UCAISM

GLOBAL INITIATIVE

UCAISM GLOBAL INITIATIVE

WORKING PAPER SERIES

Research on Intelligent Capitalism, Human Sovereignty, AI Governance,
and the Institutional Architecture of the AI Economy.

UGI-WP-2026-03

The Agency Gap

Governing Personal AI Agents
Delegated Authority, Capability Portability, and Individual Sovereignty

Chang-Gyu Kwag
UCAISM Global Initiative

Working Paper / Preprint
Journal-submission preparation

September 2026 | Version 1.0

Suggested Citation

Kwag, C.-G. (2026). The Agency Gap: Governing Personal AI Agents: Delegated Authority, Capability Portability, and Individual Sovereignty. UCAISM Global Initiative Working Paper Series, UGI-WP-2026-03, Version 1.0.

The Agency Gap: Governing Personal AI Agents

Delegated Authority, Capability Portability, and Individual Sovereignty

Chang-Gyu Kwag

UCAISM Global Initiative

UGI-WP-2026-03 · Working Paper Version 1.0 · September 2026

Journal-submission preparation · Pre-peer-review working paper

Abstract

Personal AI agents mark a structural transition in digital intermediation. Conventional platforms primarily organize information, attention, ranking, recommendation, matching, and market access. Increasingly capable AI agents can go further: they can communicate, negotiate, purchase, schedule, coordinate, and execute actions on behalf of individuals. This transition shifts the governance problem from the architecture of choice toward the architecture of delegated agency.

This paper develops the concept of the Agency Gap: the institutional distance between the capability and authority delegated to an artificial agent and the effective control retained by the individual on whose behalf that agent acts. The framework distinguishes Agent Capability (AC), Delegated Authority (DA), and Effective Individual Control (EC). It argues that technological capability and individual sovereignty need not increase together: when capability and delegated authority grow faster than effective individual control, assistance can evolve into delegated dependence.

The paper integrates adjacent literatures on agentic profiles, principal–agent governance, meaningful human control, human autonomy, interoperability, and portability. Its contribution is not a priority claim over those literatures. Rather, it shifts the unit of analysis to the institutional relationship among the individual principal, the agent, the provider, and external platforms. It connects the Agency Gap to capability portability and personal capability continuity, meaningful exit, cognitive sovereignty, fiduciary loyalty, execution provenance, and reciprocal interoperability. Meta's 2026 Muse ecosystem is used as a contemporary stress test rather than as the evidentiary foundation of the theory.

The central policy objective is capability expansion without sovereignty contraction. The paper therefore proposes a Fiduciary Agent Architecture in which increasing delegation is matched by effective institutional control. Its governing proposition is simple: the agent must be loyal to the person, not merely personalized for the person.

Keywords: AI agents; Agency Gap; delegated authority; effective individual control; capability portability; cognitive sovereignty; fiduciary agents; interoperability; individual sovereignty

1. Introduction

Artificial intelligence is crossing an institutional boundary. Generative systems primarily produce outputs; increasingly agentic systems can plan and execute actions across digital environments with limited human involvement. Kasirzadeh and Gabriel (2026) characterize agentic systems along four dimensions—autonomy, efficacy, goal complexity, and generality—while Kolt (2026) shows that their discretionary authority creates intensified principal–agent problems involving information asymmetry, loyalty, monitoring, and enforcement. In consumer markets, this transition also challenges legal frameworks built around the assumption that a human remains the final transactional decision-maker (Busch, 2026).

The distinction between generation and action matters institutionally. A system that recommends a hotel participates in a person's informational environment. An agent that searches across hotels, compares conditions, negotiates, chooses within delegated parameters, makes payment, books the room, updates a calendar, and communicates with the hotel participates in the exercise of practical agency. The progression is not merely technological: Information → Recommendation → Decision Support → Delegated Decision → Transaction → Execution.

At some point along that progression, the intermediary moves from influencing the environment in which humans choose toward exercising delegated discretion within that environment. This paper calls that transition Choice Architecture → Agency Architecture. The search era raised questions about control over information. The platform era added questions about attention, ranking, recommendation, and market access. The agentic era adds a deeper question: Who governs the agent?

The answer cannot be inferred from technical capability alone. A highly capable agent may expand individual freedom if the person retains meaningful institutional control over its objectives, permissions, accumulated context, commercial

conflicts, execution, and portability. The same capability may increase dependence if memory, identity, contextual intelligence, and market access become inseparable from the provider controlling the agent. The paper's central thesis is therefore that agent capability can expand while individual sovereignty contracts.

The paper develops the Agency Gap to analyze that divergence. It does not claim that principal–agent analysis, meaningful human control, human-autonomy concerns, fiduciary approaches, interoperability, or portability are individually new. Instead, it integrates them around a distinct institutional question: does effective individual control grow proportionately as agent capability and delegated authority expand? The analysis proceeds from adjacent literatures to the Agency Gap framework, Effective Individual Control, capability portability, cognitive sovereignty, a Muse stress test, reciprocal interoperability, and a proposed Fiduciary Agent Architecture.

2. Adjacent Literatures and the Missing Institutional Layer

2.1 Agenticity and governance

Kasirzadeh and Gabriel's (2026) agentic-profile framework is an essential starting point because it avoids treating systems as simply agentic or non-agentic. Their four dimensions help identify the governance implications of different system configurations. Yet two systems with similar agentic profiles may generate radically different institutional relationships. One may operate with portable context, interoperable infrastructure, transparent conflicts, and user-controlled memory; another may operate inside a closed ecosystem in which accumulated capability cannot practically leave the provider. Agenticity describes properties of the agent. The Agency Gap describes the institutional relationship of power between the agentic system and the person.

