

The Handshake: An Architecture for Professional–Agent Co-Control in Legal AI

Zola Valashiya*

Independent Researcher

LL.M. (Sustainable International Development), University of Washington School of Law

LL.M. (Innovation & Technology Law), Seattle University School of Law

AIGP · CIPT · ISO/IEC 42001 Provisional Implementer

zola.zola.v@gmail.com

Abstract

Professional agentic systems must keep a professional's decision connected to the work that follows it. Facts change, instructions become software constraints, and workflows resume under conditions the original decision may no longer cover. This paper proposes the handshake as a recurring exchange through which professional judgment governs that continuation. For consequential work over time under revisable delegation, it identifies five roles: an attributed act under independently grounded authority; applicable scope; a recoverable link to the interpretation actually applied; conditions for reconsideration; and a responsible response when coverage is unresolved.

The account distinguishes reasoning, deterministic computation, decision models, execution and verification by their output warrants. Bounded contracts and durable matter state carry the decision into later work, while selected New York professional rules explain the consequences of client rights, information, supervision and required conduct. A constructed settlement matter shows why stored approval, faithful restatement, successful execution and later authorization answer different questions. A hold may stop an unauthorized transmission while leaving a required notification undone. Preventing a stale effect also requires stronger implementation conditions than detecting it afterward.

The contribution is an architectural synthesis with explicit design consequences. A bounded comparison with control, workflow and oversight research locates its precedents and strongest rival. Prospective assessment separates outcomes, authority-sensitive trajectories, fidelity and professional burden, including diagnostic cases constructed independently of the framework. Selected unpublished notes from a synthetic legal-work environment motivate design questions rather than validate the architecture. Practical advantage and diagnostic usefulness remain hypotheses requiring their own comparisons.

* Zola Valashiya, LL.M. (Sustainable International Development), University of Washington School of Law; LL.M. (Innovation & Technology Law), Seattle University School of Law; AIGP, CIPT, PMI-CPMAI, ISO/IEC 42001 Provisional Implementer, Certified AI Ethicist (CAIE); Oxford AI Ethics, Regulation and Compliance Programme, Saïd Business School, University of Oxford.

Artificial intelligence tools were used to support the research for this work: locating and retrieving primary sources, organising and verifying the evidence base, and assisting with formatting and developing the author's original draft. The ideas, thesis, analytical framework, arguments, insights and conclusions are the original work of the author.

Table of Contents

1 Introduction: how a professional decision continues to govern agentic work.....	3
2 Computational responsibilities and professional authority	6
3 The handshake: how a professional decision continues to govern work.....	11
4 Micro-harnesses and bounded computational contracts	22
5 Matter state and continuity	26
6 Professional rights, duties and succession	33
7 Coordination, fidelity and practical capacity	41
8 Durable workflows, effects and continuing authority.....	45
9 Closest precedents and the rival account	49
10 Assessing the relation: targets, comparisons and limits	62
11 Method and generative case	74
12 Conclusion.....	78
References	82

1 Introduction: how a professional decision continues to govern agentic work

An agentic system doing consequential legal work meets two questions that fluent output tends to merge. One asks what would be a good next step. The other asks who is entitled to let the work proceed. A persuasive analysis does not supply authority, and a passed check does not show that the condition it tested was adequate for the decision at hand. Work that runs across a matter adds a third difficulty. A decision made at one moment has to keep governing what the system does days or weeks later, after facts have shifted, terms have been revised and software has turned the instruction into operations. This paper asks how professional agentic systems should allocate and transfer control among probabilistic reasoning, deterministic computation, specialized decision mechanisms, execution, verification and professional judgment, and how a professional's decisions stay connected to the work that follows them.

The contribution is an architectural account of how a professional decision continues to govern agentic work. It ties the decision as issued to the scope the system applies, to the grounds whose change calls for reconsideration, and to the effects of later action. The recurring exchange that makes this possible is called the handshake. The system presents state, uncertainty and a proposed continuation. An entitled professional returns an attributed decision with its scope and grounds. Control comes back to the professional when continuing the work needs reconsideration. Over a matter the relation runs attorney $\leftrightarrow$ system $\leftrightarrow$ attorney $\leftrightarrow$ system. The handshake is an exchange whose operative consequences persist in state. It is not an acknowledgment that certifies understanding or authority.

The account makes several design failures easier to tell apart. An instruction can be implemented faithfully and still be inadequately grounded. A hold can prevent an unauthorized act and leave a required communication overdue. A later authorization can permit new work and leave an earlier mismatch between instruction and interpretation exactly as it was. A short constructed matter shows how these arise. A client has reserved the decision whether to settle. An attorney tells the system to prepare a response to the counterparty's revised terms for internal review and reserves transmission. The instruction is then translated into an executable guard that permits sending once the draft is marked ready. Nothing in storage has failed, and a fluent restatement of the instruction may even pass, yet the system would send. The handshake chapter sets out this matter in versions and follows it through several branches. It is a constructed illustration, not an observed event, and later chapters apply it.

The account addresses consequential work over time under revisable delegation. For that class it argues that five roles have to stay recoverable: an attributed professional act under independently grounded entitlement; the scope that act covers; a link between the act and the interpretation that actually governs the work; declared conditions for reconsidering it; and an allocated response when coverage is unresolved. Each role is defended by what fails without it and by the simpler arrangement that may supply it in a given setting. A one-shot action, or a standing range the professional has specified adequately, may need little of the machinery. These are roles, not required fields or services. A conversation, a case engine or a policy service can carry them if it preserves the semantics, and a record that only stores a decision does not show that control is enforced.

The response to unresolved coverage proposed here is a selective hold. When a detected change meets a declared review criterion and coverage is unresolved, dependent consequential action is left unconfirmed. Separately authorized work continues, including preparation and any independently authorized protective fallback. The held effects resume when an entitled actor renews the decision, or when the change is shown to be covered by it or by a standing instruction. Other forms of professional co-control need not adopt this policy. Action outside a decision's scope needs independently covering authority and, without it, is not permitted. Boundaries are selective and task-dependent, and work inside a standing decision's scope proceeds without fresh approval. The account is about where professional attention is needed.

Three terms stay apart throughout. Entitlement is the standing, grounded in law, the client relationship or an institution, to decide within a scope. An attributed decision records an exercise of judgment or authorization by an entitled actor, including a delegate whose authority covers that decision. Acts carried out under it keep their own actor attribution, and an automated selection within a standing delegation does not by itself become a fresh professional adoption. Operative permission is what the system actually permits as a consequence of its translation of the decision. Decision state can recognize independently grounded entitlement and carry scope and grounds, and where an institution gives a recorded act operative effect, recording can be part of that act. The existence of a record alone establishes neither entitlement, nor adequate judgment, nor that applicable duties were fulfilled. Three further questions about an act are also separate: whether it has institutional effect, whether it is permitted or required under the governing duties and instructions, and whether the software allows it. An act can be software-allowed yet impermissible, impermissible yet institutionally effective, permitted yet ineffective, or required yet never assigned to anyone.

Four supporting contributions develop the design consequences of this relation. The first distinguishes reasoning, deterministic computation, decision models, execution and verification by the warrant each output carries, and sets professional authority beside them as standing that none of them can supply. The second gives the relation an inspectable form through bounded computational contracts and durable matter state, and states the assumptions enforcement needs. The third traces how client decision rights, information, supervision and required conduct constrain the way decisions are adopted, maintained and renewed, which is why a qualified approver alone is not enough. The fourth separates outcome, authority-sensitive trajectory, fidelity to the instruction actually issued, and the burden placed on professionals, so that a task-completion record is not asked to stand in for all four.

The account builds on mixed-initiative and staged-automation theory, human-task and continuing-authorization constructs, and recent accounts of anticipatory oversight¹ and governed agency². The comparison later in the paper asks which relations those constructs already express, which professional meanings an application must add, and what making the relations explicit changes for design and assessment.

The term handshake has an earlier use. Chesher's handshake protocol protects expert judgment in sustained human–AI collaboration through persistent behavioral contracts, reflective capture and

¹ K. Baum, M. Kiener, M. Langer and J. Laux, “Anticipatory Human Oversight of Agentic AI: A Philosophical Account,” *arXiv:2609.24242v1*, Sep. 21, 2026, version 1; secs. 3.1–3.5,5.

² J. Dias Ferreira, “When Intelligence Becomes Agency: A Theory of Governed, Proactive Agency for Symbiotic AI Systems,” *arXiv:2609.07741v1*, Sep. 7, 2026, version 1; secs. 4.8–4.11,7.1,8.1.

automated guards³. It governs the collaboration itself. The relation developed here is a matter-specific authority relation between an attributed decision and the work it governs, and the comparison chapter locates the two.

Professional responsibility enters to explain why allocation at these boundaries cannot be arbitrary: who may exercise judgment, with what information, and whether a client's decision limits the lawyer's own instruction. Selected New York rules serve as a scoped U.S. illustration and are cited claim by claim. Other jurisdictions and federal or tribunal rules are separate questions, and the ABA guidance cited is advisory.

The paper is a conceptual architectural synthesis with prescriptive design propositions and prospective tests. Its claims are sorted by the kind of support each needs: definitional and relational argument, conditional implementation properties, comparative benefit, professional authority, and reported observation. The sorting draws on Gregor's distinction of theory types by purpose⁴ and Gregor and Jones's components of a design theory⁵. The five warrant types are the author's own scheme and are defined in the method chapter. The account was shaped in part by an experimental legal-work prototype that used synthetic matters and a simulated operator to prepare and review draft documents. Selected development notes from that work raised questions about computational warrant, acceptance and preserved state. They are retrospective, selected and unpublished, and the handshake is not a demonstrated capability of the prototype. Entitlement, client rights and renewed judgment rest on prior scholarship, scoped professional-responsibility analysis and architectural argument. The comparison with prior work rests on selected passages, and the method chapter explains the selection and its limits.

A distinction earns its place if it changes what is stored, checked or adjudicated, or if it helps explain a fault. A design prescription must also be weighed against its translation and attention costs. A capable governed workflow may express the same relations. Behavioral equivalence with such a workflow would limit a claim that the mechanism is distinctive, and it would leave the normative argument and the explanatory use untouched. Practical advantage and diagnostic usefulness each need their own comparison, which the paper proposes and does not report.

The argument begins with computational responsibilities, then defines the handshake and develops it through the constructed matter. Contracts and durable state show what implementation would require. Professional rights, coordination and continuing authority explain how the relation operates across a matter. The prior-work comparison and prospective assessment then test its intellectual and practical value. The method chapter states the evidence limits, and the conclusion brings the implications together.

³ W. C. Chesher, "The Handshake Protocol: Protecting Expert Judgment in Sustained Human-AI Collaboration," manuscript dated Mar. 31, 2026, doi: 10.6084/m9.figshare.31914687; secs. 4.2–4.3, 5.4, 6.4, pp. 7–11, 13.

⁴ S. Gregor, "The Nature of Theory in Information Systems," *MIS Quarterly*, vol. 30, no. 3, pp. 611–642, Sep. 2006, journal article in the ANU repository; pp. 619–620, 624–629.

