary validation	Limit behavior by decision class and consequence.	Named reviewers approve a versioned boundary; out-of-bound tests decline.
Uncertainty and safe decline	Route unsupported or conflicting cases to humans.	Decline thresholds are calibrated on held-out and prospective cases.
Human confirmation	Preserve meaningful review for consequential outputs.	Reviewer can inspect evidence, disagree, record rationale, and halt action.
Deterministic action gates	Constrain tools and downstream effects outside the language model.	Unique identity, least privilege, allow-listed actions, limits, and rollback are tested.
Audit logging	Support reconstruction and oversight.	Questions, sources, outputs, approvals, overrides, and actions are time-stamped and retention-controlled.
Drift and recertification	Detect changes in policy, data, or model behavior.	Triggers, review frequency, owners, and retirement criteria are specified before use.
Incident response	Contain failures and protect continuity.	Manual fallback, kill switch, investigation procedure, and notification duties are rehearsed.

7. Illustrative scenarios

The following scenarios are synthetic. They are mechanism illustrations derived from founder-built demonstrations, not case studies, user reports, pipeline evidence, production outcomes, or claims of product readiness.

7.1 Finance evidence packet

An insurer's acquisition team asks whether a proposed structure may affect a lending covenant. In shadow mode, the system retrieves the current facility definition, the latest approved calculation, and prior documented treatment of a similar item. It identifies a judgment-dependent assumption and routes a cited draft to the controller. The controller may confirm, correct, or reject it. The system does not provide an externally usable covenant conclusion or execute a transaction. The testable mechanism is reduced search and framing time, not autonomous financial judgment.

7.2 Compliance agenda triage

A recurring review forum contains routine, exceptional, and unprecedented items. Before the meeting, a Level 1 system groups items by precedent coverage, cites comparable decisions, and highlights conflicts or missing evidence. The compliance officer reviews all outputs and chooses the agenda. The hypothesis is that routine context can be preassembled while genuine disagreement and novelty receive synchronous attention.

7.3 Succession support

A new clinical-operations director faces a staffing exception previously considered under several policy versions. A lineage-aware system retrieves the prior cases, conditions, outcomes, and superseding policy. The director retains authority and consults current clinical and legal requirements. The mechanism is memory retrieval under explicit authorization, not preservation of a predecessor's personality or automatic reuse of an old exception.

8. Quantitative model and the McKinsey estimate

8.1 Reproduction of the 2019 estimate

De Smet et al. (2019) reported about 530,000 manager-days and approximately USD 250 million in annual wages for ineffective decision-making time in a typical Fortune 500 company. Their stated assumptions were 56,400 employees, 20% managers, 220 working days per year, 37% of manager time spent making decisions, and 58% of that decision time used ineffectively. Applying those assumptions gives:

$$56{,}400 \times 0.20 \times 220 \times 0.37 \times 0.58 = 532{,}551 \text{ manager-days}$$

The result rounds to McKinsey's approximately 530,000 days. The USD 250 million figure is McKinsey's wage estimate, based in part on salary data sources identified in its note. It is not a WisdomTwin estimate, not a regulated-industry estimate, not a cost-benefit analysis, and not evidence of recoverable savings.

> Author-derived transformation. Multiplying 37% by 58% gives 21.46% of manager working time. If those annual assumptions are applied uniformly to a five-day week, 5 × 0.37 × 0.58 = 1.073 workdays per week. Thus, 'about one day a week' is a transparent arithmetic restatement of McKinsey's assumptions. It was not reported as a weekly result by McKinsey and should not be presented as a measured universal manager experience.

8.2 Non-overlapping model of potentially recovered time

A product evaluation must avoid adding meetings, messages, and interruptions when the categories overlap. For role i, define mutually exclusive weekly baseline hours: M_i for meeting attendance, A_i for asynchronous communication performed outside meetings, and Q_i for separately observed coordination or reorientation time not already counted in M_i or A_i. Let f be the eligible share of each category, g the realized gross reduction among eligible hours, and V_i the review, validation, correction, and administration burden introduced by the system.

$$R_i = M_i f^M{}_i g^M{}_i + A_i f^A{}_i g^A{}_i + Q_i f^Q{}_i g^Q{}_i - V_i$$

R_i is net weekly time recovered. The products fg are effective reduction rates, bounded from zero to one. Positive R_i does not establish productivity, financial return, or deep-work conversion. Those outcomes require separate measures. Decision waiting time should also be reported separately, preferably as the distribution of JL for pre-specified decision classes rather than converted into employee hours.

