International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: https://ijwos.com OPEN ACCESS E-ISSN : 3049-2424
Volume 3 Issue 7 , July 2026 DOI: https://doi.org/10.71366/ijwos03072654084

The Governance Half-Life of Agentic Artificial Intelligence Controls: A Dynamic Assurance Framework for Control Decay, Residual Risk, and Revalidation

Kwan Hong TAN[1]
[1] *Singapore University of Social Sciences, 463 Clementi Road, Singapore 599494.*

Article Info

Article History:

Published: 13 July 2026

Publication Issue:
Volume 3, Issue 7
July-2026

Page Number:
217-229

Corresponding Author:
Kwan Hong TAN

Abstract:

Agentic artificial intelligence can plan, use tools, communicate with other systems, and execute actions across changing environments. Its governance problem is therefore temporal: a control that was adequate at deployment may become materially weaker after model updates, data drift, new tool permissions, adversarial adaptation, staff turnover, or regulatory change. This paper develops the governance half-life construct to measure the rate at which an artificial intelligence control loses effective coverage after validation. Using an integrative multidisciplinary review and design-science theory building, the study combines artificial intelligence risk management, software reliability, cybersecurity, human factors, organizational learning, and regulatory governance. The proposed model represents control effectiveness as a decaying function, derives a governance half-life and a maximum risk-tolerable revalidation interval, and combines scheduled, event-triggered, and continuous assurance. A worked analytical demonstration shows why low-risk knowledge agents may tolerate periodic review while transaction, eligibility, and safety-relevant agents require much shorter assurance cycles. The framework contributes a control-decay ledger, portfolio-level aggregation, testable propositions, and an implementation maturity model. It shifts governance from one-time approval and annual audit toward risk-sensitive evidence renewal. The framework is intended for empirical calibration rather than treated as a universal numerical standard, but it offers organizations a practical way to align agent autonomy, control durability, residual risk, and revalidation cadence.

Keywords: agentic artificial intelligence, continuous assurance, control decay, governance half-life, residual risk, risk-based revalidation

1. Introduction

Artificial intelligence governance has historically been organized around relatively stable artifacts: a model is documented, tested, approved, deployed, and periodically audited. Agentic systems strain this logic. They can decompose goals, select tools, retain state, call external services, coordinate with other agents, and act with limited supervision. Research on reasoning-and-action agents, reflective agents, multi-agent orchestration, and interactive benchmarks demonstrates that capability increasingly emerges from a system of models, prompts, memories, tools, permissions, and environmental feedback rather than from a single static model [12]–[18]. The governance object is therefore not merely a trained model. It is a changing sociotechnical action system whose risk can move even when the underlying model weights remain unchanged.

Contemporary standards already require lifecycle risk management, monitoring, documentation, accountability, and post-deployment review. The National Institute of Standards and Technology structures artificial intelligence risk management around Govern, Map, Measure, and Manage [1], [2]; ISO/IEC 42001 embeds artificial intelligence within a management-system cycle [3]; ISO/IEC 23894 provides risk-management guidance [4]; and the European Union Artificial Intelligence Act requires risk management, logging, human oversight, accuracy, cybersecurity, and post-

market monitoring for covered systems [5]. Singapore's 2026 Model Artificial Intelligence Governance Framework for Agentic Artificial Intelligence moves further by addressing accountability, meaningful human control, technical safeguards, and value-chain responsibilities for agents [8]. These instruments establish what organizations should govern, but they do not yet provide a general analytical measure for how quickly a previously validated control becomes stale.

This paper addresses that gap through the concept of governance half-life: the elapsed time over which the effective coverage of a control falls to one-half of its validated level under specified operating conditions. The construct does not imply that controls decay mechanically or at a constant empirical rate. Rather, it offers a disciplined approximation that makes three neglected questions explicit: How durable is a control after approval? Which changes accelerate its loss of coverage? When must evidence be renewed before residual risk exceeds tolerance? The framework is designed for agentic systems but applies to any adaptive, highly coupled artificial intelligence deployment in which technical and organizational conditions change faster than conventional audit cycles.

The study asks three research questions. First, what multidisciplinary mechanisms cause control effectiveness to decline after deployment? Second, how can control decay, residual risk, and revalidation cadence be represented in a parsimonious analytical model? Third, what organizational architecture can operationalize the model without converting governance into indiscriminate compliance friction? The paper answers these questions through an integrative review and design-science method, then provides a dynamic assurance model, illustrative calculations, testable propositions, and an implementation pathway.

The central argument is that trustworthy autonomy depends not only on the strength of controls at time zero, but also on their durability and renewal rate. A strong control with a short half-life may be less protective than a moderately strong control that is continuously monitored, readily recalibrated, and supported by competent human reviewers. Conversely, permanent monitoring does not eliminate the need for design constraints, because observability can reveal failure without preventing irreversible action. Governance quality is thus a function of initial control coverage, decay rate, inherent exposure, detectability, reversibility, and restoration capability.