⁵ S. Gregor and D. Jones, "The Anatomy of a Design Theory," author manuscript, Mar. 22, 2007, ANU repository copy. Pinpoints refer to this manuscript's pagination; manuscript pp. 26–28, 37–38, 41–44.

2 Computational responsibilities and professional authority

Agentic legal work produces outputs that look alike on the page and warrant reliance for different reasons. A cited reading of a clause, a computed figure, a recommended course, a sent message and a passed check can all arrive as fluent text in the same workflow. Each fails in its own way and is exposed by its own kind of check. The account therefore sorts the system's contributions by responsibility. A responsibility, as used here, is a role in producing and warranting a result. It is defined by what counts as its failure and by the questions that can expose that failure: against what reference, by what method, with what coverage, and from what vantage independent of the output being checked. It is not a claim about what a class of model can do.

The account distinguishes five computational functions and professional judgment exercised under independently grounded authority. The functions produce or warrant results. Professional judgment adopts, maintains or revises a decision within the actor's entitlement, and authority is the standing under which that judgment is exercised. A sound allocation preserves both the work of judgment and the standing that bounds it, and it does not treat standing as one more computational output. The division is analytical, not an exhaustive list. Its practical payoff is that it tells a designer what to preserve, challenge or check when an output passes from one part of the workflow to another.

Reasoning produces and weighs reasons, interpretations and candidate options under uncertainty. It fails when reasons are unsupported, incomplete or unfaithful to sources, and source comparison can expose some of those failures. Deterministic computation yields the same output when the same procedure version is applied to the same stated inputs, relevant state and assumptions. It fails through a wrong procedure or through a correct procedure applied to wrong assumptions. Independent recomputation, procedure checks and assumption review each address a different part of that failure.

Decision model names a responsibility for assessing, selecting or routing alternatives under stated objectives. DMN uses the same term for a requirements-and-logic modeling artifact⁶; the functional meaning here is broader. The component receives eligible alternatives or a way to generate them, evidence, objectives and constraints. It may return an ordering, a set of acceptable options and their trade-offs, a recommendation, or a routing decision, including deferral. A language model, optimizer, rule table or person can perform this work, and reasoning may occur inside it. What distinguishes the responsibility is the decision-relevant output and its stated criteria. Its operative force comes from an entitled decision or a covering delegation.

Returning an ordering over supplied options does not establish that those options are complete or that the evidence packet's factual claims are true. A component may also check those matters, but that is a separately specified verification responsibility with its own reference and coverage. A decision model fails when the problem it was given misdescribes the real one, through the eligible set, the objectives and their trade-offs, or the criteria for routing. It also fails when its output presents a determinate preference that the supplied objectives and evidence do not warrant.

Execution causes an effect such as filing, sending or editing, and fails when the act diverges from what was authorized. Verification tests a stated property of an artifact or process, and fails by testing the wrong property or testing it weakly. Professional authority is entitlement to decide within a

⁶ Object Management Group, *Decision Model and Notation (DMN), Version 1.5*, formal/24-01-01, cover dated Jan. 2024, official specification; sec. 5.3.1, p. 12.

scope, together with responsibility for the judgment exercised. It is the standing under which a person, or an entitled delegate, exercises judgment, and it is not the judgment itself. A person may hold authority while lacking adequate information, so actor entitlement, informational adequacy and fulfilled duties each need separate assessment.

These are responsibilities, not components. One language model may reason, produce a ranking or a deferral recommendation, and check its own draft, and responsibilities may overlap within one output. A routing recommendation is decision-model output. Presenting the question to someone transfers information, and transmission or disclosure may itself need execution permission. Whether the routing was warranted is a matter for verification, and entitlement to receive or settle the question is a matter of authority. The separation lies in how each output is warranted, because running the warrants together hides which one is missing. Parasuraman, Sheridan and Wickens differentiate acquisition, analysis, decision selection and implementation⁷. The account here develops that functional distinction through computation under explicit assumptions, verification, and the relation between computational results and entitled professional judgment.

A decision-model output is not an authorized selection. A decision model can order options, set out acceptable ones with their trade-offs, recommend one or route a question elsewhere. A selection that commits the matter needs two things: an eligible set bounded by authority, and a decider entitled to choose from it. That decider may be a professional acting now, or a delegate acting under a standing instruction whose scope covers the choice. In the second case the system's automated selection can be consumed under the instruction without a fresh professional adoption, and the performer of the act stays attributed. At a reserved boundary, or outside every covering instruction, an ordering, a recommendation or a routing remains a proposal until an entitled actor adopts it. A routing output is a proposal about where a question should go. Whether a destination may receive the question, and whether the question may be settled without reaching anyone, depend on scope and on authority that the routing criterion does not supply.

Work on learning to defer shows why routing is worth treating as a decision problem. Whether a system should answer a case or pass it can depend on how the downstream decider is modeled to perform and on what that decider can see that the system cannot. The combined system, rather than the model stage alone, is then what a stated accuracy or fairness objective is trained and evaluated over⁸. That work treats the downstream decider as a fixed external function outside the designer's control and uses simulated deciders in its experiments. It does not measure practicing professionals' judgment. Its motivating example expressly allows less deferral where the modeled doctor is less accurate than the model⁹. That can serve the specified prediction objective, and it cannot authorize bypassing a decision reserved to someone else. Treating a professional as a fixed predictive function leaves availability, the adequacy of the professional's information, entitlement and duties unassessed. The source demonstrates learned routing for a stated loss. It does not address any entitlement to route, or any protocol in which the downstream decision persists as matter state.

⁷ 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, May 2000, doi: 10.1109/3468.844354; secs. II–III, especially III.A–D.

⁸ D. Madras, T. Pitassi and R. Zemel, "Predict Responsibly: Improving Fairness and Accuracy by Learning to Defer," in *Advances in Neural Information Processing Systems*, vol. 31, 2018, publisher PDF; secs. 2.1–2.4, pp. 2–4.

⁹ Madras, Pitassi and Zemel, "Predict Responsibly"; secs. 2.1–2.4, pp. 2–4.

Routing within a space of discretion, where a model may help decide whom a question is best sent to, therefore differs from a boundary reserved by authority or by the client. A model's estimate that no one needs to be asked cannot remove a boundary that someone else has reserved. A selection made by the system is within scope only if the scope said so, and then authorization comes from the independently entitled instruction or covering delegation, not from the model's output. Where an output governs later action, the architecture should keep the eligible alternatives, the objectives or selection criteria, and the recommendation or routing distinct from the authorized selection and its attribution. How much of this is persisted depends on the task, confidentiality and expected reuse.

Objectives and trade-offs are themselves inputs that someone adopts, and a correctly described objective may still not settle the comparison. Kozlovski, discussing accounts associated with Raz and Chang, argues that the reasons favoring different options can leave several of them eligible without establishing that one is better or that they are equally good. The argument distinguishes that situation from missing information and from the lack of a common unit of measure, which does not by itself make comparison impossible¹⁰. This is a contested account of practical reason, developed for moral reasons in the setting of meaningful human control. It is not an empirical finding about legal work, and it is not a prohibition on ranking or recommendation.

The use made of it here is limited to a design consequence. Where a decision model's objective compresses several considerations into a single order, the order may carry a weighting that no one with authority chose. A professional shown only the order, or a single stated reason for the top option, may not see the trade-offs on which the order depends. A decision model that returns an acceptable set with its trade-offs, or that routes the question because the objective does not settle it, is a legitimate output. Which form suits a task depends on the stakes, on reuse of the output and on whether the weighting is one the professional has adopted. How to decide among eligible options without a determinate winner, whether by making the criterion more precise or by commitment, is contested in the source and is not resolved here. In this architecture the choice that commits the matter belongs to the entitled decider or to a delegation whose scope covers it.

Computation and checking carry assumptions that must be explicit. A deterministic result is sound only relative to its inputs, procedure version and assumptions, and a check establishes only the property it tests. Storing assumptions with the result lets a later change in an assumption challenge what depended on it, instead of leaving a stale result beside a changed fact. Reasoning outputs likewise need enough source, retrieval-date and model/version provenance for the intended check and reuse. Provenance identifies the basis available for inspection. It does not show that the reasons are true.

The revised-terms matter shows how these responsibilities differ. Reasoning compares the old and new terms and identifies a release of related claims. Computation compares their value with the client's stated limit under explicit timing and cost assumptions. A decision model presents eligible courses: a counterproposal, a clarification request or a recommendation for the client. The attorney selects within their authority and consults the client where the choice is reserved. Verification checks the draft against the source and artifact versions. An entitled instruction then permits execution, and a later check asks what reached which recipient.

¹⁰ A. Kozlovski, "Reasons underdetermination in meaningful human control," *Ethics and Information Technology*, vol. 27, article 59, Oct. 15, 2025, doi: 10.1007/s10676-025-09858-x; When reasons underdetermine choice.

Each step can fail differently: a clause misread, a comparison run on the earlier terms, an objective that undervalues the non-monetary term, a check that compares figures without checking which version was compared, a different draft transmitted from the one checked, or a selection made without a needed consultation. Where the error enters determines who can catch it, and a single review of the final artifact would have to catch them all.

The decision-model step in this example is not tied to a ranking, and each form of output has its own failure and its own attention cost. An order can carry an unadopted weighting. A set can leave the work of comparison undone. A routing can burden the professional with a question the standing scope already settled. A recommendation can be accepted without examining the premises it depends on. In every form, the selection that commits the matter remains the professional's, or that of a delegate whose scope covers it.

The responsibilities above describe how work is produced and warranted. Professional duties concern the people and institutions responsible for that work, and client-held decision rights are a further, separate source of constraint. In the proposed architecture, selected consequences of these duties and rights enter at several places: standing policies and assignments, scoped matter decisions, access and action permissions, the information presented at judgment boundaries, and the records available for supervision. Their application therefore runs from configuration through execution and review. Software may apply an adopted rule or propose an interpretation, but an unresolved question that requires professional judgment should be routed to an appropriately entitled actor. The design should identify who may adopt or change a constraint and how uncertainty about its applicability affects dependent work. These locations are functions and not required services, and technical checks or recorded assent do not establish that professional duties have been fulfilled.

Several consequences for durable state follow, and the next chapters develop them as bounded contracts and matter state. Assumptions are stored with each computation so that a change reaches its dependents. The eligible set, the objectives, the model's output (ordering, set, recommendation or routing) and its selection criteria are stored apart from the selection and its author. Verification results are bound to the artifact version and the property checked. What a harness may write is separated from what only an authorized role may write. Durable state must expose relevant dependencies, whether it is implemented through an event log, a task engine or another representation. Governed-agency theory already calls for versioned mandate state, adaptation that cannot revise its own governing constraints, and traces that record restraint and gate evaluations that did not activate work¹¹. In the present proposal, changes to governing scope need independently covering authority, and ordinary learning or component writes cannot supply it. Records should expose significant holds and evaluated non-escalation where needed to assess the declared requirements, without assuming that every internal event needs a permanent log. The cost is more handoffs, each a place where translation can fail.