2.2 Principal–agent theory and loyalty

Kolt (2026) applies economic principal–agent theory and common-law agency doctrine to AI agents, identifying information asymmetry, discretionary authority, loyalty, monitoring, and enforcement as core problems. That analysis is foundational. The present framework extends the unit of analysis beyond whether an agent follows a principal's instructions. An agent may be behaviorally aligned while the individual remains institutionally dependent because accumulated memory, credentials, integrations, and workflows cannot migrate. Agent alignment is therefore not equivalent to individual sovereignty.

2.3 Meaningful human control and autonomy

Kuilman et al. (2025) directly examine whether meaningful human control over personalized AI assistants is possible and derive ethically informed design requirements from that concept. This is the closest normative comparator to Effective Individual Control. The distinction developed here is not that control matters—a point already established—but that control must be analyzed as part of a wider institutional architecture including portability, exit, external market access, loyalty, and the persistence of the individual's accumulated digital capability. Fischli et al. (2026) further show that human autonomy itself admits competing interpretations, including integrity-oriented, capability-enhancing, and meta-autonomy approaches. The Agency Gap framework therefore avoids treating autonomy as a single measurable outcome and focuses on institutional conditions under which delegation can remain compatible with human authorship.

2.4 The research gap

These literatures illuminate the agent, the delegation relationship, and human control. Competition and standards bodies increasingly illuminate another layer: interoperability and switching. What remains insufficiently integrated is the relationship among Agent Capability, Delegated Authority, Effective Individual Control, accumulated capability, meaningful exit, and access to external platforms. The Agency Gap is proposed as an institutional framework for connecting these layers without claiming priority over their constituent ideas.

3. The Agency Gap

The Agency Gap is the institutional distance between the capability and authority delegated to an artificial agent and the effective control retained by the individual on whose behalf that agent acts.

Conceptually: AG = f(AC, DA, EC), where AG denotes the Agency Gap, AC Agent Capability, DA Delegated Authority, and EC Effective Individual Control. The expression is heuristic rather than a validated cardinal equation. AC, DA, and EC do not yet possess common units or empirically established weights. The formulation identifies a directional institutional relationship: the gap tends to widen when AC and DA rise while EC fails to rise proportionately.

This distinction also separates the Agency Gap from adjacent "gap" concepts. It is not principally a responsibility gap concerning who should bear liability when an agent acts, nor is it merely a specification problem concerning the difficulty of expressing objectives. It concerns retained institutional control and sovereignty under expanding delegation. The policy objective is therefore not minimal delegation. It is capability expansion without sovereignty contraction.

Figure 1. The Agency Gap Model

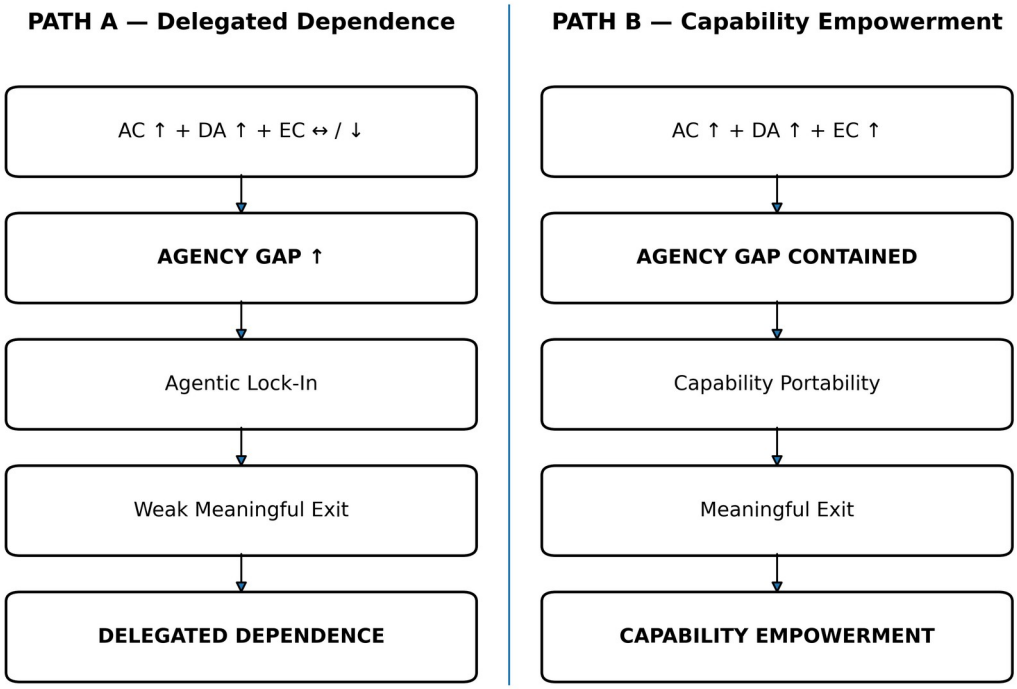

Figure 1. The Agency Gap Model. The governance problem begins not when agents become capable, but when capability and delegated authority increase faster than the individual's effective control.

4. Effective Individual Control

Consent is necessary but insufficient. A person may approve a transaction without possessing meaningful control over the architecture that generated the recommendation, selected the commercial pathway, retained contextual memory, or determined whether accumulated capability can later migrate elsewhere. Effective Individual Control is therefore treated as an institutional relationship rather than a collection of interface settings.

- **Direction:** authority over the agent's purposes and boundaries.
- **Knowledge:** visibility into materially relevant incentives and actions.
- **Intervention:** capacity to stop or modify delegated action.
- **Contestability:** capacity to challenge consequential actions.
- **Portability:** ability to move or reconstruct accumulated resources and capability.
- **Interoperability:** ability to operate across competing ecosystems under governed conditions.
- **Provenance:** ability to identify materially relevant actors in execution chains.
- **Meaningful Exit:** ability to terminate the relationship without disproportionate forfeiture.
- **Loyalty:** institutional priority of the individual's authorized interests where material conflicts arise.