2. Literature Review and Theoretical Foundations

2.1 Agentic systems as evolving action infrastructures.

Foundation models create general capabilities that can be recombined across tasks, but agentic systems add planning, memory, tool use, and feedback loops [19]. ReAct interleaves reasoning and action [12], Reflexion uses feedback to improve subsequent attempts [13], AutoGen coordinates multiple conversational agents [17], and agent benchmarks reveal substantial performance variation across environments and tasks [14]–[16]. These architectures produce value because they adapt, yet the same adaptability expands the control surface. A prompt change can alter planning behavior; an application programming interface update can change available actions; retrieval data can shift; a tool can inherit broader permissions; or one agent can amplify the error of another. The system's operational identity is therefore versioned and relational, not fixed.

Risk taxonomies for large language models identify misinformation, harmful outputs, discrimination, privacy leakage, malicious use, and human-computer interaction failures [20]. In agentic settings, these hazards become executable. Prompt injection can redirect tool use; excessive agency can produce unauthorized action; insecure output handling can propagate commands; supply-chain changes can introduce vulnerabilities; and unbounded consumption can create financial or environmental externalities [9]–[11]. Static testing remains necessary, but it only establishes performance and control coverage under a bounded configuration and time window.

2.2 From model drift to control drift.

Machine-learning engineering has long recognized that production systems accumulate hidden technical debt through entangled dependencies, unstable data, configuration complexity, feedback loops, and changes in the surrounding

system [21]. Production-readiness rubrics accordingly emphasize monitoring, data tests, model tests, infrastructure tests, and reproducibility [22], while software-engineering studies show that machine-learning systems differ from conventional software because data and models are tightly coupled and error behavior can be non-monotonic [23]. Concept-drift research further demonstrates that relationships between inputs and outcomes can change over time [24]. Governance half-life extends these insights from model performance to control performance. A fairness threshold may lose meaning after population shift; a permission policy may become porous after a new integration; an escalation rule may fail after work redesign; and a documentation artifact may no longer describe the deployed system.

Documentation instruments such as model cards and datasheets improve transparency by recording intended uses, limitations, performance characteristics, and dataset properties [25], [26]. Internal algorithmic auditing adds structured evaluation and accountability throughout the development lifecycle [27]. Yet documentation itself can decay when system changes are not reflected in the record. The abstraction trap identified in sociotechnical fairness research is especially relevant: controls may appear complete at the technical layer while omitting institutional context, affected stakeholders, or the conditions under which the system is actually used [28]. Accountable algorithms consequently require procedural safeguards, reviewability, and mechanisms for contestation rather than technical explanation alone [29].

2.3 Human control and competence decay.

Human oversight is often treated as a stable safeguard, but human-factors research suggests otherwise. Bainbridge's ironies of automation show that automation can leave humans responsible for rare, difficult interventions while reducing the practice needed to perform them [30]. Automation may be used, misused, disused, or abused [31], and appropriate levels of automation depend on the type of function and the allocation of responsibility between human and machine [32]. Trust must be calibrated to system capability rather than maximized [33]. As systems become more autonomous, situation awareness and meaningful intervention can deteriorate, particularly when operators monitor long periods of normal performance and face abrupt exceptions [34].

Automation bias can cause users to follow erroneous recommendations or omit independent checks [35]. Interventions that require users to reason rather than merely accept advice can improve appropriate reliance, although they may also impose cognitive cost [36]. Algorithm aversion and algorithm appreciation further show that reliance is context-sensitive rather than uniformly excessive or deficient [37], [38]. These findings imply a human-control decay mechanism: review skill, domain knowledge, attention, role clarity, and confidence can weaken after deployment. A nominal human-in-the-loop control may thus retain its organizational label while losing practical effectiveness.

2.4 Organizational memory, technical debt, and governance renewal.

Technical debt research frames short-term design expedients as obligations that impose future cost [39], [40]. Organizational-learning research likewise treats knowledge acquisition, distribution, interpretation, and memory as processes that can strengthen or degrade performance [41], [42]. Agentic governance combines both dynamics. Organizations may approve an agent faster than they build durable ownership, incident response, model inventories, reviewer competence, or change-management routines. The result is not simply a missing control but a control whose effectiveness depends on tacit knowledge held by a small number of people. Turnover, vendor substitution, outsourcing, or restructuring can sharply reduce coverage even when formal policies remain unchanged.