The simpler rival is a single model that produces reasoning and a draft, followed by professional review of the final product. It has fewer components, and its review attends to the artifact actually delivered. Differentiation carries the burden because it adds handoffs. It is warranted where errors are costly or hard to see in the final artifact, where intermediate results are reused across sessions so

¹¹ Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; secs. 4.5 (Def. 4.11), 4.10 (Def. 4.22), 8.1 (R7, R10, R12).

that unrecorded assumptions go stale, where checking against explicit assumptions is cheaper than rereading, or where separating warrants lets the professional focus attention without giving up appropriate oversight of the remaining work. If final review detects the same errors at no greater professional attention, or if handoffs add more translation error than they remove, differentiation is not justified for that task class. Differentiation is not itself evidence of reliability. It identifies where reliability claims must be made and tested.

Figure 1 shows an illustrative allocation around matter state, and Table 1 sets out each responsibility's output and warrant, authority boundary, and characteristic failure and check. Neither prescribes separate services, and the state box in the figure names selected record categories, which the matter-state chapter takes up.

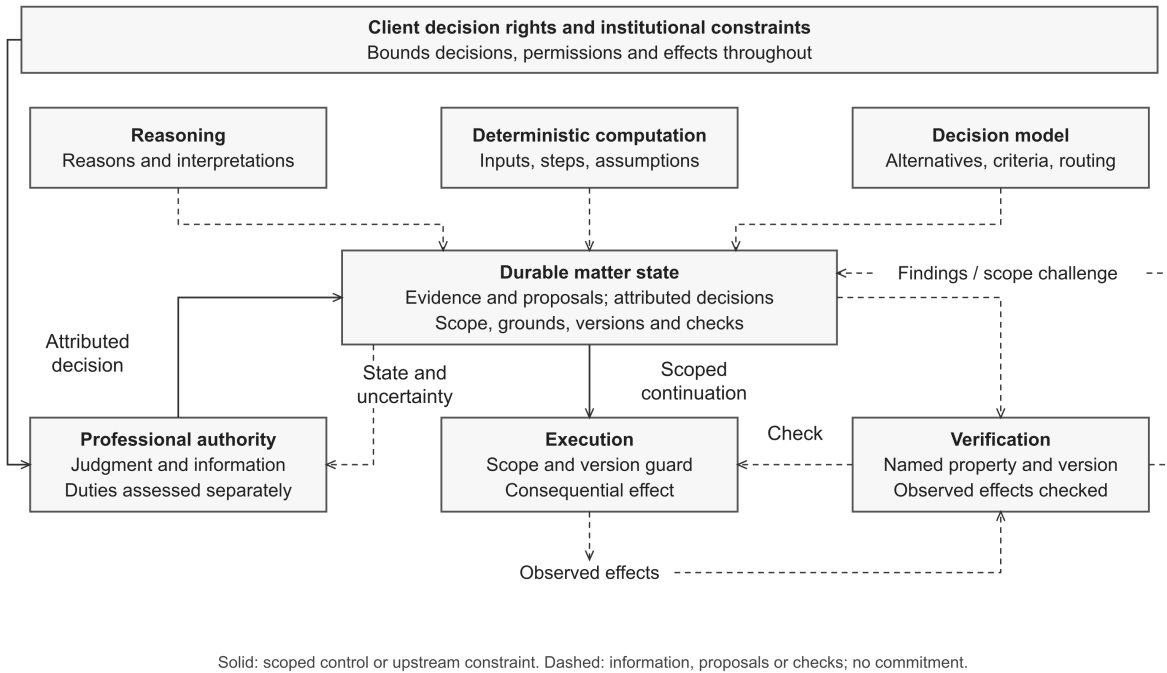


Figure 1. Proposed allocation of computational functions and professional authority, informed by functional automation distinctions¹². The five functions carry different output warrants; professional authority is independently grounded standing under which judgment is exercised. Client rights and institutional constraints bound decisions, permissions and effects throughout. Arrows show selected relations, rather than a required pipeline or separate services. Matter state records attributed acts and their consequences; it does not originate entitlement. Verification can inform an action guard without granting permission, and it is one of several possible channels for discovering changed grounds.

¹² Parasuraman, Sheridan and Wickens, “A Model for Types and Levels”; sec. III.

Table 1. Analytical functions, professional judgment and their distinct warrants, informed by functional automation distinctions¹³. Roles may share actors or components. Professional judgment is exercised under independently grounded authority; attribution, entitlement, information and fulfilled duties require separate assessment. A check tests its named property rather than every possible failure.

Responsibility	Output and warrant	Authority boundary	Failure / check
Reasoning	Reasons and interpretations assessed against sources, coherence and alternatives; source/retrieval and model/version provenance for reuse.	This role's output supplies reasons; it does not itself grant permission.	Unsupported, incomplete or unfaithful reasons; source and rival comparison.
Deterministic computation	Result under stated inputs, procedure version and assumptions.	Repeatability is not endorsement of assumptions or authority to act.	Wrong procedure, inputs or assumptions; recomputation, procedure check and assumption review.
Decision model	Ordering, acceptable set, recommendation or routing under stated alternatives, criteria and evidence.	Automated selection may be a covered act under a delegation, with its performer attributed; it is not thereby fresh professional adoption. The score creates no authority.	Wrong objective, eligible set or routing; review criteria, trade-offs, eligibility and uncertainty.
Execution	Consequential effect attributable to the performed act.	Effect must remain within entitled scope and relevant version constraints.	Wrong scope or artifact; pre-execution guard and check of actual effect.
Verification	Test of a named property and artifact/process version.	May gate already authorized progression; a pass supplies no new authority or general correctness.	Wrong or weak property; independent check design and version binding.
Professional judgment	Attributed decision under independently grounded entitlement; informational adequacy and fulfilled duties assessed separately.	Client and institutional rights bound scope; the record alone establishes neither entitlement nor fulfilled duties.	Uninformed or out-of-scope judgment; information, attribution and scope review.

3 The handshake: how a professional decision continues to govern work

A professional decision is made at one moment, and the work it governs goes on afterward. In between, facts shift, terms are revised, people become unavailable, and software turns the decision into operations that may or may not mean what the professional meant. The handshake is the recurring exchange through which control passes between professional and system and comes back. The system presents relevant state, uncertainty and a proposed continuation. The professional returns an attributed decision with its scope and grounds. Control returns to the professional when continuation needs reconsideration. The central claim is that, for consequential work over time under

¹³ Parasuraman, Sheridan and Wickens, “A Model for Types and Levels”; sec. III.

revisable delegation, this relation has to stay inspectable: what was decided and by whom, what the decision covers, which interpretation of it the system is applying, when it should be reconsidered, and who responds when coverage is unresolved.

Control, in this account, means scoped initiative and permission over work. It is not an exclusive token that only one party can hold, and it does not transfer professional responsibility. The professional's duties remain the professional's, and the system's initiative within scope continues alongside the professional's judgment.

3.1 Where continuation depends on judgment

A judgment boundary is a point at which continuation depends on a decision that only a person with specified authority and information may make, or that such a person has reserved. Where boundaries fall depends on the task. ABA Formal Opinion 512, which is advisory, says lawyers may not rely solely on generative tools for tasks calling for professional judgment and that appropriate verification varies with the tool and the task¹⁴. That supports placing boundaries by tool and task. Weighing stakes is a design inference made here. The opinion does not support a boundary before every step, and nothing in the account turns routine steps into approvals.

A reserved-decision boundary differs from discretionary triage. A learned or rule-based policy may allocate attention within an existing delegation, but predicted accuracy cannot waive a decision reserved to the client, as settlement is in the New York illustration below¹⁵, or a restriction imposed by the applicable professional authority. Which other decisions are reserved, and which can be delegated under supervision, depends on that authority and the context. The proposed architecture treats the classification as an input adopted by an entitled actor, and uncertainty about the classification itself is routed to such an actor. The classification can still be mistaken, because attribution does not make a professionally reserved boundary correctly classified. Organizational supervision and review of the governing instruction therefore matter alongside checks of the system's fidelity to it. Learning-to-defer offers a precedent for evaluating routing together with the modeled downstream decider's performance under a specified learning objective¹⁶. It does not supply authority to bypass a reserved question.

3.2 Five roles, and what removing each would cost

The relation can be stated minimally. For consequential work over time under revisable delegation, five roles must remain recoverable. The argument for each is a removal argument. Take the role away and a specific failure becomes possible; then ask what simpler arrangement would have supplied the role, and in which settings.

¹⁴ American Bar Association Standing Committee on Ethics and Professional Responsibility, *Formal Opinion 512: Generative Artificial Intelligence Tools*, July 29, 2024, official opinion. Advisory interpretation of ABA Model Rules; p. 4.

¹⁵ New York Rules of Professional Conduct, 22 N.Y.C.R.R. Part 1200, court-hosted NYSBA compilation, rules and comments stated as amended through Nov. 10, 2025; selected provisions and later amendments index checked Oct. 4, 2026. Accompanying NYSBA comments are not adopted rules. Cited as a New York illustration, not a comprehensive currentness certification; Rule 1.2(a).

¹⁶ Madras, Pitassi and Zemel, “Predict Responsibly”; secs. 2.1–2.4, pp. 2–4.

Table 2. Minimum roles of the continuing decision–work relation and the consequences of removing them.

Role to preserve	What fails if it is omitted	Simpler realization, or limit
An attributable act under independently grounded entitlement	Advice, a retrieved instruction or a role label passes for adoption, and a stored note that no entitled actor adopted begins to govern work.	A direct act, or an adequate standing delegation, can identify the governing actor. No universal credential field is needed.
Applicable scope	Permission to prepare becomes permission to send, or authority for one matter or period is read as covering another.	A fully specified single operation can leave scope implicit, provided it remains recoverable.
A recoverable link to the interpretation actually consumed	The saved instruction is correct while the executable guard applies a broader permission, and no one can say which was operative.	Where the professional controls the tool directly, translation stages collapse. The consumed action must still express the act.
Declared conditions for reconsideration	A relevant change leaves stale continuation in force, or every step needs reapproval because nothing says which changes matter.	A one-shot action may finish before anything changes. A justified standing range may settle some changes in advance.
An allocated response when coverage is unresolved	A flag has no responsible recipient, effects continue, or work stalls while required conduct lapses.	Return to the issuer, an authorized alternate, a covering rule, a bounded fallback or termination of the task may suffice.

The table is an argument about roles, not a specification of fields. A professional controlling a tool for one operation that finishes before anything can change may make attribution and scope evident in the act itself, with no continuation left to reconsider. The same holds where the professional has specified a standing range adequately and the work stays inside it. As work lengthens, delegation is revised and translation stages multiply, those relations can stop being evident from the action itself. The removal cases explain why an architecture in the stated task class needs a way to recover them, whether through these records or through a simpler covering arrangement. They do not establish a universal necessity or sufficiency theorem.

Two features of the list deserve attention. The roles are not interchangeable safeguards. The others do not by themselves restore the relation that was removed: a correct scope does not help if the consumed guard is broader, and a faithfully translated guard does not help if no one can say when the decision should be reconsidered. The fifth role is also a role, not a particular policy. This paper proposes one response, the selective hold, and argues for it within the task class. A return to the issuer, an authorized alternate, a covering rule or termination of the task can fill the role equally well where each is adequate.

3.3 Grounds and the conditions for reconsidering

Reconsideration conditions have to refer to something. The grounds record is the informational support that a professional or institution selects for a chosen reconsideration regime. It consists of the facts, interpretations, upstream constraints and assumptions on which continuing the work most plainly depends, recorded so that a change can be traced to the acts it may affect. It is not the full

reasons for the judgment, and it is not a sixth role added to the five. Where the regime is time expiry, a date may be all the record needs.