Table 2. Illustrative sensitivity analysis for one role, hours per week

Scenario	M	A	Q	Effective reductions (M / A / Q)	V	Net R
Conservative	4	5	1	20% / 15% / 10%	1.00	0.65
Central illustration	7	7	2	30% / 25% / 20%	1.25	3.00
Upper illustration	10	9	3	40% / 35% / 30%	1.50	6.55

Note. All values are synthetic inputs chosen to show model behavior. They are not published averages, predictions, pilot results, or commitments. The categories must be defined so an hour cannot enter more than one baseline bucket. Validation burden includes all time spent checking, correcting, escalating, and governing outputs.

9. Propositions

The framework yields the following falsifiable propositions for recurring, bounded decision classes.

P1. Holding information readiness and decision complexity constant, greater concentration of decision rights in a scarce role will be associated with higher median judgment latency.

P2. Evidence-grounded role-specific assistance will reduce judgment latency relative to the usual workflow when precedent coverage is high and the authorized decision boundary is stable.

P3. Net time recovery will be mediated by validation burden. Low first-pass fidelity or poor evidence presentation can eliminate or reverse gross time savings.

P4. Source provenance and permission enforcement will improve auditability and reduce unauthorized disclosure risk, but will not by themselves improve decision correctness.

P5. Record incompleteness, historical bias, and policy drift will reduce fidelity unless lineage, expiry, contestability, and revalidation controls are active.

P6. For novel or high-consequence decisions, appropriate safe decline may increase measured handling time while reducing unsupported recommendations; latency and safety should therefore be evaluated jointly.

10. Empirical evaluation agenda

10.1 Study progression

Evaluation should begin with a narrow decision taxonomy and a retrospective feasibility study. Researchers identify candidate decision classes, define readiness criteria, assess record coverage, and exclude legally or ethically unsuitable data. Historical hold-out testing can estimate evidence retrieval, recommendation agreement, and decline behavior, but it cannot establish real-world time recovery.

The next phase should be prospective shadow mode. The system produces outputs without exposing them to operational users or changing decisions. Blinded domain reviewers compare outputs with actual decisions and assess evidence completeness, material error, bias, and appropriate decline. Only if pre-specified safety thresholds are met should the study proceed to assisted mode, in which authorized users may view cited recommendations but retain all material authority. Bounded action should be considered only after a separate risk assessment and successful assisted-mode evidence.

Where enough teams or decision classes exist, a stepped-wedge or randomized rollout can improve causal inference. Otherwise, a pre-specified interrupted time-series or matched within-role design may be feasible. Baseline and intervention windows should be long enough to cover workflow cycles and policy events. Sample size should be based on baseline latency dispersion and the smallest operationally meaningful effect, not on a generic target.

10.2 Outcomes and analysis

The primary operational outcome should be median judgment latency for a pre-specified low- or moderate-risk decision class. Quality requires a separate, blinded adjudication rubric and an agreed non-inferiority margin. Latency distributions are typically skewed, so medians, quantiles, bootstrapped confidence intervals, or suitable mixed-effects models are preferable to an unqualified mean. Analyses should report overrides, missingness, learning effects, and policy changes.

Table 3. Minimum pilot measurement set

Domain	Measure	Operational definition
Speed	Judgment latency	Time from protocol-defined readiness to first substantive authorized attention, by decision class.
Quality	Material agreement	Blinded reviewers judge whether an output requires no material change, minor change, or major correction.
Safety	Unsafe recommendation rate	Outputs that would violate policy, law, decision rights, or a pre-specified harm criterion.
Calibration	Decline performance	Precision and recall of safe decline against expert labels, stratified by consequence.
Authorization	Permission integrity	Any source or output exposed beyond effective user and purpose permissions; target is zero.
Efficiency	Net time recovered	Mutually exclusive baseline reductions minus all review, correction, and governance time.
Human factors	Reliance and contestability	Acceptance of correct and incorrect advice, override behavior, and perceived ability to challenge.