2.5 Governance standards and the missing temporal metric.

Current frameworks converge on risk-based governance, documentation, monitoring, human oversight, accountability, and continuous improvement [1]–[8]. The Singapore agentic framework explicitly recognizes that agents operate through value chains involving model developers, platform providers, tooling providers, system providers, deployers, and end users [8]. Public-sector research similarly finds that agent oversight requires continuous visibility, cross-departmental coordination, security, evaluation, and auditing rather than siloed episodic review [49]. However,

organizations still need a way to translate these principles into assurance timing. Annual audits may be excessive for a stable low-risk agent yet dangerously slow for a rapidly changing high-impact agent. The literature therefore supports a temporal construct that is risk-sensitive, empirically calibratable, and compatible with existing governance systems.

2.6 Sustainability and externality drift.

Control decay is not limited to accuracy, security, or compliance. Training and deploying large artificial intelligence systems can consume substantial energy, and inference-heavy general-purpose systems may be materially more resource-intensive than task-specific alternatives [43]–[45]. A deployment approved under one workload can exceed its cost or carbon budget after usage growth, agent loops, retries, or multi-agent orchestration. Environmental and financial controls therefore require the same dynamic treatment as safety controls. A system can remain technically functional while becoming economically or environmentally misaligned.

Prior work by Tan argues that enterprise artificial intelligence value depends on the interaction between productivity and governance rather than capability alone [46], that agentic organizations transform human roles toward supervision and accountability [47], and that bounded autonomy requires evidence-generating, reversible, and accountable control architectures [48]. The present study extends these contributions by focusing on the temporal durability of controls. Its novelty lies not in asserting that controls must be monitored, but in formalizing control decay, deriving a risk-tolerable validation interval, and connecting those measures to a practical assurance operating model.

Table 1. Static-governance failure modes and the governance half-life response

Failure mode	Why static approval becomes stale	Dynamic assurance response
Model and prompt change	Behavior changes without a new business approval	Version binding, regression tests, event-triggered revalidation
Data and context drift	Performance, fairness, or relevance shifts after deployment	Drift indicators, subgroup checks, tolerance-based review
Tool and permission expansion	The agent can act beyond the originally assessed boundary	Least privilege, capability inventory, authorization tests
Threat adaptation	Attackers discover new prompt, tool, or supply-chain paths	Red-team refresh, security telemetry, incident-triggered reset
Human competence loss	Reviewers retain responsibility but lose intervention skill	Drills, rotation, minimum practice, escalation testing
Organizational and legal change	Ownership, workflows, or duties change while policy text remains	Owner confirmation, legal trigger, evidence refresh
Usage and externality growth	Cost, energy, and social impact exceed approved assumptions	Budget and impact thresholds, workload and carbon monitoring

3. Method

This study uses an integrative conceptual review and design-science theory-building approach. An integrative review is appropriate because the research problem crosses artificial intelligence engineering, cybersecurity, human factors, organizational behavior, law, risk management, and sustainability. The evidence base included foundational research on automation and technical debt, contemporary work on foundation-model agents and production machine learning, international standards, regulatory instruments, and recent agentic-governance guidance. Sources were selected for conceptual relevance, methodological influence, or direct operational authority rather than for statistical aggregation. The method does not claim systematic-review exhaustiveness.

The design process followed four stages. First, the problem was reframed from static control adequacy to time-varying control effectiveness. Second, recurring decay mechanisms were synthesized into technical, adversarial, operational, human, institutional, and externality dimensions. Third, these mechanisms were expressed as a compact analytical model connecting control coverage, decay, inherent risk, and revalidation. Fourth, the model was stress-tested through four hypothetical agent scenarios representing increasing action scope and consequence. The scenarios are analytical demonstrations, not empirical observations, and their parameters are deliberately transparent so that organizations can replace them with calibrated values.

Construct validity was pursued through alignment with established concepts. Initial control coverage corresponds to the proportion of relevant risk pathways effectively addressed at validation. Decay rate generalizes model drift, technical debt, threat evolution, human skill degradation, and organizational forgetting. Inherent risk reflects exposure before controls, while residual risk reflects the uncovered portion after controls. Revalidation includes testing, approval, evidence refresh, role confirmation, and restoration actions. The model is intentionally parsimonious: it supports comparison and scheduling but does not replace domain-specific safety cases, legal analysis, statistical validation, or cybersecurity testing.

4. Results and Discussion

4.1 The governance half-life model.

For control i, let $C_i(t)$ denote effective control coverage at time t after validation, where $0 \le C_i(t) \le 1$. Under a first-order approximation, coverage decays exponentially from validated coverage $C_{i,0}$ at rate $\lambda_i > 0$:

$$C_i(t) = C_{i,0} \exp(-\lambda_i t) \quad (1)$$

The exponential form is not presented as a universal law. It is a practical baseline because it preserves positivity, permits comparison across controls, and can be updated as evidence accumulates. A stepwise, piecewise-linear, or hazard-based model may be superior for particular systems. The governance half-life h_i is the time at which effective coverage falls to one-half of its validated level:

$$h_i = \ln(2) / \lambda_i \quad (2)$$

A short half-life signals rapid staleness, not necessarily weak initial design. For example, a well-designed authorization control can have high initial coverage but a short half-life if tool permissions, staff roles, and application interfaces change frequently. Conversely, a stable low-risk retrieval agent may have a longer half-life even with more modest controls.