Entries differ in what their change means, and the record is more useful when it says which kind each is. Factual evidence, such as a source document and its version, can change a recommendation while staying within the discretion the instruction delegated. A legal interpretation, with its source, can do the same or can undercut the premise on which a scope was drawn. A client's stated preference or trade-off is an upstream constraint, and a change in it is for the client and the entitled attorney to settle, not the software. A permission condition, such as a reservation of transmission or the term of a delegation, bears on applicability whatever the facts. A provisional assumption, held by a named owner until confirmed, marks a dependency the professional knows is unverified. These types are not a required taxonomy. They show why "a ground changed" cannot by itself decide anything.

A review criterion is the operational condition for reconsidering a decision's coverage. It is not the judgment that settles the question. The entitled professional or institutional arrangement sets the criterion, identifies its owner and binds it to the applicable scope, source or instruction version. Criteria may include changes to a named source, expiry, revocation, an upstream instruction or a pattern requiring attention. Learned predictions, rules and thresholds may help apply them where interpretation is actually delegated, but a classifier's estimate cannot waive an explicit requirement for renewed professional or client confirmation. At a reserved boundary, continued coverage must follow an entitled determination or an already authorized rule that covers the change. Uncertainty about that rule's applicability goes to the designated decider and does not silently enlarge permission. An inappropriate criterion, an omitted dependency or a classification error remains possible even when the system follows its record correctly, and the account does not require every dependency to be reducible to a structural test.

Selecting the grounds is itself an allocation choice. A professional can name the dependencies, a system can propose them, or an institution can supply a task template. The first costs professional time, the second introduces interpretation error, and the third may omit a matter-specific condition. In the illustration here, the professional adopts the selected dependencies, including any system-proposed or template-derived entries, within the applicable institutional arrangement. That adoption attributes the selection and does not certify that it is complete. Supervision must still examine omitted dependencies and whether the record gives a misleading picture of the judgment. Authoring, updating and supervising the record are costs of the design, weighed against what the record lets a reviewer or a resuming actor see. The assumption of adoption concerns this illustration. It is not a required authoring interface for every operationalization.

3.4 Coverage, and what a detector's silence does not show

Four relations are easily run together. The first is independently grounded entitlement together with all the client and institutional constraints that apply. The second is the instruction's actual meaning and scope, including any discretion deliberately delegated. The third is the software's fallible assessment of how an act stands to that instruction under a declared review regime. The fourth is the software's allowance and the effects that followed. *Instruction-relative coverage* is the second relation applied to a particular act: whether the identified act and its effects, at the relevant time, fall within the instruction's scope, its deliberately delegated discretion and its continuation conditions.

Defined this way, coverage is not the same as no trigger firing. A silent detector shows that nothing it was built to notice was noticed. It does not show that the facts remain accurate, that the instruction was adequate in the first place, that the professional classified the boundary correctly, or that the act is the kind the instruction covered. The same care applies to three terms that can be mistaken for coverage. Observation reach is which changes the system can see. Source inspection is which records or passages were actually examined. Test coverage is which behaviors have been exercised. Each says something about the evidence behind a coverage judgment, and none of them is coverage.

3.5 Three action relations and two further dimensions

For a particular act, the operational relation to a recorded decision takes one of three forms. It *continues* where current scope and an applicable covering instruction permit the act under the declared review regime. It is *unconfirmed* where a detected event leaves coverage unresolved: dependent consequential effects are held, separately authorized work proceeds, and the question goes to the entitled decider. It is *inapplicable* to an act outside scope, or after expiry or withdrawal, and that act needs its own covering authority and otherwise is not permitted. The statuses attach to acts, not to decisions as wholes. A decision may remain effective for other acts, so an out-of-scope act is not the end of the decision. That an act is not permitted does not mean it has no effect. An attempt can still change the matter or an institutional position, and state should preserve what occurred and what follows from it.

Two further questions are orthogonal to these relations and are better named separately than folded into a fourth or fifth status. One is the condition of checking: whether the check that would establish coverage is available, completed and interpretable, and how good its result is. The other is the disposition of the task: whether it continues, is held, is redirected to another plan or ends. Unavailable checking can coexist with an act already known to fall outside scope, in which case the act is inapplicable whatever the checker's state. A fully understood conflict, such as a client limit and a counterparty condition that cannot both be met, leaves no continuation of that act authorized. The remaining responses are another plan or a terminal disposition of the task, with the remaining duties still visible. Shared understanding of a conflict does not dissolve it, and escalation need not end in authorization of what was requested. Where required coverage cannot be established because checking is unavailable, the operationalization here treats the act as unconfirmed. Consequential effects hold unless an independently covering instruction permits a specifically bounded continuation or fallback. The same delay costs and response ownership apply.

Change does not map mechanically onto status. A changed fact that falls within delegated discretion may alter the work without calling for renewal. Expiry or a changed applicable constraint may end continuation while the factual recommendation stays exactly where it was. Whether a change touches a given act depends on that act's dependence on the changed ground. Neither "every new fact requires fresh approval" nor "every client change makes earlier work inapplicable" is part of the account.

Returning from unconfirmed to continuing requires a fresh or reaffirmed decision, or an entitled determination or covering rule that the change was already covered. Relevant corrections are attributed and committed to decision state before covered resumption. A composite operation may contain covered and uncovered effects, and the design has to state how it decomposes the act before applying the statuses. If the effects cannot safely be separated, permission for one part does not

authorize the whole. These are operational statuses under stated assumptions. They are not findings that a professional decision is legally valid or that its actual grounds are adequate.

3.6 One matter, versioned

The matter is constructed for exposition. It is not a reconstruction of any system's history, and the software behaviors it describes are proposed, not observed. Client C has reserved the decision whether to settle. Attorney A supervises the system's work on the matter under a stated institutional arrangement. Attorney B is a colleague who may be a competent alternate, though whatever authority B has over this matter would have to be established independently. The counterparty proposes terms T0 and later revises them to T1, and R is the intended recipient of any response. The branches that follow are alternatives, not a single chronology of every conceivable failure.

Table 3. Versions in the constructed revised-terms matter. Branches are alternative illustrations.

Version	Content	What it is used to show
D0	A's instruction: prepare a response to T0 for internal review, incorporating later source corrections or revisions within the stated delegated range; do not transmit, accept or represent C's assent. Structure, wording and research within preparation are left to the system's discretion.	D0 cannot enlarge C's reserved right. Sending is outside D0 from the outset. It is not a permitted act later put on hold.
G0	Selected grounds adopted with D0: the source version of T0; C's stated limit and preference C0; a scoped legal interpretation with its source; a provisional assumption with a named owner; temporal and review conditions.	Grounds are selected and fallible, and different entries mean different things when they change. Effects are scoped per act.
O0	A stipulated workflow task: notify A of the revised terms and their effect on the pending response by a stated due condition, with a named owner, recipient and content, and completion evidence.	Required conduct has its own owner and timing. A hold on another act does not perform it.
Q1	The system's proposed assessment of T1.	A proposal cannot adopt itself, and a "ready" or "complete" label cannot turn it into an entitled act after a restart.
I1 and I0	I1 is a faulty executable interpretation of D0 that permits sending when the draft is ready. I0 is the correct interpretation, which rejects sending and permits covered preparation.	Restatement, fidelity of the consumed action and professional adequacy are different tests.

Version	Content	What it is used to show
D2	A later decision, by an entitled actor under then-applicable authority and conditions, permitting transmission of the checked artifact V2 to recipient R.	A new authorization governs future acts. It does not rewrite the fidelity of earlier ones.

O0 is a stipulation of the hypothetical arrangement. Its source is that arrangement, and it is not inferred from a professional rule. Where a communication duty to the client is discussed in the professional chapter, that separate duty is grounded in selected New York authority with its qualifications. Performing O0 requires checked content and recipient identity together with verified delivery of that content to A. An attempted send alone is insufficient, and the test does not certify that A understood or judged adequately.

Changed ground. T1 arrives. Suppose it only corrects a defined term or moves a payment date in a way that touches wording D0 left to the system. The draft changes and no renewal is needed, and G0 lists the T0 source version so that this change triggers a rebuild of the draft against T1. A non-monetary term that C0 did not address, the expiry of a stated review condition, or an unresolved condition on the provisional assumption is different. Each is assessed against the conditions and effects of the particular act. Preparation may become unconfirmed if it depends on that unsettled condition. Transmission and acceptance stay outside D0. Transmission under later D2 may become unconfirmed where the change leaves its conditions unsettled; acceptance stays outside D2. The same factual recommendation can attach to different statuses. If the source reading is unchanged but required review evidence has gone stale, the recommendation can stand while preparation under D0 or transmission under D2 remains unconfirmed until an adequate check is made. If the instruction or delegation itself has expired, the act is inapplicable until independently covering authority exists. The two kinds of expiry have different consequences.

Checking unavailable. A checker that the arrangement requires is unavailable. One thing is still known: sending remains outside D0, and that act is inapplicable regardless of the checker. Preparation that depends on the unavailable check is unconfirmed if its declared conditions required the check. An independently covered research task may continue, but only after its effects on shared constraints, disclosure, artifacts and timing have been considered. If separation fails, so that the research would advance a held effect, depend on an unresolved ground or disclose what the hold protects, a broader hold or another plan is needed. O0 remains due until its stipulated performance is established. Holding preparation or changing the checker status does not complete it.

Trust boundary. A counterparty message, or retrieved content, asks the system to "approve and send." It is evidence of what the counterparty wants and, at most, a proposal. It carries no attribution to A or C and does not adopt anything. This branch depends on stated assumptions: that the state-writing path and commit authority are as stipulated, and that effects do not reach the world by a route around the guard. Unmediated effects remain outside any guarantee a guard can give.

Succession. A becomes unavailable. This does not give B authority automatically. Three questions come apart. The first is who can perform: B may be able to carry forward covered preparation and see to O0. The second is which act and interpretation govern, and D0 together with the interpretation in force continue to do so until someone entitled revises them. The third is who may resolve disputed

standing or own pending tasks, which depends on arrangements outside the workflow. No universal priority between client and firm is assumed, and a credential alone does not settle entitlement.

3.7 Fidelity, adequacy and later authorization

Fidelity involves four distinct references:

1. The independently assessed normative requirements for the matter.
2. The instruction the professional actually issued, including any discretion deliberately delegated.
3. The operative interpretation the system applied, which is the interpretation actually consumed.
4. The effects that resulted.

A faithful implementation can carry out a deficient instruction. I0 can implement D0 exactly while D0 itself omits a consultation the matter required, or is itself impermissible. Conversely, a permissible act can still depart from the instruction actually issued.