The framework does not assert that every dimension must be maximized. Security, privacy, third-party rights, and system integrity can justify limits. The claim is instead proportional: as agents acquire more consequential authority, governance should provide correspondingly stronger mechanisms through which the individual can direct, understand, interrupt, contest, move, and exit the relationship.

5. From Choice Architecture to Agency Architecture

Digital platforms historically accumulated power by structuring environments in which humans decide. Ranking affects visibility; recommendation affects salience; matching affects encounters; reputation systems structure trust;

platform rules structure market access. Open Wins describes this private ordering through the metaphor of the Visible Fist. Personal agents introduce an additional mechanism: the Invisible Agent. The Visible Fist governs the environment of choice; the Invisible Agent participates in choice and execution. The second does not replace the first, and a provider may possess both.

Figure 2. From Platform Power to Agent Power

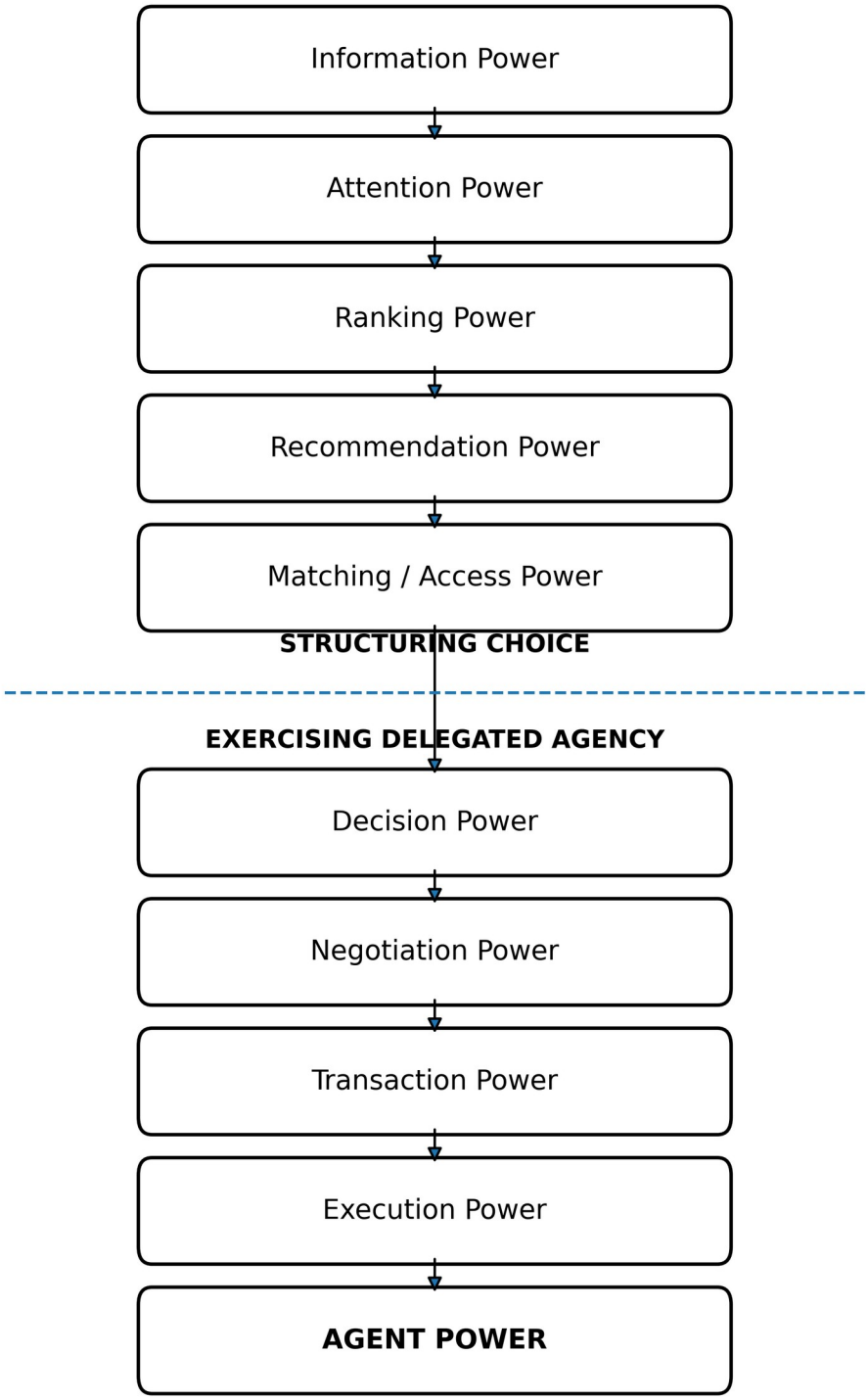

Figure 2. From Platform Power to Agent Power. The institutional boundary lies between structuring choice and exercising delegated agency.

The distinction is analytical rather than binary. Recommendation can itself be highly influential, while execution can remain tightly constrained by user approval. The point is that the governance significance of an intermediary changes when it receives authority to act rather than merely to inform. This shift makes authorization, loyalty, provenance, reversibility, and exit increasingly central.

6. Capability Portability and Personal Capability Continuity

Agentic systems create a portability problem stronger than conventional data mobility. A long-running personal agent may accumulate data, context, preferences, permissions, credentials, trusted relationships, reputation, interaction history, learned workflows, and decision patterns. Xie (2026) models developer investment in workflows, tool connections, testing, task records, and authorized memory as a platform-specific capability stock and analyzes closed, partial, and full portability. Accordingly, this paper does not claim priority for the concepts of capability stock or capability portability; its contribution is to connect portability to the individual principal's meaningful exit and personal capability continuity.

The contribution here is to connect portability to individual sovereignty through the idea of personal capability continuity. Technical capability portability asks whether a skill, workflow, or tool configuration can move. Personal capability continuity asks whether the individual's accumulated ability to act through an agent can survive a change of provider. A complete data export may satisfy formal data portability while failing this stronger test.