The decay rate is determined by conditions rather than assigned by technology category alone. A general specification is:

$$\lambda_i = \lambda_{\text{ref}} \exp(\beta_1 V_M + \beta_2 V_D + \beta_3 V_T + \beta_4 A + \beta_5 K + \beta_6 O - \beta_7 M - \beta_8 R) \quad (3)$$

Here V_M is model-version volatility, V_D data or context drift, V_T threat volatility, A action and permission scope, K dependency coupling, O organizational turnover, M monitoring quality, and R recovery readiness. The beta coefficients are empirically estimable sensitivities. The exponential link ensures a positive decay rate and allows interacting drivers to accelerate decay. Monitoring and recovery reduce effective decay because they detect divergence sooner, preserve evidence, and shorten the time during which ineffective controls remain uncorrected. They do not, however, make the underlying system intrinsically safe.

Table 2. Governance half-life variables and candidate operational indicators

Symbol	Construct	Illustrative indicators
$C_{i,0}$	Validated control coverage	Test coverage, policy-path coverage, reviewer readiness, rollback proof
λ_i	Control-decay rate	Release frequency, drift, threat change, permission change, staff turnover
B_i	Inherent risk	Exposure, impact, affected population, privilege, irreversibility
θ_i	Risk tolerance	Board-approved limit, regulatory threshold, service-level boundary
M	Monitoring quality	Telemetry completeness, detection delay, alert precision, evidence retention
R	Recovery readiness	Rollback time, containment success, fallback availability, incident rehearsal

4.2 Residual risk and the maximum revalidation interval.

Let B_i represent inherent risk for the controlled activity, combining exposure and consequence before control. Residual risk is:

$$RR_i(t) = B_i\,[1 - C_i(t)] \tag{4}$$

Given risk tolerance θ_i, the maximum scheduled interval T_i^* before revalidation is the time at which residual risk reaches tolerance:

$$T_i^* = \frac{1}{\lambda_i} \ln\left[\frac{C_{i,0}}{1-(\frac{\theta_i}{B_i})}\right] \tag{5}$$

This expression is defined when $\theta_i < B_i$ and initial coverage is sufficient to place residual risk below tolerance. If initial residual risk already exceeds tolerance, the system should not be approved on the assumed configuration. T_i^* is a

ceiling rather than a guarantee: event-triggered revalidation can occur earlier, and continuous controls may be required for high-impact actions.

4.3 Control restoration and evidence renewal.

Revalidation should not be reduced to rerunning a benchmark. At intervention time τ_j, the organization may restore coverage by updating policy, reducing permissions, retraining reviewers, changing prompts or models, strengthening tests, correcting data, or proving rollback capability. The post-intervention coverage can be represented as a bounded jump:

$$C_i(\tau_j^+) = \min \{1, C_i(\tau_j^-) + \Delta C_{ij}\} \quad (6)$$

A control-decay ledger should record $C_{i,0}$, λ_i, evidence date, system version, owner, dependencies, risk tolerance, scheduled ceiling, event triggers, and restoration history. This creates a durable link between governance decisions and the configuration to which they applied. It also supports regulatory logging, internal audit, incident investigation, and board reporting without requiring every stakeholder to interpret raw technical telemetry.

4.4 Dynamic Assurance Scheduling.

The proposed operating model combines three clocks. The first is a scheduled clock derived from T_i^*. The second is an event clock activated by material changes: model or prompt update, new tool or data source, permission expansion, vendor change, workflow redesign, reviewer turnover, incident, regulatory change, significant drift, or breach of cost and environmental thresholds. The third is a continuous clock for telemetry such as action logs, exception rates, policy violations, latency, spend, energy, fairness indicators, and rollback success. Scheduled assurance prevents indefinite staleness; event assurance responds to discontinuities; and continuous assurance detects deterioration between formal reviews.

Assurance intensity should be proportionate to action consequence and reversibility. An agent that drafts internal text may be monitored primarily for confidentiality and quality. A procurement agent that recommends suppliers requires conflict, authorization, and market-integrity controls. A customer-facing transaction agent needs real-time limits, identity controls, rollback, and fraud monitoring. An eligibility or safety-relevant agent should operate under strict action boundaries, independent review, contestability, and evidence retention. The same foundation model can therefore sit within different governance half-lives depending on deployment context.