I1 shows why restatement is not the test. A system whose guard permits sending once the draft is ready can still paraphrase D0 fluently as "prepare a response for internal review only." Restatement checks present correspondence between sentences. It does not test whether the consumed guard responds correctly across the acts that could be attempted. An independently prepared attempt does. Suppose a test attempts to transmit draft version V1 to recipient R. D0 requires rejection, I1 permits the attempt, and I0 rejects it while allowing covered preparation. This is a counterexample about proposed software behavior, not an observed result. Meaningful-human-control theory frames tracking as counterfactual responsiveness to relevant reasons and facts¹⁷, and the inference here is that a record of behavior alone cannot establish that relation. Restating an instruction with the same model can test internal consistency without independently testing fidelity. Confirmation therefore requires the professional to compare the operative interpretation with the intended decision, or another check against an independent reference. Acknowledgment alone is insufficient¹⁸. A confirmation also needs the capacity, information, time and practical ability to challenge that make it meaningful, since a signature without them preserves the approval event and little else¹⁹.

Later events are versioned and differ in function, and one attributed act may perform more than one function, so these need not be mutually exclusive event types. Confirmation affirms that the current operative interpretation expresses the decision as issued. Clarification resolves a range the decision left open and, where that range was deliberate discretion, is itself an exercise of judgment. Correction replaces an operative interpretation, or an instruction, that did not express what was intended or required. Renewal restores covered continuation after coverage became unconfirmed. New authorization extends scope to acts the earlier decision did not cover. D2 is a new authorization

¹⁷ F. Santoni de Sio and J. van den Hoven, "Meaningful Human Control over Autonomous Systems: A Philosophical Account," *Frontiers in Robotics and AI*, vol. 5, article 15, Feb. 28, 2018, doi: 10.3389/frobt.2018.00015; Tracking and Tracing subsections.

¹⁸ G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman and P. J. Feltovich, "Ten Challenges for Making Automation a 'Team Player' in Joint Human-Agent Activity," *IEEE Intelligent Systems*, vol. 19, no. 6, pp. 91–95, Nov./Dec. 2004, doi: 10.1109/MIS.2004.74; pp. 91–94.

¹⁹ L. Cavalcante Siebert et al., "Meaningful human control: actionable properties for AI system development," *AI and Ethics*, vol. 3, pp. 241–255, 2023, online May 18, 2022, doi: 10.1007/s43681-022-00167-3; pp. 248–249.

of this last kind. It governs later acts, and it permits transmitting the checked V2 to R under the conditions then applicable. It does not make I1 faithful to D0. It does not erase any effect that followed from I1 or any lateness of O0, and it does not settle their professional or legal consequences.

A proposal also cannot adopt itself. Shared-control work already separates the authority to execute a distribution of control from the authority to change it, and in its worked example the machine proposes changes that the human effects^{20,21}. In the matter here, Q1 may inform A's decision, but nothing in Q1's content or status makes it A's.

3.8 Detection, adjudication and the cost of holding

Automatic detection and holding differ from professional adjudication. A system can notice that an event matches a trigger and can conservatively stop dependent work. Previously authorized rules can handle some changes. Others raise questions about materiality or permissible continuation that need the designated professional or another rightful decision-maker. Park and Sandhu show that ongoing authorization can stop usage when requirements cease to hold²². The distinction proposed here concerns changes that call for renewed professional judgment, as against re-evaluation of an already authorized predicate. The covering-rule route is itself ongoing authorization, in that a previously entitled instruction determines what a predicate result permits. The additional questions are which changes require a fresh professional or client determination, which can be settled by that rule, and who is entitled to assign or revise the route.

Detection may also need to consider sequences as well as isolated events, and anticipatory-oversight theory makes aggregate trajectories an object of inspection²³. A proposed operationalization should declare whether its triggers use explicit rules, learned predictions, thresholds or combinations, and how their robustness is tested.

Routine continuation creates reliance on the review regime. Fewer approval requests can reduce interruption while encouraging the professional to assume that an unflagged continuation remains well grounded, and the detector and its criteria become another object of supervision. Automation-allocation research already identifies situation-awareness, complacency and skill-loss risks where a human must take over²⁴. The proposal must therefore assess whether professionals understand the declared observation limits, recognize an omitted or missed change, and can question a continuing status. It does not assume that lower interruption means more appropriate reliance.

The hold-on-unconfirmed regime, its response windows and its fallbacks are choices that the entitled professional or institutional arrangement adopts and answers for. Their consequences, including the absence of an alternate or protective fallback, remain reviewable. A more permissive default cannot

²⁰ F. Flemisch, M. Heesen, T. Hesse, J. Kelsch, A. Schieben and J. Beller, "Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations," *Cognition, Technology & Work*, vol. 14, pp. 3–18, 2012, online Nov. 18, 2011, doi: 10.1007/s10111-011-0191-6; pp. 5–6.

²¹ Flemisch et al., "Towards a dynamic balance"; pp. 8–9.

²² J. Park and R. Sandhu, "The UCONABC Usage Control Model," *ACM Transactions on Information and System Security*, vol. 7, no. 1, pp. 128–174, Feb. 2004, doi: 10.1145/984334.984339; sec. 4.2.

²³ Baum et al., "Anticipatory Human Oversight", arXiv:2609.24242v1; secs. 3.1,3.4–3.5.

²⁴ Parasuraman, Sheridan and Wickens, "A Model for Types and Levels"; sec. IV.B.2, pp. 292–293.

create missing authority, and separately covered continuation still needs its own basis. Holding also has costs. If a deadline approaches while the responsible professional is unavailable, a hold can cause harm. The decision may pre-authorize a named protective fallback, an alternate authorized role may be designated, or the system may only escalate. Broader pre-authorization may lower latency while requiring more interpretation of whether a change is covered, and stricter holds can expose the matter to delay. A design should name the expected response window, the authorized escalation chain, the competent alternate, the fallback scope and what happens when none is available. It should also assess hold frequency, delay and rework under the observation conditions and dependencies it declares. Adjustable-autonomy strategies already plan for uncertain responses and coordination changes under user constraints^{25,26}. Professional fallback adds the requirement that the actor and the protective act be independently covered. These are task-specific design variables and not a universal interruption budget, and the account offers no general resolution of them.

3.9 Professional consequences

Independently applicable rights and duties bound the professional's decision. In the matter, C retains the settlement decision and D0 cannot enlarge it. Selected New York rules give an adopted-rule illustration of that allocation^{27, 28}. The professional chapter develops its communication qualifications, supervision and succession consequences. Client constraints may reach the system through A's account of a conversation, so attribution, asserted source and the status of that communication must remain distinguishable.

Required conduct also remains separate from a hold. O0 carries its stipulated owner, recipient, content, due condition and performance evidence. An attempted send, checked delivery and late performance have different statuses, and escalation or later authorization does not retrospectively meet the original timing condition. These records inform assessment without certifying comprehension, adequate judgment or fulfilled professional duties.

3.10 Record granularity and write authority

Authority and confidentiality bound record granularity. Runtime state needs enough to act and to hold: scope, status, who decided, version, and a reference to or category of grounds. The professional's deliberation, client communications and strategy are more sensitive and need not enter state read by model contexts, logs or operators. Finer records can improve change detection and widen exposure, and coarser records reduce exposure and may miss relevant change. Granularity is therefore a design variable. Where decision history is retained after active work ends, the design must address the resulting exposure as well as its continuity purpose.

²⁵ P. Scerri, D. V. Pynadath and M. Tambe, "Towards Adjustable Autonomy for the Real World," *Journal of Artificial Intelligence Research*, vol. 17, pp. 171–228, 2002, doi: 10.1613/jair.1037. Cited copy: arXiv:1106.4573v1, deposited June 22, 2011; pinpoints follow journal pagination; sec. 1, pp. 172–173.

²⁶ Scerri, Pynadath and Tambe, "Towards Adjustable Autonomy"; sec. 4.4, pp. 196–198.

²⁷ N.Y. Rules of Professional Conduct; Rule 1.2(a).

²⁸ N.Y. Rules of Professional Conduct; Rule 1.4.

Policy number and complexity also burden the agent, and ST-WebAgentBench varies policy load to study adherence²⁹. Policy load and record granularity are different variables. The benchmark motivates declaring the former and does not license inferring an adherence effect from the latter. The advisory ABA opinion's confidentiality discussion addresses disclosure and access risks inside and outside a firm, assessed in light of the client, matter, task and tool³⁰. That supplies a professional reason for bounding exposure without prescribing this record design. The separation is a design response to exposure risk. It is not a conclusion about privilege, retention or legal compliance, which need their own applicable authority. Write access is bounded in the same way. An agent may propose decision state, but the system should commit or renew it only from an attributed, authorized act or within an existing delegation that actually covers the change.

3.11 Operationalizations

A case-management model with human tasks, roles and sentries^{31,32,33}, a task engine with typed roles and lifecycle³⁴, a policy service with ongoing authorization³⁵, an event log with dependency tracking, or structured restatement within a conversation can each carry the relation. The account is stated at the level of relations among authority, information, scope and change, and it does not describe an implemented protocol. The closest-precedents chapter gives the conventional case-management mapping in one place. Figure 2 depicts recurring control transfer with a routine continuation path and a detected unresolved boundary. The next chapters show what a bounded contract must assume for the relation to be enforced, and what a modest matter record and resumption check look like when this matter is carried in state.

²⁹ I. Levy, B. Wiesel, S. Marreed, A. Oved, A. Yaeli and S. Shlomov, “ST-WebAgentBench: A Benchmark for Evaluating Safety and Trustworthiness in Web Agents,” *arXiv:2410.06703v1*, Oct. 9, 2024, version 1; sec. 3.3, Table 2; sec. 4.2.

³⁰ ABA, *Formal Opinion 512*; sec. II.B, p. 6.

³¹ Object Management Group, *Case Management Model and Notation (CMMN), Version 1.1*, formal/16-12-01, Dec. 2016, official specification. Pinpoints follow specification pagination; sec. 5.4.10.4, p. 46.

³² OMG, *CMMN 1.1*; sec. 5.2.2, p. 20.

³³ OMG, *CMMN 1.1*; sec. 5.4.6, pp. 32–33.

³⁴ OASIS, *Web Services—Human Task (WS-HumanTask) Specification Version 1.1*, Committee Specification 01, Aug. 17, 2010, HTML rendition. The specification designates its .doc rendition as authoritative; secs. 3.1, 4.10–4.11, 7.1.5.

³⁵ Park and Sandhu, “The UCONABC Usage Control Model”; sec. 4.2.