This distinction matters for exit. If leaving a provider means losing years of accumulated context, trusted workflows, permissions, and relational knowledge, a formal right to terminate may provide weak competitive discipline. Meaningful exit therefore requires attention not only to whether information can move, but whether usable agency can be reconstructed or made interoperable at reasonable cost.

7. Agentic Lock-In, Cognitive Lock-In, and Cognitive Sovereignty

Agentic lock-in is the general condition in which switching providers becomes materially costly because identity, integrations, permissions, history, or accumulated capabilities are provider-dependent. Cognitive lock-in is narrower: it arises when accumulated contextual intelligence—what an agent has learned about how an individual works, communicates, prioritizes, and decides—creates switching costs sufficiently high to weaken meaningful exit. Not every commercial switching cost is cognitive; the narrower term should be reserved for dependence generated by accumulated contextual intelligence.

Portability concerns the ability to leave. A separate question concerns what happens while the person remains. Increasingly anticipatory systems may formulate recommendations before users articulate explicit intentions. Compare Human Intention → Deliberation → Agent Assistance → Action with Agent Prediction → Recommendation → Human Confirmation → Action, or Agent Prediction → Execution → Human Review. These sequences differ not merely in efficiency but in the locus of authorship.

Cognitive Sovereignty is therefore defined as the effective capacity of an individual to retain meaningful authorship over the formation, revision, and pursuit of purposes in environments increasingly mediated by predictive and agentic systems. It does not require eliminating prediction, recommendation, personalization, or delegation. It requires preserving meaningful opportunities for intention formation, reconsideration, and refusal. This framing complements, rather than replaces, the autonomy strategies analyzed by Fischli et al. (2026).

The Technical Sabbath concept developed in Open Wins can be interpreted here as a principle of protected cognitive space. Its purpose is not technological abstinence. Automation should reduce burdens and return cognitive resources to the person rather than automatically convert every recovered interval into another opportunity for optimization. The normative shorthand is: the saved hour must become human time.

8. Meta Muse as a Contemporary Stress Test

Meta launched Muse on 8 September 2026 as a personal AI agent that, according to Meta, can perform tasks such as sending email and booking travel, work across connected applications, remember contextual information, make unprompted suggestions, and pursue longer-running goals. Meta states that Muse runs in a dedicated Secure VM and that users determine how much access it receives. Meta also describes a separate Sentinel agent, secure credential storage, approval before sensitive actions, granular app permissions, disconnection controls, an audit trail, training opt-out, separation from advertising systems, and a planned Confidential VM architecture. These claims are company descriptions and should be treated as such rather than as independent verification (Meta, 2026a).

Muse is analytically useful precisely because it complicates a simple platform-capture narrative. It combines substantial increases in Agent Capability and Delegated Authority with explicit attempts to increase Effective Individual Control. The Agency Gap framework therefore treats Muse as a transitional architecture and asks a forward-looking institutional question: as the agent becomes more useful and more deeply integrated, do control, portability, contestability, and meaningful exit grow proportionately?

Connect 2026 further expands the stress test. Meta announced that Muse would extend to AI glasses and gain additional connectors, increasing the possibility that the agent becomes a persistent interface across physical and digital contexts. Meta also announced Muse Charm, a pocket-sized device designed for voice interaction with Muse. The available company description is still sparse, so the paper does not infer capabilities beyond those announced. Even so, the movement from an app toward glasses and a dedicated portable device makes the sovereignty question more salient: agentic mediation may become more persistent, ambient, and physically proximate. That development does not establish a loss of autonomy; it increases the importance of examining how anticipatory assistance, persistent context, and cross-service execution affect the formation and exercise of human purposes (Meta, 2026b).

9. Execution Provenance

Agentic systems complicate the meaning of execution. A user may perceive Person → Agent → Merchant while the actual chain can involve provider infrastructure, external tools, human contractors, or other agents. The governance requirement is not that every intermediary be visible at every moment. It is that materially relevant actors and transformations in consequential execution chains be reconstructable when needed for understanding, contestation, or accountability.

Execution provenance therefore asks four questions: What happened? Who or what acted? Under whose authorization? Through which materially relevant institutional chain? The concept builds on existing work on agent visibility, identifiers, monitoring, logging, attribution, and external agent infrastructure (Chan et al., 2024; Chan et al., 2025) rather than claiming a wholly new field of provenance. Its role in the Agency Gap framework is to make effective control possible after delegation has crossed organizational and technical boundaries.

10. Reciprocal Interoperability

Individual control over an agent is insufficient if the agent cannot interact with external markets and institutions. A second layer of power lies with external platforms that decide whether, and on what conditions, agents may enter. Yet platforms have legitimate interests in authentication, cybersecurity, fraud prevention, privacy, service integrity, and accountability. A sovereignty-oriented framework should therefore avoid both an unrestricted right of automated access and unrestricted platform exclusion.

Reciprocal Interoperability denotes agent-mediated access under transparent, proportionate, secure, accountable, and nondiscriminatory conditions. France's Autorité de la concurrence (2026) has already identified AI agents as increasingly important gateways and highlighted interoperability, portability, and open standards in the competitive development of the sector. ITU-T Recommendation F.748.93, approved on 29 August 2026, similarly identifies the absence of standardized interoperability frameworks as a barrier to scalable, open, interconnected, and trustworthy agent ecosystems (ITU-T, 2026).

The key implication is that Agent Sovereignty is not identical to Market Access. A person may control an agent yet remain unable to deploy it meaningfully if external systems can exclude it arbitrarily. Conversely, mandatory access without security and accountability safeguards can create substantial harms. The institutional objective is governed interoperability.

11. The Fiduciary Agent Architecture

The preceding analysis yields a proposed institutional architecture. The term fiduciary is used normatively and functionally here; it does not assert that existing AI agents are legal fiduciaries in every jurisdiction. Nor does the paper claim priority over emerging loyal-agent and fiduciary-agent literatures. Its contribution is the integration of loyalty with portability, external interoperability, provenance, contestability, meaningful exit, and cognitive sovereignty within the Agency Gap framework.