Domain	Measure	Operational definition
Equity	Subgroup error analysis	Material error and decline rates across relevant groups, roles, and decision contexts where lawful to assess.
Operations	Incident and fallback performance	Detection, containment, recovery, and continuity during injected and real failures.
Downstream	Implementation and outcome quality	Completion, rework, adverse events, and outcome measures defined for the decision class.

Telemetry involving workers should be proportionate, transparent, and minimized. The measurement plan should be reviewed by privacy, employment, security, legal, records, model-risk, and worker-representation functions as applicable. A preregistered protocol, versioned model and data manifests, and an independent audit path would make later claims more credible.

11. Regulated settings and legal boundaries

Regulated organizations may have mature controls and high costs of unsupported decisions, but this does not make them automatically suitable early adopters. Their data can be more sensitive, the consequences of error higher, and the required validation more demanding. The appropriate claim is conditional: regulated settings create both a strong potential need and a high evidentiary threshold.

Privacy obligations are jurisdiction- and use-specific. Under the European Union's General Data Protection Regulation, purpose limitation, data minimization, lawfulness, transparency, security, and data-subject rights can be relevant to employee and customer records (European Parliament & Council, 2016). Canada's PIPEDA applies to commercial activities and to employee personal information in federal works, undertakings, and businesses; provincial coverage differs (Office of the Privacy Commissioner of Canada, 2025). HIPAA applies to protected health information held or transmitted by covered entities and business associates, not to employee communications generally (U.S. Department of Health and Human Services, 2025).

The EU AI Act classifies specified employment and worker-management uses as high risk and imposes duties that depend on role, use case, and implementation timing (European Parliament & Council, 2024). Banking guidance also requires careful scoping. The 2026 U.S. interagency model-risk guidance states that generative and agentic AI are outside that document's scope while indicating that broader risk-governance practices should guide tools not covered (Board of Governors of the Federal Reserve System et al., 2026). OSFI's Guideline E-23 takes effect for Canadian federally regulated financial institutions on May 1, 2027, while OSFI's July 2026 technology-risk bulletin directly addresses generative and agentic AI (OSFI, 2025, 2026).

A deployment may also implicate employment law, labor agreements, works councils, professional duties, records retention, litigation hold, confidentiality, privilege, intellectual property, discrimination law, sector rules, and procurement commitments. Human review does not automatically resolve those issues. Responsibility and authority must be allocated by applicable law, organizational policy, contract, and professional standard.

12. Risks, limitations, and discussion

The central technical risk is false fidelity: a fluent response may look like the role's settled view while resting on incomplete, conflicting, obsolete, or biased evidence. Citation can make an error inspectable but not correct. Confidence labels may be poorly calibrated, and anthropomorphic framing may encourage overreliance. A narrow interface, consequence-based authority, cognitive forcing, and prominent uncertainty are therefore preferable to unrestricted persona simulation.

The central organizational risk is institutionalizing the past. Historical decisions can contain discrimination, workarounds, path dependence, and strategic choices that should not be repeated. A role also changes when a

new incumbent is hired to change policy. Governance must allow explicit discontinuity rather than treating consistency as the objective.

The central measurement risk is mistaking displaced activity for recovered value. Time saved in one meeting may reappear as review, exception handling, new communication, or additional work. Even genuine net time recovery may not become concentrated work, wellbeing, or improved outcomes. These downstream effects require direct measurement.

This paper has further limitations. It is authored by a founder with a financial interest in the concept. The literature review is purposive and may omit relevant traditions in organizational theory, law, information systems, and human-computer interaction. The framework has not been tested in production. The examples are synthetic. The proposed controls can conflict, for example when audit retention and deletion duties point in different directions. Finally, the name AI Judgment Twin may overstate representational fidelity unless every use is accompanied by the narrower definition provided here.

13. Conclusion

Judgment latency isolates a measurable organizational interval: the wait between a decision being ready and the required role attending to it. The construct is useful only when readiness, decision class, authority, consequence, and quality are specified. It should not be reduced to a slogan about faster decisions.

A role-specific, evidence-grounded AI system could, in principle, reduce that interval for recurring bounded decisions by assembling precedent, exposing conflict, drafting recommendations, and escalating novelty. It could also amplify bias, disclose protected information, create false authority, or add review burden. The difference depends on admissible data, precise permissions, calibrated decline, meaningful human control, independent testing, and continuous governance.