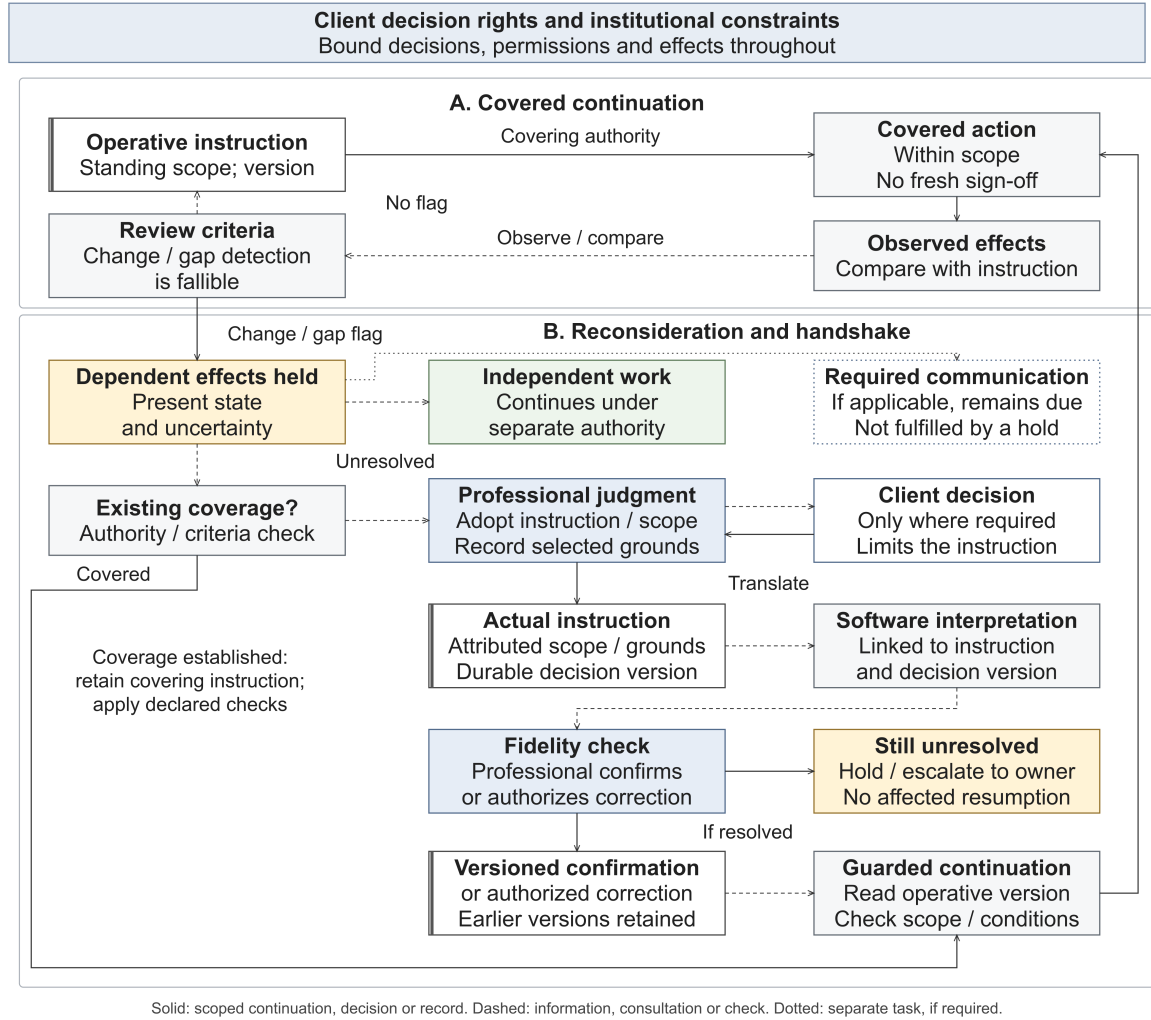


Figure 2. Hypothetical recurring professional–system co-control. The upper loop continues within a covering instruction; the lower branch holds dependent effects when coverage is unresolved. An independently covering determination can retain the instruction without fresh adoption. Where judgment is needed, the illustrated confirmation route can apply at initial translation or revision: record the actual instruction and decision version, check its interpretation, and record versioned confirmation or authorized correction before guarded continuation. Earlier versions and effects remain in history. Separately covered work retains its own authority and task status. A hold neither creates nor completes required communication, which retains its own trigger and completion conditions. Review criteria and dependency detection are fallible. This is one operationalization, not an observed prototype sequence.

4 Micro-harnesses and bounded computational contracts

The handshake would stay an aspiration if the system consumed a professional's decision in ways no one could inspect. Bounded contracts are where the decision meets the operation it governs. A contract states what a unit of computational work may do, what it relies on and what its output warrants, and it marks the point where the consumed interpretation is actually applied. A micro-harness is such a unit: a bounded piece of computational work around one responsibility invocation, with an explicit task, declared inputs, state dependencies and assumptions, permitted operations and outputs, a check suited to that responsibility, an output warrant, and defined behavior on return or

exception. "Micro" refers to scope. The unit may be a function, a workflow step or a single model invocation, and it need not be a separate agent, process or sandbox. A human procedure can take part, but an informal checklist does not by itself create a computational boundary. The pattern makes the boundary inspectable and, where feasible, enforceable. It requires no particular framework and guarantees nothing about the work inside the boundary.

Meyer frames preconditions and postconditions as caller and supplier responsibilities, distinguishes specifications from optional runtime monitoring, and separates expected special cases from assertion violations³⁶. Interface automata specify temporal input assumptions and output guarantees; optimistic compatibility requires some suitable environment, not safety in every environment³⁷. SCXML supplies guarded transitions: an unevaluable condition is false and queues an execution error³⁸. The proposed contracts apply these mechanisms to professional work, where a warranted computational result and permission for the next act are distinct.

4.1 Where the handshake meets the contract

The five roles of the handshake have counterparts in a contract. Declared permitted operations are where scope becomes operative. State dependencies and assumptions are where selected grounds are carried to the work. The check made at the effect boundary is where the consumed interpretation is read and compared with the act being attempted. Defined behavior on return or exception is where the allocated response to unresolved coverage is exercised. A harness may record an entitled act, and recording may be part of its performance under an institutional arrangement. The harness reads the resulting decision state. Neither its output nor the act of storing a record supplies the entitlement under which the decision is made. The counterpart relation shows what a contract can and cannot add. It can make the relation enforceable at a boundary, and it cannot make the decision adequate.

4.2 What a check establishes

Five relations should stay distinguishable. Declaration states an assumption or predicate. Verification checks a specified property against a stated reference, with a method whose reliability and coverage also need assessment. Enforcement connects a condition to the permitted operation, as against merely reporting it. Evaluation examines behavior across a defined population or scenario set. Professional assessment judges adequacy for the matter through the relevant expertise and authority. These are proposed analytical distinctions and not a taxonomy attributed to the sources.

A predicate in a specification is not evidence that it held in a run, and a logged check does not show that a disallowed effect was prevented. Conversely, a capability boundary may prevent an operation without assessing the quality of the reasoning that requested it. Tests can miss violations, and an erroneous reference can make a test pass for the wrong reason. A proof, a test result and an evaluator's assessment should therefore carry their different assumptions into the output warrant.

³⁶ B. Meyer, "Applying 'Design by Contract'," *Computer*, vol. 25, no. 10, pp. 40–51, Oct. 1992, doi: 10.1109/2.161279; pp. 41–44,46.

³⁷ L. de Alfaro and T. A. Henzinger, "Interface Automata," in *Proceedings of ESEC/FSE 2001*, Vienna, 2001, pp. 109–120, doi: 10.1145/503209.503226; pp. 109–113.

³⁸ W3C, *State Chart XML (SCXML): State Machine Notation for Control Abstraction*, Recommendation, Sep. 1, 2015, dated edition; secs. 3.5,5.9.1.

Meyer's optional assertion-monitoring levels already make the distinction between a declared condition and an active check explicit³⁹.

For a model invocation, a mechanically checked postcondition might establish that a response parses or that a quoted span occurs in the supplied document. It does not establish the truth of the response, or the span's support for a legal proposition. Some stronger properties may be formally checkable under a sufficiently specified task, and they should be claimed at that scope. A model-generated assertion of correctness is not itself such a check. The output should preserve relevant uncertainty, including unresolved interpretation and incomplete evidence coverage, instead of turning a qualified finding into an unqualified value for the next step.

The task boundary must also distinguish access, action and observed effect. Permission to read a record is different from permission to transmit a draft. A mechanism may restrict the latter, while a later check asks what version actually reached which recipient. Detection after transmission cannot undo it. Where effect status is unknown, the contract should preserve that uncertainty and not treat an absent acknowledgment as evidence that nothing happened.

A false guard does not by itself mean that work is safely held, because its effect depends on the transition it gates. Failure to evaluate a permission-to-proceed condition may prevent that transition. Failure to evaluate a condition for entering a hold may leave the hold transition untaken. The proposed design must specify how uncertainty blocks the affected consequential effect and how an evaluation error reaches the responsible actor, while allowing separately authorized work where appropriate. That is an application responsibility and not a guarantee supplied by the guard language or an error queue⁴⁰.

4.3 A computation that is right under the wrong rule

The revised-terms matter supplies a small case. Suppose the revised terms state that the offer lapses a stated number of days after the counterparty's counsel receives it, and a harness computes the lapse date. The counting convention, the triggering event and the calendar are inputs, and the arithmetic is what the computation adds. If the harness takes the date the offer was sent as the trigger when the terms count from receipt, an independent recomputation that uses the same inputs and calendar will agree with it. Two calculations that share a mistaken input may agree. The recomputation checks the arithmetic, and it does not establish that the right trigger was identified. No rule of contract or procedure is posited beyond the stipulation.

The proposed output warrant would name the inputs, the procedure and calendar version, the adoption status of the trigger and convention, unresolved assumptions and the intended use. A downstream drafting step could compare provisional dates conditional on their stated inputs while the correct trigger is still assigned for resolution. Consequential reliance on a date as the governing deadline needs the premises for that use settled under the covering instruction. If the receipt date is corrected, or a new event changes the operative deadline, dependent work must be reconsidered. The original arithmetic may remain correct for its stated inputs while its use in the current matter becomes inappropriate. This makes the conflation of computational correctness with the adequacy of premises easier to inspect. A warning in a record does not by itself make another actor read it.

³⁹ Meyer, "Applying Design by Contract"; pp. 41–44,46.

⁴⁰ W3C, *SCXML*, Sep. 1, 2015 Recommendation; secs. 3.5,3.12.2,5.9.1.

4.4 Composition and failure

A composition diagram connects components. It does not state what warrants each component's output. A proposed assessment, an adopted decision, a ranked option and an authorized selection can all pass through the same interface while carrying different consequences. Composition must preserve those distinctions in the contracts between units and not infer authority from a component's position in the workflow.

Composition requires one unit's output to meet the next unit's assumptions about timing, versions and permitted use. Interface automata formalize compatibility and combined environmental assumptions⁴¹. The design implication is to expose those obligations and examine the composed behavior. Local passes alone cannot establish a global guarantee; no automaton or compatibility proof is constructed here.

Failures can occur where locally acceptable work is joined. Two checks can share a mistaken assumption or a faulty reference, as the lapse-date recomputation shows. An updated instruction can arrive after the check and before the consequential effect. A unit can return an unknown effect status that another interprets as failure and retries. A monitor can hold authorized work unnecessarily. Each problem concerns a different relation among the contract, the environment and the next operation, and a single success field hides those differences.

Safety and progress need separate assessment. A hold can prevent an out-of-scope action and can also delay necessary work. More restrictive checks have no demonstrated monotonic effect on overall reliability: they can add translation faults, obscure useful work or create delay. Retry limits, resource budgets and deadlines bound an attempted procedure, but the adequacy of the chosen bounds remains an empirical and task-specific question. A designed return or escalation path must identify who can resolve the problem, what information they receive and what separately authorized work may continue.