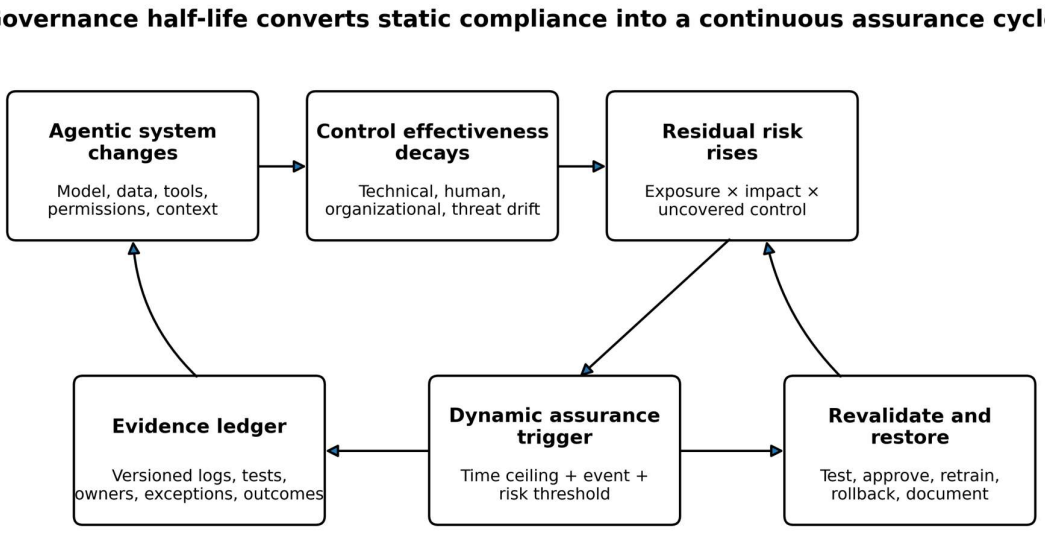

Figure 1. Dynamic assurance cycle governed by control decay and evidence renewal

4.5 Illustrative analytical demonstration.

Table 3 applies the equations to four hypothetical agent types. The values are not empirical estimates. They demonstrate how identical high initial coverage can produce different assurance schedules when inherent risk and decay differ. The low-risk knowledge-retrieval agent has a half-life of approximately 19.8 weeks and reaches its illustrative tolerance after 8.6 weeks. The procurement agent reaches tolerance in about 1.4 weeks. The transaction agent crosses tolerance in under one week, while the regulated eligibility agent requires a ceiling measured in days. These differences arise before any assumption of catastrophic model failure; they follow from ordinary control decay interacting with higher consequence.

Table 3. Illustrative governance half-lives and risk-tolerable revalidation ceilings

Agent scenario	B_i	$C_{i,0}$	λ_i / week	Half-life h_i	Maximum T_i^*
Internal knowledge retrieval	0.25	0.92	0.035	19.8 weeks	8.64 weeks
Procurement recommendation	0.55	0.90	0.070	9.9 weeks	1.36 weeks
Customer transaction execution	0.75	0.94	0.120	5.8 weeks	0.68 weeks
Regulated eligibility decision	0.90	0.97	0.160	4.3 weeks	0.39 weeks

Figure 2 shows the sensitivity of residual risk to the decay rate. With the same initial coverage and inherent risk, high-decay controls cross a fixed tolerance much earlier. This highlights why governance maturity cannot be measured only by the existence of policies or the number of controls. The rate at which controls are invalidated by change is equally important. Organizations with complex vendor ecosystems, frequent releases, broad permissions, and high staff turnover may need more assurance capacity even when their written frameworks appear comprehensive.