The next credible step is therefore not a claim of hours recovered or enterprise value created. It is a transparent, staged evaluation that begins in shadow mode, treats safety and decision quality as constraints, and publishes results whether or not the propositions are supported.

Declarations

Conflict of interest. The author is Co-Founder and Chief Executive Officer of WisdomTwin, Inc., is involved in developing the proposed architecture, and has a financial interest in its adoption.

Product and evidence status. WisdomTwin is pre-revenue, has USD $0 product revenue, no production users, and no paying customers. The company has five founder-built synthetic demonstrations. No demonstration, pipeline activity, or client outcome is presented as product traction or empirical validation in this paper.

Funding. No external research funding was reported for the preparation of this working paper.

Data availability. No new empirical dataset was collected or analyzed. The arithmetic in Section 8 is reproduced from stated source assumptions and shown in full.

Ethics. The paper reports no research involving human participants, personal records, or live organizational deployments. All scenarios are synthetic.

AI-assistance disclosure. Generative AI tools assisted with source discovery, fact-checking, drafting, editing, and document production. Citations and quantitative claims retained in this version were checked against publisher pages, versions of record, or official institutional sources on September 3, 2026. The named author remains responsible for the argument, disclosures, and final submitted text.

Peer-review status. This working paper has not been peer reviewed.

Version history

Version	Date	Status and material changes
1.0	September 3, 2026	Initial pre-publication draft; references flagged as unverified.
2.0	September 3, 2026	Fact-checked conceptual revision; source caveats corrected; Bodnar item removed; McKinsey arithmetic reproduced and relabeled; architecture claims converted to design requirements; propositions and evaluation protocol added; legal scope narrowed.

References

Allen, J. A., & Lehmann-Willenbrock, N. (2023). The key features of workplace meetings: Conceptualizing the why, how, and what of meetings at work. Organizational Psychology Review, 13(4), 355–378. https://doi.org/10.1177/20413866221129231

Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, & Office of the Comptroller of the Currency. (2026, April 17). Revised guidance on model risk management (SR 26-2). https://www.federalreserve.gov/supervisionreg/srletters/SR2602.htm

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), Article 188, 1–21. https://doi.org/10.1145/3449287

De Smet, A., Jost, G., & Weiss, L. (2019, May 1). Three keys to faster, better decisions. McKinsey Quarterly. https://www.mckinsey.com/capabilities/people-and-organization/our-insights/three-keys-to-faster-better-decisions

European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union, L 119, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Gopalsamy, L. P. (2026). Decision latency as a first-class performance metric in AI-native engineering organizations. Journal of Information Systems Engineering and Management, 11(2s), 445–453. https://doi.org/10.52783/jisem.v11i2s.14423

Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In F.-J. Kahlen, S. Flumerfelt, & A. Alves (Eds.), Transdisciplinary perspectives on complex systems (pp. 85–113). Springer. https://doi.org/10.1007/978-3-319-38756-7_4

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the digital twin: A systematic literature review. CIRP Journal of Manufacturing Science and Technology, 29, 36–52. https://doi.org/10.1016/j.cirpj.2020.02.002

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. American Psychologist, 64(6), 515–526. https://doi.org/10.1037/a0016755

Leroy, S. (2009). Why is it so hard to do my work? The challenge of attention residue when switching between work tasks. Organizational Behavior and Human Decision Processes, 109(2), 168–181. https://doi.org/10.1016/j.obhdp.2009.04.002

Lin, Y., Chen, L., Ali, A., Nugent, C., Cleland, I., Li, R., Ding, J., & Ning, H. (2024). Human digital twin: A survey. Journal of Cloud Computing, 13, Article 131. https://doi.org/10.1186/s13677-024-00691-z

Mankins, M., Brahm, C., & Caimi, G. (2014). Your scarcest resource. Harvard Business Review, 92(5), 74–80, 133. https://pubmed.ncbi.nlm.nih.gov/24956871/

Mark, G., Gonzalez, V. M., & Harris, J. (2005). No task left behind? Examining the nature of fragmented work. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 321–330). Association for Computing Machinery. https://doi.org/10.1145/1054972.1055017