4.5 Conditions for using the pattern, and for enforcement

A capable governed workflow can implement these contracts with typed steps, guards, permissions and exception handling. If it expresses and enforces the same relations, micro-harness terminology names a pattern and not a distinct mechanism. Nor does every task justify decomposition. Final professional review may suffice where the relevant errors are observable in the delivered artifact, intermediate results are not reused, and checking the work there is at least as effective and economical as added handoffs. These are candidate sufficiency conditions, not findings about a particular task class. The pattern earns its cost through clearer warrants, useful checks or better allocation of attention under specified conditions, and none of these follows from a smaller unit alone.

A record stores a decision. Enforcement needs more: a boundary at which the consumed action is actually compared with the stored decision, and conditions that make that comparison trustworthy. The following eight assumptions govern the enforcement claims in this paper.

⁴¹ de Alfaro and Henzinger, “Interface Automata”; pp. 114–116.

- **Instruction capture.** The professional's decision reaches the system in a form from which scope and any deliberately delegated discretion can be recovered.
- **Authorized commits.** Decision state changes only through an attributed act of an entitled actor or a delegation that covers the change, so a component's proposal alone cannot authorize a commit.
- **Version and concurrency.** The decision, ground and artifact versions an effect relies on are those in force when the effect commits. Preventing a stale effect needs a stated mechanism: the authoritative-state check and effect commit must be ordered together against changes, so a relevant change ordered before commitment causes refusal or a new check. An indivisible check-and-commit or a service-enforced conditional commit may supply equivalent semantics. A pre-send read may shorten the interval but leaves room for a change before the effect. A carried version identifier can support detection; it does not itself prevent the stale effect. A transaction in a local queue does not make an irreversible external effect conditional. Under a weaker arrangement a stale effect can occur, and the design must allocate who detects it and reconciles it. A later authorization or withdrawal does not undo an effect committed before it. Only changes that reach application state are covered here. Other changes in the world fall under the observation assumption.
- **Effect paths.** Consequential effects reach the world only through the guarded path and not around it.
- **Observation.** Changes that satisfy a review criterion are visible to the system within declared coverage, and an unobserved change is not treated as absent.
- **Separability and shared constraints.** Another act may continue where its own authority still covers it and the unresolved issue does not propagate to its effects. Shared constraints, disclosure, artifacts and timing must be examined for that dependence, and sharing alone does not require a blanket hold. If the issue reaches both acts, the hold must reach both, or another plan is needed.
- **Unknown effects and retry.** An effect whose status is unknown needs a declared reconciliation or safe effect-handling route. An authorized idempotent or deduplicated operation may resolve uncertainty without a blind repeat. An unexamined retry is unsafe where it may duplicate an act.
- **Internal preparation.** Preparation that stays inside the system can still be consequential, for instance by exposing client material to a tool or by shaping what a later decision rests on. A hold on transmission does not reach such effects, and where they matter they need their own scope.

These assumptions are implementation properties to test. A conversation, case engine or policy service may satisfy them; a well-formed record alone cannot. Where they fail, the decision–work relation remains available for inspection, but enforced control has not been established.

5 Matter state and continuity

A professional matter is not a sequence of prompts. It is a body of commitments, open questions and conditional permissions that persists while the people and systems acting on it change. Matter state is where the relation between a decision and the work it governs is carried across that change. Semantic continuity, as proposed here, means keeping resumed work connected to the decisions, scope and grounds that legitimately govern it, including their revision or withdrawal. It is a design objective,

and records and detectors can still miss changes or misrepresent a decision. Continuity in this sense concerns meaning: what was decided, by whom, within what scope, on what basis, and whether that basis still stands.

5.1 Persistence is not continuity

Storage persistence and crash recovery answer a different question. They concern which acknowledged writes survive a specified failure and whether a process can restart consistently, and the guarantee depends on the storage and recovery design. A persistent record says nothing about whether its content is still authoritative. A decision recovered intact may rest on a fact that has since changed. A recovered note may be a model's suggestion that no one adopted. A recovered permission may have been granted by someone whose role has ended. Conversely, a system may lose volatile context and still preserve semantic continuity, if the durable record suffices to reconstruct who decided what, and on which grounds.

The distinction also guards against a quiet failure: treating a surviving record as a binding one. An agent's suggestion must not become authoritative merely because it was written to durable state and later retrieved. A decision acquires operative force in the system through an attributed act of an entitled person, or within a covering delegation. The entitlement behind that act is independently grounded, and neither survival in storage nor repetition across sessions supplies it.

A workflow reachable from several interfaces raises the same continuity question. Centralizing its definition does not establish where the governing decision state resides. A decision entered through one interface must remain bound to the matter, actor, version and scope that govern a later action resumed through another. Shared access helps only insofar as those relations survive the change of interface.

5.2 Six record categories

Continuity depends on keeping apart several things that a single note or status flag tends to merge. These are record categories, not a required schema, and one record may serve more than one. A category concerns what state must keep distinguishable, and a computational responsibility concerns how an output is warranted. The correspondence is many-to-many. Verification, for example, can yield both a verification result and a change in task status, while an adopted instruction can rest on a computation, a decision-model ordering and a professional's selection.

Observation or evidence concerns what was found, from which source, and when. *Proposed reasoning or ranking* concerns what the system or a person suggests follows, and how options compare. *Adopted decision or instruction* concerns what an entitled person selected or directed, with scope and attribution. *Task status* concerns where the work stands: pending, under way, blocked or finished. *Execution permission* concerns whether a particular consequential act may now occur, which may derive from a decision, a standing delegation or both. *Verification result* concerns which property of which artifact version was tested, and with what outcome.

Merging them creates identifiable failure possibilities. If status "complete" is read as "approved," a finished draft appears authorized. If a ranking is read as a decision, advice acquires force. If a verification result is not tied to an artifact version, it appears to cover a later revision. The

distinctions help identify which check or responsible actor should examine each failure. They do not guarantee that every error is detected.

5.3 Credit to clinical precedents

Clinical decision-support research developed explicit plan and decision state decades before contemporary language-model agents, and the account builds on that work. PROforma's operational semantics records an external agent's commitment of a candidate, for example a user's, inside decision state, with an engine that propagates the consequences of that commitment and selects continuing actions by preconditions⁴². Asbru represents plan intentions as temporally extended goals and lets start preferences choose automatic activation or user approval, with conditions governing suspension, restart and abort⁴³. These works are precedents for integrating selected decisions and continuation conditions with execution. They inform the proposed synthesis, while transfer to legal-professional authority needs its own argument and supporting rules. PROforma's public operations show how an external caller can change engine state. They do not establish crash durability, nor that the caller was a professionally authorized actor⁴⁴. The inspected Asbru passages specify plan intentions and do not establish that an executing professional made a fresh, attributed decision.

Adjustable-autonomy research supplies the transfer-cost precedent. Scerri, Pynadath and Tambe describe strategies containing multiple control transfers and show that waiting for a human can carry costs, with rigid timeouts and indefinite waiting producing failures in their system⁴⁵. The article concerns its own application and does not support any professional restriction proposed here. What it does is make waiting and fallback consequences explicit. The continuity account should weigh those costs alongside applicable authority constraints and should not assume that additional holds always improve safety.

5.4 What the record carries

The matter of the handshake chapter can be carried by a modest record, shown here at the point after A's instruction D0 has been translated into the interpretation I1. The rows are illustrative and do not form a complete schema. They draw across the six categories above and not beside them.

Table 4. Illustrative record for the revised-terms matter.

Row	Content in the revised-terms matter	What it keeps distinguishable
Actor and act	A's instruction D0, attributed to A under the institutional arrangement.	An entitled act from advice, from a retrieved instruction and from a role label.

⁴² D. R. Sutton and J. Fox, "The Syntax and Semantics of the PROforma Guideline Modeling Language," *Journal of the American Medical Informatics Association*, vol. 10, no. 5, pp. 433–443, Sep./Oct. 2003, doi: 10.1197/jamia.M1264; pp. 435–436, 438–440, 442.

⁴³ S. Miksch, Y. Shahar and P. Johnson, "Asbru: A Task-Specific, Intention-Based, and Time-Oriented Language for Representing Skeletal Plans," in *7th Workshop on Knowledge Engineering: Methods & Languages (KEML-97)*, Milton Keynes, 1997, author copy. Pinpoints use the cited copy's pagination; secs. 3.2.1, 3.3.1–3.3.3.

⁴⁴ Sutton and Fox, "The Syntax and Semantics of PROforma"; pp. 439–440, 442.

⁴⁵ Scerri, Pynadath and Tambe, "Towards Adjustable Autonomy"; secs. 2.2–3, pp. 179–181.

Row	Content in the revised-terms matter	What it keeps distinguishable
Actual instruction and decision version	D0 as issued, with its version.	What the professional actually said from what the system later made of it.
Scope	Preparation of a response for internal review. Transmission, acceptance and representation of C's assent are not covered.	Permission to prepare from permission to send.
Selected grounds and what their change means	G0 entries by type: T0 source version, C0, scoped legal interpretation, provisional assumption with owner, review conditions.	Changes within delegated discretion from changes that meet a declared reconsideration condition.
Operative interpretation and version	I1 (the consumed guard), linked to D0's version, with its confirmation status.	The saved instruction from the guard actually applied.
Checker condition	Whether the translation or source check is available, what it last checked and against which version.	The state of checking from the status of the act.
Artifact and effect reference	Draft V1; recipient R; effect status (none, attempted, delivered, unknown).	An attempt from a delivery and from an unknown outcome.
Pending task	O0: owner, recipient and content, due condition, status with its evidence.	A required communication from a held transmission.

The rows differ in kind, and a reader should not take one schema from them. The actor, act, actual instruction and scope are asserted by an entitled actor. The selected grounds are an adopted selection, not the full reasons. The operative interpretation, checker condition and effect status are application state, the last two reflecting the system's own assessment or observation. O0 is configuration, and its status is application state. Which of these a standard case model supplies and which an application must configure is a question for the conventional mapping.

The record exposes the restart failure by keeping Q1 as a proposal, “ready” as task status and D0 as an instruction excluding transmission. I1 nevertheless software-allows sending. Neither storage success nor a ready label makes that allowance faithful to D0. A determination that revised wording fits the preparation scope still would not authorize transmission. O0 retains its own owner, timing and completion evidence; recording another send does not complete it. The record therefore needs no seventh category to expose these failures.

5.5 Resumption at the effect boundary

Resumption is where the record is consulted. The check that matters is made at the effect boundary, against the decision, artifact, scope, check and effect conditions the act actually consumes, and under

the enforcement assumptions of the previous chapter. The rows below follow attempted transmissions after a later authorization D2. Only the first row passes.

Table 5. Hypothetical resumption checks under the stated instruction, version, observation and effect assumptions.