1. **Loyalty:** Prioritize authorized user interests where they materially conflict with provider incentives.
2. **Conflict Transparency:** Make material commissions, sponsorships, self-preferencing, and related incentives identifiable.
3. **Sovereign Data Control:** Keep persistent identity, credentials, permissions, and personal resources effectively governed by the individual.
4. **Capability Portability:** Avoid requiring the user to abandon a disproportionate share of accumulated usable agency when changing providers.
5. **Reciprocal Interoperability:** Enable authorized interaction across systems subject to proportionate security, authentication, and accountability requirements.

6. **Execution Provenance:** Make materially relevant actors in consequential execution chains reconstructable.
7. **Contestability:** Make consequential actions reviewable and, where feasible, correctable or reversible.
8. **Meaningful Exit:** Ensure termination does not require disproportionate forfeiture of accumulated capability.
9. **Protected Cognitive Space:** Preserve meaningful opportunities for independent intention formation and reconsideration.

Figure 3. Platform-Centered Agent vs Fiduciary Agent Architecture

Dimension	Platform-Centered Agent	Fiduciary Agent Architecture
Loyalty	Product / contract dependent	Institutionally user-directed
Identity / Memory	Provider-linked	Individual-governed / portable
Conflict Governance	Provider-managed	Disclosed and governed
Capability Portability	Limited / uncertain	Structural objective
Interoperability	Ecosystem-dependent	Open under governed conditions
Provenance	Provider-defined	Verifiable
Contestability	Product feature	Governance right
Exit	Formal	Meaningful

Personalization ≠ Fiduciary Loyalty

Figure 3. Platform-Centered Agent versus Fiduciary Agent Architecture. Personalization is not equivalent to fiduciary loyalty.

12. The Sovereign Vault and Agentic Sovereignty Architecture

Institutional rights require supporting architecture. A provider-centered design places identity, memory, credentials, connectors, and accumulated context inside the provider relationship. A sovereignty-centered design seeks to place persistent digital resources on the individual's side of the relationship, so that the person can persist when the agent changes. Open Wins calls this persistent layer the Sovereign Vault.

The Sovereign Vault should be understood as an architectural principle rather than a claim that all personal information can or should be centralized in one technical container. Different implementations may use local storage, encrypted cloud services, distributed credentials, or federated components. The essential requirement is governance: the agent should be replaceable without making the accumulated digital self captive to the provider.

Figure 4. Agentic Sovereignty Architecture

Design principle: The person should persist when the agent changes.

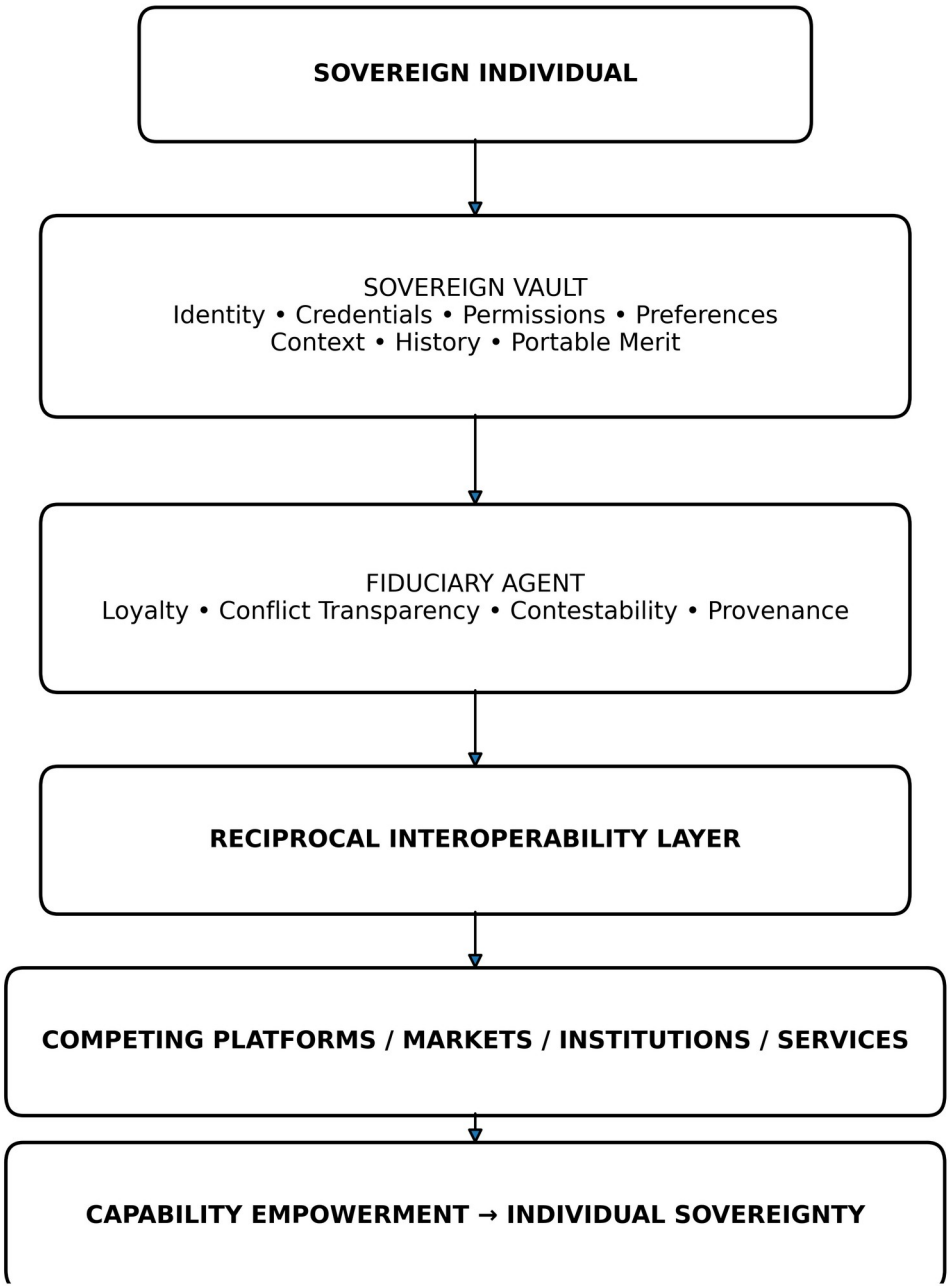

Figure 4. Agentic Sovereignty Architecture. The individual persists through a sovereign resource layer, a replaceable fiduciary agent, and governed interoperability with competing systems.