Mark, G., Gudith, D., & Klocke, U. (2008). The cost of interrupted work: More speed and stress. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 107–110). Association for Computing Machinery. https://doi.org/10.1145/1357054.1357072

Microsoft. (2023, May 9). Will AI fix work? Work Trend Index Annual Report. https://www.microsoft.com/en-us/worklab/work-trend-index/will-ai-fix-work

Microsoft. (2025, June 17). Breaking down the infinite workday. Work Trend Index Special Report. https://www.microsoft.com/en-us/worklab/work-trend-index/breaking-down-infinite-workday

National Institute of Standards and Technology. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). https://doi.org/10.6028/NIST.AI.600-1

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. Organization Science, 5(1), 14–37. https://doi.org/10.1287/orsc.5.1.14

Office of the Privacy Commissioner of Canada. (2025, May 29). Privacy in the workplace. https://www.priv.gc.ca/en/privacy-topics/employers-and-employees/02_05_d_17/

Office of the Superintendent of Financial Institutions. (2025, September 11). Guideline E-23: Model risk management (2027). https://www.osfi-bsif.gc.ca/en/guidance/guidance-library/guideline-e-23-model-risk-management-2027

Office of the Superintendent of Financial Institutions. (2026, July). Generative and agentic artificial intelligence: Implications for technology, cyber security, and operational resilience. https://www.osfi-bsif.gc.ca/en/risks/technology-cyber-risk-management/technology-risk-bulletin/generative-agentic-artificial-intelligence-implications-technology-cyber-security-operational

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery. https://doi.org/10.1145/3586183.3606763

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253. https://doi.org/10.1518/001872097778543886

Perlow, L. A. (1999). The time famine: Toward a sociology of work time. Administrative Science Quarterly, 44(1), 57–81. https://doi.org/10.2307/2667031

Puranik, H., Koopman, J., & Vough, H. C. (2020). Pardon the interruption: An integrative review and future research agenda for research on work interruptions. Journal of Management, 46(6), 806–842. https://doi.org/10.1177/0149206319887428

Rogelberg, S. G., Allen, J. A., Shanock, L., Scott, C., & Shuffler, M. (2010). Employee satisfaction with meetings: A contemporary facet of job satisfaction. Human Resource Management, 49(2), 149–172. https://doi.org/10.1002/hrm.20339

Rogers, P., & Blenko, M. (2006). Who has the D? How clear decision roles enhance organizational performance. Harvard Business Review, 84(1), 52–61, 131. https://hbr.org/2006/01/who-has-the-d-how-clear-decision-roles-enhance-organizational-performance

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

U.S. Department of Health and Human Services. (2025, March 14). Summary of the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

Walsh, J. P., & Ungson, G. R. (1991). Organizational memory. Academy of Management Review, 16(1), 57–91. https://doi.org/10.5465/amr.1991.4278992

Appendix A. Glossary

AI Judgment Twin. A proposed role-specific, evidence-grounded decision-support representation limited by purpose, permissions, validated decision classes, and authority level. It is not a personal replica or officeholder.

Decision episode. A bounded instance containing a question or trigger, evidence state, authorized judgment, action or recommendation, and outcome where available.

Decision latency. The broader elapsed time across recognition, information assembly, coordination, approval, and implementation.

Judgment latency. The elapsed time from protocol-defined decision readiness to first substantive attention by the authorized role or forum.

Safe decline. A refusal to recommend or act because the request is outside the validated boundary, evidence is insufficient or conflicting, or consequence requires human attention.

Shadow mode. Prospective evaluation in which system outputs are recorded for comparison but do not influence operational users or decisions.

Validation burden. All human and organizational time required to inspect, correct, approve, escalate, govern, and audit system outputs.

Appendix B. Minimum decision-readiness record

A study protocol should specify the following fields before latency measurement begins.

- Stable decision-class identifier and consequence tier.
- Named role or forum holding decision authority.
- Minimum required evidence packet and the rule for declaring it complete.
- Ready timestamp, attention timestamp, disposition timestamp, and execution timestamp.
- Outcome-quality rubric, reviewers, blinding method, and adjudication path.
- Version identifiers for policy, source corpus, retrieval system, model, prompts, and action gates.
- Overrides, declines, missing evidence, incidents, and downstream rework.