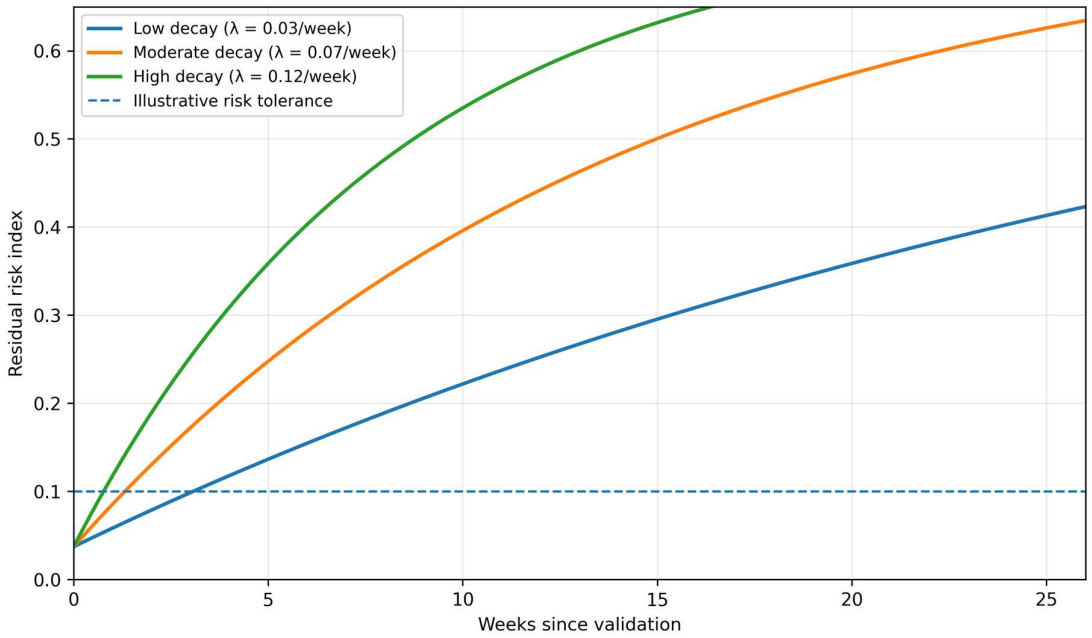

Figure 2. Sensitivity of residual risk to alternative control-decay rates

4.6 Portfolio aggregation.

Enterprises govern portfolios of agents, not isolated controls. A simple portfolio indicator can be formed as the exposure-weighted harmonic mean of control half-lives:

$$h_{\text{portfolio}} = [\sum_i (\frac{w_i}{h_i})]^{-1}, \qquad \sum_i w_i = 1 \tag{7}$$

The harmonic mean is appropriate because short half-lives exert greater influence than long ones. However, portfolio aggregation must not conceal critical controls. A high-impact agent with a very short half-life should retain a separate escalation rule even when its exposure weight is small. Portfolio dashboards should therefore combine aggregate indicators with red-flag thresholds for irreversibility, legal significance, vulnerable populations, security privilege, and systemic coupling.

4.7 Theoretical propositions.

Proposition 1: Greater action scope, permission breadth, and irreversibility will be associated with shorter governance half-lives, controlling for initial control quality.

Proposition 2: Dependency and threat volatility will accelerate control decay more strongly in multi-agent systems than in single-agent systems because failures can propagate through coordination pathways.

Proposition 3: Observability, evidence retention, and tested recovery will lengthen effective half-life by reducing undetected divergence and restoration delay, but they will not substitute for preventive constraints.

Proposition 4: Human-oversight effectiveness will exhibit nonlinear decline when intervention frequency falls below the level needed to preserve reviewer competence and situation awareness.

Proposition 5: Dynamic assurance schedules will produce lower cumulative residual risk than fixed annual review at comparable governance effort in volatile deployments.

Proposition 6: Portfolio risk will be disproportionately determined by high-exposure controls with the shortest half-lives rather than by average control maturity.

4.8 Managerial and policy implications.

The framework reframes governance capacity as a renewable operational resource. Boards should ask not only whether an agent was approved, but when the evidence will expire and what events invalidate it. Chief risk officers can use the control-decay ledger to allocate assurance resources to fast-decaying, high-consequence controls. Engineering teams can integrate revalidation triggers into release management. Human-resource leaders can protect reviewer competence through rotation, drills, and minimum intervention practice. Procurement teams can require vendors to disclose model, tool, and policy changes that reset assurance assumptions. Sustainability teams can treat growth in inference loops and resource consumption as governance events rather than merely cost variances.

For regulators and standard setters, governance half-life offers a way to operationalize proportionality. Prescriptive calendar frequencies are often either too slow for volatile systems or unnecessarily burdensome for stable low-risk uses. A risk-tolerable interval tied to inherent risk, control coverage, and decay drivers can support more adaptive supervision. Regulators need not adopt the formula directly; they can require organizations to justify review cadence, identify invalidating events, retain evidence, and demonstrate that controls remain effective between approvals.

4.9 Implementation maturity model.

Organizations can implement the framework incrementally. At Level 1, controls are documented at deployment but reviewed mainly after incidents. At Level 2, systems have owners, inventories, scheduled reviews, and basic change notifications. At Level 3, controls have estimated half-lives, event triggers, versioned evidence, and tested rollback.

At Level 4, decay rates are calibrated from incidents, drift, changes, and assurance outcomes; schedules are risk-optimized; and portfolio exposure is governed dynamically. Maturity should be evidenced by operating data rather than self-description.

Table 4. Governance half-life implementation maturity model

Level	Operating characteristics	Evidence of maturity
1. Reactive	Approval at deployment; review after incidents	Basic policy, incident record, named system owner
2. Scheduled	Inventory, periodic review, change notification	Review calendar, version register, role assignments
3. Dynamic	Estimated half-life, event triggers, continuous telemetry	Control-decay ledger, trigger logs, rollback tests
4. Calibrated	Empirical decay estimation and portfolio optimization	Longitudinal data, validated coefficients, risk-adjusted capacity

4.10 Limitations and research agenda.

The model simplifies complex control behavior into a scalar coverage measure and an initial exponential decay assumption. Real controls can fail abruptly, improve through learning, interact with one another, or decay differently across stakeholder groups. The illustrative parameters are not validated. Future research should estimate decay rates from change logs, incidents, audit findings, override patterns, reviewer performance, and policy exceptions. Longitudinal field studies could compare fixed and dynamic assurance schedules. Experiments could test how intervention frequency affects reviewer competence. Sector-specific studies should examine whether financial, healthcare, education, public-sector, and industrial agents exhibit distinct decay profiles. Research should also study equity-sensitive half-lives, because a control may remain effective on average while deteriorating faster for underrepresented groups.