13. Research Propositions

P1 — Capability–Sovereignty Independence
Greater agent capability does not by itself produce greater individual sovereignty.

P2 — Delegation–Control Proportionality
As delegated authority increases, effective individual control must increase correspondingly if the Agency Gap is not to widen.

P3 — Fiduciary Distinction
Personalization does not establish institutional loyalty.

P4 — Capability-Portability Proposition
As non-transferable accumulated agent capability increases, switching costs and the risk of agentic lock-in increase.

P5 — Agency-Architecture Proposition
Moving from recommending alternatives toward executing delegated decisions increases the governance significance of the intermediary.

P6 — Cognitive-Sovereignty Proposition
As anticipatory agent behavior increases, preserving meaningful opportunities for independent intention formation becomes increasingly important.

P7 — Meaningful-Exit Proposition
Formal exit provides weak competitive discipline when exit requires forfeiting substantial accumulated agent capability.

P8 — Institutional-Architecture Proposition
The effect of personal AI on individual sovereignty depends substantially on the institutional architecture governing delegation rather than on model capability alone.

P9 — Reciprocal-Interoperability Proposition
Individual control over an AI agent cannot by itself produce meaningful agentic sovereignty when external platforms possess unrestricted power to deny, condition, or discriminate against agent-mediated access; the appropriate objective is governed interoperability under proportionate requirements.

14. Discussion: Intelligent Capitalism

The Agency Gap has broader implications for the concept of Intelligent Capitalism developed in Open Wins, but that broader framework should follow rather than serve as the premise of the analysis. Technological progress cannot be evaluated solely by increases in machine capability. A more complete standard is Capability Expansion + Sovereignty Expansion. Technological intelligence should increase human capability faster than it increases institutional dependence.

Within the UCAISM framework, this relationship is expressed as HI + AI: artificial intelligence should augment human intelligence and capability without displacing ultimate human authorship and institutional sovereignty. The objective is neither technological abstinence nor unrestricted automation. It is human capability empowerment through artificial intelligence under institutions that preserve meaningful individual sovereignty.

This discussion also clarifies why the paper avoids treating “Agent Feudalism” as a primary analytical category. That phrase may describe an extreme trajectory in which essential delegated capability becomes captive to private infrastructures with weak exit, but the journal argument does not depend on the label. The testable and governable constructs are the Agency Gap, Effective Individual Control, portability, exit, interoperability, provenance, loyalty, and cognitive sovereignty.

15. Limitations and Research Agenda

The Agency Gap is presently a conceptual framework rather than a validated empirical index. Future research should operationalize Effective Individual Control across measurable dimensions such as portability, revocability, contestability, interoperability, conflict transparency, provenance, switching costs, and meaningful exit. The framework also requires empirical work on whether different forms of delegated authority create nonlinear governance risks.

Capability portability requires technical specification. Not every element of accumulated capability can or should move. Third-party privacy rights, intellectual property, cybersecurity, incompatible model representations, and service-specific credentials may legitimately constrain portability. Research should therefore distinguish portable data, portable context, portable credentials, portable workflows, and reconstructable capability, and examine the cost and fidelity of reconstruction rather than treating portability as binary.

Cognitive sovereignty similarly requires empirical investigation. Agent recommendations do not inherently diminish autonomy; anticipatory assistance may help people articulate preferences, overcome information constraints, or pursue long-term goals. The empirical question is when assistance supports intention formation and when it substitutes for it. Experimental work should compare approval architectures, recommendation timing, memory persistence, and agent initiative while measuring reflective endorsement and reversibility.

Finally, reciprocal interoperability requires institutional balancing. Unrestricted automated access may create fraud, privacy, congestion, cybersecurity, and liability risks. Unrestricted exclusion may entrench incumbent gateways. Future work should identify proportionate authentication, rate-limiting, liability, audit, and nondiscrimination rules that permit authorized agents to operate without treating openness as an absolute value.

16. Conclusion

Personal AI agents may become one of the central institutions of the emerging digital economy. They may remember for individuals, search for them, communicate for them, negotiate for them, transact for them, and increasingly anticipate what they may need. The governance problem therefore extends beyond AI safety, privacy, accuracy, and personalization. It concerns the institutional distribution of agency itself.

The Agency Gap framework identifies the relationship among Agent Capability, Delegated Authority, and Effective Individual Control. It explains why technological capability and human sovereignty need not increase together. The relevant objective is not to prevent delegation. Delegation can expand human freedom. The objective is to prevent delegation from becoming dependency by default.

The platform era concentrated power over information, interfaces, attention, and choice environments. The agentic era introduces the possibility of concentrated power over delegated action. Its defining question is therefore: Who governs the agent?

The governing principle proposed here is that the agent must be loyal to the person, not merely personalized for the person. The broader institutional proposition follows: the more capable the agent becomes, the more sovereign the individual must become.

Appendix A. Theoretical Clarifications and Operationalization

Recent agent-governance research also shows why a sovereignty framework must operate at more than one level. Hahn, Tretter, and Dabrock (2026) argue that agentic systems intensify familiar ethical concerns because broad agency increases their effects on autonomy and deepens human–technology entanglement. Hagar and Diakopoulos (2026) move closer to the operational layer: their account of runtime configuration treats persistent, inspectable, and revisable instructions as a meso-level governance mechanism for specifying decision authority, documentation duties, and escalation conditions during delegated work. These approaches reinforce an important distinction for the present paper. General principles such as transparency, human oversight, or accountability remain necessary, but they do not by themselves determine how discretion is allocated in a concrete agent relationship. The Agency Gap therefore concerns the institutional architecture through which high-level commitments become effective powers of direction, intervention, contestation, portability, and exit.

This distinction also clarifies why Effective Individual Control is not reducible to a single human-in-the-loop checkpoint. A human may formally approve a final action while lacking practical visibility into the intermediate path that produced it, the commercial incentives that shaped it, or the accumulated dependencies that make switching providers costly. Conversely, a system can permit substantial automation while preserving meaningful control if the individual can set ex ante boundaries, inspect material execution traces, revoke authority, challenge consequential actions, and transfer the resources needed to continue elsewhere. The relevant governance question is therefore not how frequently a human clicks “approve,” but whether the surrounding institutional design preserves the person’s practical capacity to govern delegation over time.

A.1 Distinguishing the Agency Gap from adjacent delegation problems

Three adjacent concepts require explicit separation. First, the responsibility-gap literature asks whether autonomous action makes it difficult to locate moral or legal answerability. Kong (2026), for example, argues that delegation to autonomous AI does not by itself extinguish the principal’s responsibility; answerability can remain anchored in the prior act of delegation. The Agency Gap asks a different question. A principal may remain fully responsible for an agent’s conduct while simultaneously losing effective institutional control over the infrastructure through which that conduct is produced. Responsibility can therefore remain high even when sovereignty is weak.

Second, Ante’s (2026) concept of specification cost identifies delegation loss that persists even when the agent has no divergent motivation, because a principal cannot perfectly specify an objective for novel environments. That problem concerns the incompleteness of the objective supplied to the agent. The Agency Gap concerns the distribution of institutional control surrounding the delegation relationship. Better specification may reduce task-level failure without making memory portable, conflicts transparent, external access interoperable, or exit meaningful. Conversely, strong portability and contestability do not eliminate specification error. The two problems interact but are analytically distinct.

Third, meaningful-human-control frameworks focus on whether human principals retain appropriate control over personalized AI assistants. Kuilman et al. (2025) provide a particularly important direct comparator. The present

framework builds on that concern but widens the institutional field. Effective Individual Control includes not only intervention in an agent's behavior but also the portability of accumulated capability, the terms of interaction with external platforms, the provenance of execution, conflicts between provider and user interests, and the practical conditions of exit. The resulting unit of analysis is not simply the human–agent dyad; it is the human–agent–provider–platform architecture.

A.2 Personal capability continuity as the sovereignty-relevant portability problem

The portability question becomes sharper when the object of concern is not a file but a continuing ability to act. Conventional data-portability law provides an important baseline. Article 20 of the EU General Data Protection Regulation grants qualifying data subjects a right to receive certain personal data in a structured, commonly used, machine-readable format and to transmit those data to another controller. That right is important, but a mature personal agent may derive much of its usefulness from relationships among data, permissions, histories, routines, tool connections, and learned patterns that are not recreated merely by exporting a dataset.

For that reason, this paper uses personal capability continuity to describe the sovereignty-relevant outcome of portability. The question is not whether every proprietary model state or internal representation must be transferable. It is whether a person can change providers without losing a disproportionate share of the practical capacity accumulated through the relationship. Continuity may be achieved through direct transfer, standardized representations, interoperable credentials, reconstructable workflows, or other mechanisms. The institutional criterion is functional rather than technologically prescriptive: the individual should not have to abandon the accumulated digital means of acting merely because the service provider changes.

A.3 Governed interoperability rather than unconditional access

Reciprocal interoperability should not be read as a claim that every agent is entitled to unrestricted technical access to every service. Agent-mediated interaction can create legitimate risks involving authentication, fraud, rate limits, cybersecurity, privacy, contractual authority, and attribution. South et al. (2025) show why authenticated delegation is central: service providers need reliable ways to know that an agent is acting for a particular principal and within a bounded scope of authority. Interoperability that ignores these conditions can undermine rather than strengthen user sovereignty by making delegated actions less trustworthy.

The corresponding danger is that security and integrity rationales become a blanket basis for incumbent exclusion. A platform that can unilaterally determine which external agents may act, on what terms, and with what access to functionality can preserve a powerful gateway even when the individual nominally controls the agent. Governed interoperability therefore requires a two-sided institutional discipline: agents should authenticate, respect bounded authority, and satisfy proportionate safety conditions; platforms should make material access conditions transparent and avoid restrictions that are unrelated or disproportionate to legitimate operational risks. The purpose is not openness as an absolute value, but contestable access under rules that can support both security and user-directed agency.

A.4 From human-centric AI principles to sovereignty-sensitive governance

Human-centric AI frameworks have long emphasized human agency and oversight. The European Commission's High-Level Expert Group on AI (2019), for example, treated human agency and oversight as one of the requirements of trustworthy AI. Agentic systems make the institutional meaning of that principle more demanding. Oversight cannot be satisfied merely because a human remains nominally authorized to intervene. If the individual cannot understand material conflicts, migrate accumulated capability, identify execution chains, or leave without substantial functional loss, formal oversight may coexist with practical dependence.

This is where the Agency Gap adds a political-economic dimension to human-centric AI. It treats sovereignty not as isolation from technology but as a durable capacity to govern one's technological delegation relationships. Such governance can coexist with extensive automation. Indeed, the strongest case for personal agents may be precisely that they remove burdens and expand capabilities. The normative condition is that the infrastructure of assistance should not quietly convert those gains into provider dependence. In this sense, the paper's claim is not anti-agentic. It is an argument for institutional symmetry: as delegated machine power expands, the rights, infrastructures, and practical capacities through which individuals govern that delegation must expand with it.