5. Conclusion

Agentic artificial intelligence converts governance from a point-in-time approval problem into a problem of control durability. Models, data, tools, permissions, threats, organizations, and human capabilities all change after deployment. A control can therefore remain formally present while losing effective coverage. The governance half-life framework makes this temporal risk visible by linking validated control coverage, decay rate, inherent risk, and revalidation cadence.

The paper contributes an analytical model, a maximum risk-tolerable review interval, a three-clock assurance architecture, a control-decay ledger, portfolio aggregation, testable propositions, and a maturity pathway. Its practical message is straightforward: governance evidence should have an expiry logic. High-risk agents require continuous and event-driven assurance, while stable low-risk agents can be governed proportionately. The numerical form is a starting hypothesis for empirical calibration, not a substitute for domain expertise or legal duties. Its value lies in changing the unit of governance from controls that exist to controls that remain effective.

Acknowledgments

The author acknowledges the interdisciplinary research communities in artificial intelligence governance, software engineering, human factors, cybersecurity, and organizational learning whose work informed the conceptual synthesis.

Author Contributions

Kwan Hong TAN: conceptualization, methodology, formal analysis, investigation, visualization, writing the original draft, draft review and editing.

Funding Details

This research received no external funding.

Disclosure Statement

The author declares no financial or non-financial conflict of interest.

Data Availability

No empirical dataset was generated or analyzed. The numerical scenarios are hypothetical analytical demonstrations, and all equations and parameters required to reproduce them are reported in the manuscript.

Ethics Statement

This conceptual study did not involve human participants, animals, personal data, or confidential organizational records; therefore, ethics committee approval and informed consent were not required.

References

[1] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, Gaithersburg, MD, USA, 2023, doi: 10.6028/NIST.AI.100-1.

[2] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1, Gaithersburg, MD, USA, 2024, doi: 10.6028/NIST.AI.600-1.

[3] ISO/IEC 42001:2023, Information Technology—Artificial Intelligence—Management System, International Organization for Standardization, Geneva, Switzerland, 2023.

[4] ISO/IEC 23894:2023, Information Technology—Artificial Intelligence—Guidance on Risk Management, International Organization for Standardization, Geneva, Switzerland, 2023.

[5] European Parliament and Council of the European Union, Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence, Official Journal of the European Union, 2024.

[6] Organisation for Economic Co-operation and Development, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, Paris, France, 2019, amended 2024.

[7] UNESCO, Recommendation on the Ethics of Artificial Intelligence, Paris, France, 2021.

[8] Infocomm Media Development Authority and AI Verify Foundation, Model AI Governance Framework for Agentic AI, Singapore, 2026.

[9] Cyber Security Agency of Singapore and FAR.AI, Securing Agentic AI: A Discussion Paper, Singapore, 2025.

[10] MITRE, MITRE ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems, 2026. [Online]. Available: https://atlas.mitre.org/

[11] OWASP Foundation, OWASP Top 10 for LLM Applications 2025, 2024. [Online]. Available: https://genai.owasp.org/

[12] S. Yao et al., "ReAct: Synergizing reasoning and acting in language models," in Proc. International Conference on Learning Representations, 2023.

[13] N. Shinn, F. Cassano, A. Gopinath, K. Narasimhan, and S. Yao, "Reflexion: Language agents with verbal reinforcement learning," in Advances in Neural Information Processing Systems, vol. 36, 2023.

[14] X. Liu et al., "AgentBench: Evaluating LLMs as agents," in Proc. International Conference on Learning Representations, 2024.

[15] S. Zhou et al., "WebArena: A realistic web environment for building autonomous agents," in Proc. International Conference on Learning Representations, 2024.

[16] C. E. Jimenez et al., "SWE-bench: Can language models resolve real-world GitHub issues?" in Proc. International Conference on Learning Representations, 2024.

[17] Q. Wu et al., "AutoGen: Enabling next-gen LLM applications via multi-agent conversation," arXiv:2308.08155, 2023.

[18] Z. Xi et al., "The rise and potential of large language model based agents: A survey," arXiv:2309.07864, 2023.

[19] R. Bommasani et al., "On the opportunities and risks of foundation models," arXiv:2108.07258, 2021.

[20] L. Weidinger et al., "Taxonomy of risks posed by language models," in Proc. ACM Conference on Fairness, Accountability, and Transparency, 2022, pp. 214–229.

[21] D. Sculley et al., "Hidden technical debt in machine learning systems," in Advances in Neural Information Processing Systems, vol. 28, 2015, pp. 2503–2511.

[22] E. Breck, S. Cai, E. Nielsen, M. Salib, and D. Sculley, "The ML test score: A rubric for ML production readiness and technical debt reduction," in Proc. IEEE International Conference on Big Data, 2017, pp. 1123–1132.

[23] S. Amershi et al., "Software engineering for machine learning: A case study," in Proc. IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice, 2019, pp. 291–300, doi: 10.1109/ICSE-SEIP.2019.00042.

[24] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," ACM Computing Surveys, vol. 46, no. 4, art. 44, 2014.

[25] M. Mitchell et al., "Model cards for model reporting," in Proc. Conference on Fairness, Accountability, and Transparency, 2019, pp. 220–229.

[26] T. Gebru et al., "Datasheets for datasets," Communications of the ACM, vol. 64, no. 12, pp. 86–92, 2021.

[27] I. D. Raji et al., "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing," in Proc. ACM Conference on Fairness, Accountability, and Transparency, 2020, pp. 33–44.

[28] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi, "Fairness and abstraction in sociotechnical systems," in Proc. Conference on Fairness, Accountability, and Transparency, 2019, pp. 59–68.

[29] J. A. Kroll et al., "Accountable algorithms," University of Pennsylvania Law Review, vol. 165, no. 3, pp. 633–705, 2017.

[30] L. Bainbridge, "Ironies of automation," Automatica, vol. 19, no. 6, pp. 775–779, 1983.

[31] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," Human Factors, vol. 39, no. 2, pp. 230–253, 1997.

[32] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," IEEE Transactions on Systems, Man, and Cybernetics—Part A, vol. 30, no. 3, pp. 286–297, 2000.

[33] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," Human Factors, vol. 46, no. 1, pp. 50–80, 2004.

[34] M. R. Endsley, "From here to autonomy: Lessons learned from human–automation research," Human Factors, vol. 59, no. 1, pp. 5–27, 2017.

[35] L. J. Skitka, K. L. Mosier, and M. Burdick, "Does automation bias decision-making?" International Journal of Human-Computer Studies, vol. 51, no. 5, pp. 991–1006, 1999.

[36] Z. Buçinca, M. B. Malaya, and K. Z. Gajos, "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, vol. 5, no. CSCW1, art. 188, 2021.

[37] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Algorithm aversion: People erroneously avoid algorithms after seeing them err," Journal of Experimental Psychology: General, vol. 144, no. 1, pp. 114–126, 2015.

[38] J. M. Logg, J. A. Minson, and D. A. Moore, "Algorithm appreciation: People prefer algorithmic to human judgment," Organizational Behavior and Human Decision Processes, vol. 151, pp. 90–103, 2019.

[39] P. Kruchten, R. L. Nord, and I. Ozkaya, "Technical debt: From metaphor to theory and practice," IEEE Software, vol. 29, no. 6, pp. 18–21, 2012.

[40] Z. Li, P. Avgeriou, and P. Liang, "A systematic mapping study on technical debt and its management," Journal of Systems and Software, vol. 101, pp. 193–220, 2015.

[41] G. P. Huber, "Organizational learning: The contributing processes and the literatures," Organization Science, vol. 2, no. 1, pp. 88–115, 1991.

[42] L. Argote and E. Miron-Spektor, "Organizational learning: From experience to knowledge," Organization Science, vol. 22, no. 5, pp. 1123–1137, 2011.

[43] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in NLP," in Proc. 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 3645–3650.

[44] D. Patterson et al., "Carbon emissions and large neural network training," arXiv:2104.10350, 2021.

[45] A. S. Luccioni, Y. Jernite, and E. Strubell, "Power hungry processing: Watts driving the cost of AI deployment?" in Proc. ACM Conference on Fairness, Accountability, and Transparency, 2024, pp. 85–99, doi: 10.1145/3630106.3658542.

[46] K. H. Tan, "The AI productivity-governance frontier: A theoretical model for enterprise value creation under agentic automation," International Journal of Science and Research Archive, vol. 20, no. 1, pp. 52–60, 2026, doi: 10.30574/ijsra.2026.20.1.1434.

[47] K. H. Tan, "Navigating the supervisory economy: AI stakeholder recognition and the transformation of organizational accountability," International Journal of Research Publication and Reviews, vol. 7, no. 6, pp. 7670–7675, 2026, doi: 10.55248/gengpi.07.0626.18a10.

[48] K. H. Tan, "Governance-aware agentic AI for enterprise engineering systems: A design-science reference architecture and quantitative risk-control model," ResearchGate Preprint, 2026, doi: 10.13140/RG.2.2.18948.69768.

[49] C. Schmitz, J. Rystrøm, and J. Batzner, "Oversight structures for agentic AI in public-sector organizations," arXiv:2506.04836, 2025.