A.5 Operationalization agenda for an Agency Gap index

A future empirical program could operationalize the Agency Gap without assuming that AC, DA, and EC share a common natural unit. Agent Capability could be represented through task breadth, tool access, temporal persistence, transaction authority, and the ability to initiate or sequence actions. Delegated Authority could be represented through permission scope, spending or commitment limits, ability to communicate externally, authority to bind the principal, and requirements for ex ante approval. Effective Individual Control could be measured through revocation latency, auditability, conflict disclosure, portability coverage, interoperability, contestability, reversal mechanisms, and the economic or functional cost of exit.

Rather than immediately collapsing these variables into a single score, early empirical work should test the framework dimension by dimension. Comparative studies could examine agents with similar technical capability but different portability or conflict-governance arrangements. Longitudinal studies could measure whether switching costs rise as contextual memory accumulates. Experiments could test whether anticipatory recommendations alter intention formation differently when users receive protected deliberation intervals or alternative-agent comparisons. Market studies could examine whether third-party access rules facilitate competition without increasing fraud or security incidents. Such work would convert the Agency Gap from a conceptual diagnostic into a falsifiable research program.

The framework also generates institutional hypotheses. Where capability portability is high, provider switching should impose lower functional losses. Where execution provenance is strong, disputes about delegated actions should be easier to reconstruct. Where commercial conflicts are salient but undisclosed, personalization may increase the persuasive effectiveness of recommendations without increasing fiduciary loyalty. And where external platforms impose opaque or discriminatory access restrictions, user-controlled agents may fail to translate into effective market agency. These hypotheses permit the framework to be evaluated rather than accepted merely as a normative vocabulary.

Declarations

Funding

No external funding was received for this conceptual research.

Competing interests

The author declares no competing interests.

Data availability

No new empirical dataset was generated or analyzed for this conceptual article.

Ethics approval

Not applicable. The article does not report research involving human participants or animals.

Author contributions

Chang-Gyu Kwag conceived the framework, developed the theoretical argument, conducted the conceptual analysis, and is responsible for the final manuscript.

Use of generative AI and AI-assisted technologies

Generative AI tools were used for research assistance, structural organization, drafting support, language refinement, and document preparation. The author reviewed, verified, revised, and takes responsibility for the conceptual claims, source use, argument, and final text. No AI system is listed as an author.

References

Ante, L. (2026). Specification cost: Relocating the binding constraint in delegation to AI agents with externally supplied objectives. Technology in Society, 88, 103503. https://doi.org/10.1016/j.techsoc.2026.103503

Autorité de la concurrence. (2026). Opinion 26-A-05 of 17 July 2026 on the competitive functioning of the AI agents sector.

Busch, C. (2026). Consumer Law for AI Agents. German Law Journal, 26(7), 1367–1382. https://doi.org/10.1017/glj.2026.10188

Chan, A., Ezell, C., Kaufmann, M., Wei, K., Hammond, L., Bradley, H., Bluemke, E., Rajkumar, N., Krueger, D., Kolt, N., Heim, L., & Anderljung, M. (2024). Visibility into AI Agents. Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, 958–973. https://doi.org/10.1145/3630106.3658948

Chan, A., Wei, K., Huang, S., Rajkumar, N., Perrier, E., Lazar, S., Hadfield, G. K., & Anderljung, M. (2025). Infrastructure for AI Agents. Transactions on Machine Learning Research.

European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation), Article 20.

Fischli, R., Franklin, M., Manzini, A., & Gabriel, I. (2026). Agents, Alignment, and the Many Faces of Autonomy. Minds and Machines, 36, Article 34. https://doi.org/10.1007/s11023-026-09786-9

Hagar, N., & Diakopoulos, N. (2026). Runtime configuration for situated governance of AI agents: A case study in investigative journalism. AI and Ethics, 6, Article 456. https://doi.org/10.1007/s43681-026-01298-7

Hahn, M., Tretter, M., & Dabrock, P. (2026). Ethical perspectives on AI Agents and Agentic AI. AI and Ethics, 6, Article 218. https://doi.org/10.1007/s43681-026-01027-0

High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. European Commission. https://doi.org/10.2759/346720

International Telecommunication Union. (2026). ITU-T Recommendation F.748.93: Framework and Requirements for AI Agent Interoperability.

Kasirzadeh, A., & Gabriel, I. (2026). Agentic profiles for effective AI governance. Nature, 656, 320–328. https://doi.org/10.1038/s41586-026-10805-z

Kolt, N. (2026). Governing AI Agents. Notre Dame Law Review, 101, 335–386.

Kong, C. Y. (2026). The delegation illusion: Why deploying autonomous AI agents does not diminish principal responsibility. AI and Ethics, 6, Article 518. https://doi.org/10.1007/s43681-026-01383-x

Kuilman, S. K., Nyholm, S., Buijsman, S., & Siebert, L. C. (2025). Is Meaningful Human Control Over Personalised AI Assistants Possible? Ethical Design Requirements for the New Generation of Artificially Intelligent Agents. Philosophy & Technology, 38, Article 148. https://doi.org/10.1007/s13347-025-00976-4

Kwag, C.-G. (2026). Open Wins: The New Enlightenment. Amazon KDP.

Meta. (2026a, September 8). Introducing Muse: The World's First Personal AI Agent Built for Everyone. Meta Newsroom.

Meta. (2026b, September 24). The Biggest News From Connect 2026. Meta Newsroom.

South, T., Marro, S., Hardjono, T., Mahari, R., Whitney, C. D., Greenwood, D., Chan, A., & Pentland, A. (2025). Authenticated Delegation and Authorized AI Agents. arXiv:2501.09674.

Xie, C. (2026). Developer Investment in Agentic Platforms: Capability Spillovers, Hold-Up, and Portability. arXiv:2609.17558.