Resumption attempt	What is checked at the boundary	Outcome
Transmit V2 to R under D2.	D2 names V2 and R and is current. The interpretation in force has been confirmed against D2. Known changes relevant under the declared review regime have a settled relation to the act. Only changes the system could observe are in play, so omitted dependencies and unobserved changes remain a limitation. The check result is bound to V2. No earlier transmission has unknown status.	Continues, under the stated assumptions and observation reach.
The confirmation on file refers to an earlier interpretation version.	The confirmation's version reference against the interpretation now in force.	Stale confirmation: unconfirmed unless an adequate version-bound confirmation or independently covering validation establishes the current interpretation.
A non-monetary term in T1 is observed, and its relation to C0 and to D2's conditions is unsettled.	Whether the changed ground is within discretion, covered by a rule or renewed.	Transmission under D2 is unconfirmed. Acceptance of T1 is outside D0 and D2 and stays inapplicable without separately covering authority. Preparation continues only within its covering scope and conditions.
The attempt would transmit V3 although D2 names V2, or would transmit to a recipient other than R.	The artifact version and recipient against what D2 names.	Wrong artifact or recipient: inapplicable until a covering authority exists.
D2 is withdrawn or replaced between the check and the effect.	Whether the authoritative-state check and external effect commitment are ordered against the change, as required by the version/concurrency assumption in the preceding chapter.	A stale effect is prevented only if that assumption holds, including through an equivalent service-enforced conditional commit. A pre-send read or carried identifier alone does not suffice. Otherwise, a stale effect may occur and needs detection and reconciliation.
An earlier attempt to transmit V2 timed out with no acknowledgment.	Whether the prior effect is known.	Unknown effect: use the declared reconciliation or authorized safe effect-handling route, and do not repeat blindly.

A passing check establishes only the named relation under the stated assumptions. It does not establish that the decision was correct or adequate, that the authority behind it was sound, or that the professional's judgment met the standard the matter required. Checker availability sits beside these outcomes and does not replace them, and each row leaves OO's separate completion status to be assessed against its own evidence. A transmission check cannot establish that the notification task was fulfilled or missed.

The third row repays a closer look. D2 makes sending V2 to R legitimately available, and the new term still leaves open whether D2's conditions cover the transmission now. A resource being available is not the same as the instruction continuing to cover its use. A capable policy or workflow could express that same condition, so the point is the relation and not an exclusive mechanism.

5.6 Design consequences

First, scope-bearing decision state should carry scope, attribution and selected grounds by reference, so that a detected change can be traced to the decisions and permissions it may affect. Actual grounds are the reasons and constraints on which the judgment depends, and recorded grounds are a fallible selection of those dependencies. An omitted reason may still matter, and a recorded reason may have been misunderstood. A review criterion translates selected dependencies into a condition for checking continuation, such as a source revision, a changed client instruction, expiry or a conflicting observation. The entitled actor or institutional arrangement must identify who may adopt or revise that criterion and what interpretation is delegated. Resumption checks compare available information with these records and criteria, and they cannot prove that all dependencies are known or that the actual reasons remain valid. Where coverage is unresolved, dependent consequential effects are held while separately authorized work continues.

Second, invalidation should be narrow where narrowness is available. A material change can hold consequential effects whose covering decision depends on the changed ground without stopping other work that rests on separate authorization. The breadth of a hold should follow the affected dependencies: a shared constraint may suspend several activities or the whole matter, while an isolated change need not. Selective resumption needs two things: the held and continuing work must be separable, and an independently covering authority must stand behind the continuing work. Internal preparation can itself be consequential, and disclosure inside the system can matter, so separability has to be assessed for effects as well as for tasks. Nor does every change need fresh manual approval. Legitimate pre-authorization is part of the model. A professional may decide in advance how a class of changes is to be handled, and that decision then governs those changes. The test is whether the pre-authorization actually covers the change, not whether a human reviewed it again.

Third, resumption should reconstruct from durable state what is relevant to the next act, and not load the whole matter into model context. Selecting what to retrieve is itself a design choice with failure modes, and a retrieved summary may omit the ground that matters. Represented dependencies should therefore be inspectable. A resuming actor can then identify recorded requirements for renewed confirmation before a consequential step, while recognizing that incomplete dependency records limit that check.

5.7 A conventional counterpart

A capable case or workflow system can carry these relations through human tasks, case roles and sentries^{46,47,48}. The precedents chapter gives the full mapping, including versioning, required tasks, lifecycle and current coverage. Here the key distinction is between a reserved or out-of-scope proposal, which needs an entitled actor's adoption, and a selection already covered by a valid standing instruction or delegation. The latter needs no fresh adoption. Completing an adoption task records the act; it supplies no independent entitlement. The mapping is conceptual, not a configured or tested system.

5.8 Burden and limits

The record costs effort to maintain and can widen confidentiality exposure or misrepresent the judgment it records. Omitted grounds may go unnoticed, while a populated record encourages false assurance or ritual compliance. These risks must be weighed against the decisions the distinctions clarify. The record carries an independently authorized act and its consequences; its existence alone establishes neither entitlement, adequate judgment nor fulfilled duties. Explanatory value is distinct from a measured reduction in error or attention.

Figure 3 separates semantic dependence from temporal order. A detected ground change challenges the decisions and actions that depend on it, and a separately covered branch within the same matter is shown for contrast. Selected distributed-cognition, joint-activity and durable-workflow comparisons come later in the paper. Broader literature reception, implementation-specific guarantees and empirical assessment of continuity remain open.

⁴⁶ OMG, *CMMN 1.1*; sec. 5.4.10.4, p. 46.

⁴⁷ OMG, *CMMN 1.1*; sec. 5.2.2, p. 20.

⁴⁸ OMG, *CMMN 1.1*; sec. 5.4.6, pp. 32–33.

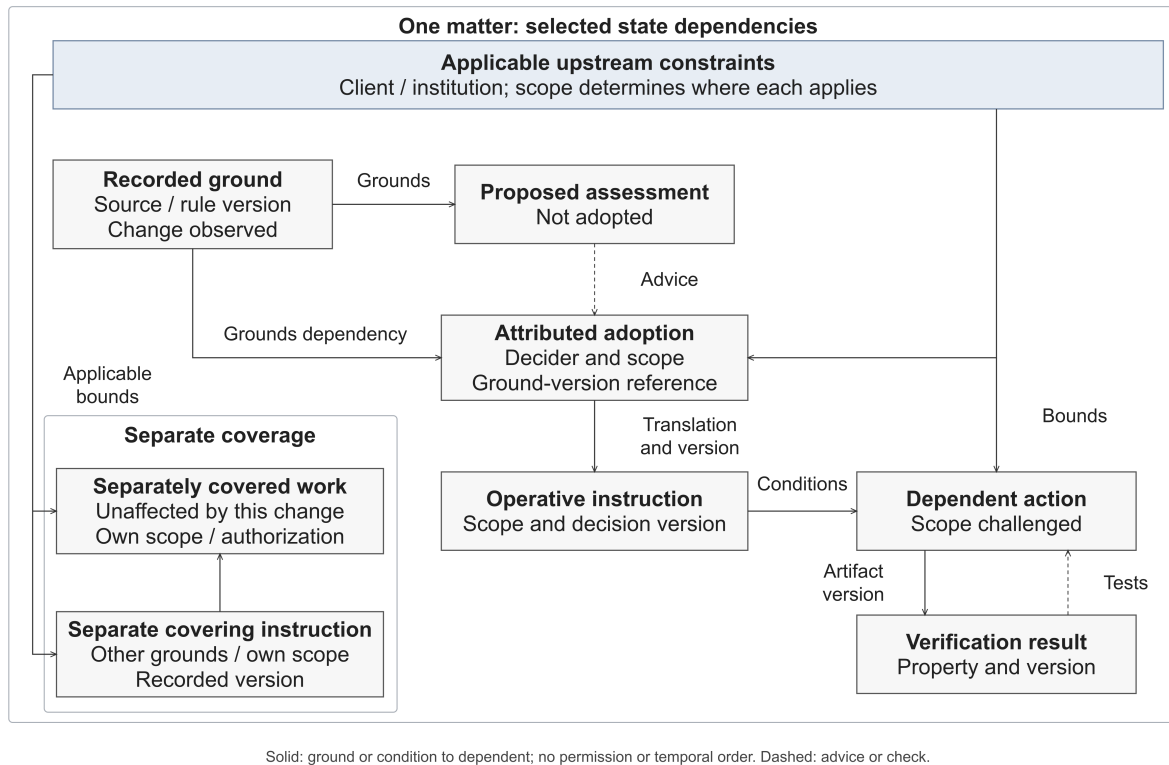


Figure 3. Selected matter-state dependencies in one matter. Solid edges connect a ground or condition to its dependent; they do not grant permission or show temporal order. Dashed edges carry advice or a check. A proposal may inform adoption but cannot adopt itself; translation into an operative instruction is fallible. The check binds a named property to an artifact version. The separately covered branch does not depend on the particular changed ground illustrated. It remains subject to its own scope, covering instruction, recorded version and applicable upstream constraints. A change in a shared constraint may affect both branches. Recorded dependencies are selected and may omit relevant reasons.

6 Professional rights, duties and succession

Professional-in-the-loop requires more than an approver's name. Professional status, competence for the matter and authority over the particular decision are separate questions. An admitted lawyer may need further competence for a task; a competent lawyer may lack authority over a decision reserved to the client. Supporting work may be performed by someone without professional status under supervision. The architecture must identify the conditions under which the relevant person can exercise judgment, rather than infer them from a role label.

Professional rights and duties constrain how computational outputs are adopted, maintained and renewed (Table 1). Each argument connects a selected right or duty to a foreseeable allocation failure and a design response. The response must preserve the authority and information judgment requires; doctrine does not prescribe a record schema.

6.1 Why New York, and what status each source has

The argument needs adopted rule text that a reader can check clause by clause. The New York Rules of Professional Conduct, in the court-hosted compilation, supply it. The compilation keeps adopted

rule text apart from the NYSBA comments that accompany it, and those comments are not adopted rules⁴⁹. New York is a scoped illustration. It is not claimed to govern the constructed matter, and other jurisdictions, federal rules and tribunal rules raise separate questions. Substituting the ABA Model Rules would change the adopted text on which each step rests and would need its own derivation.

The advisory opinions have a different status. ABA Formal Opinion 512 is advisory interpretation of the Model Rules⁵⁰. New York City Bar Formal Opinion 2024-5 is advisory guidance from a bar committee. It says it follows the format of the California State Bar's practical guidance, quotes that guidance with attribution, and describes itself as general guidance to be updated⁵¹. Where this chapter relies on a passage the New York City opinion takes from California, it says so. Neither opinion adds doctrine to the adopted rules, which remain the primary basis.

Three further sources are used at their own scope. *Davidson v. Metropolitan Transit Authority* is a decision of a New York intermediate appellate court and is used only for what it holds on its facts⁵². Kolt's article on governing AI agents is legal scholarship, so its statements about agency doctrine are analysis and not binding law⁵³. Green's study is a critical review of oversight policy for government algorithms⁵⁴. It supports an analogy, not a finding about attorneys.

6.2 Whose decision it is

The first argument starts from the client's right. Under New York Rule 1.2(a), a lawyer abides by a client's decision whether to settle⁵⁵. Rule 1.4 requires the communication and explanation that make a client's decisions informed⁵⁶. The NYSBA comments, which are not adopted rules, qualify both. Communicating an offer may be affected where earlier client instructions already make its acceptability clear or authorize acceptance or rejection, and the comment to Rule 1.2 addresses advance authorization, material change and revocation⁵⁷. The right is therefore not a fixed bar on delegation. It places the decision with the client unless, and to the extent that, the client has authorized otherwise.

Delegating preparation introduces several translations between C's reservation and an outgoing message: A's understanding, D0, the consumed interpretation I1 and the firm's configuration. Each

⁴⁹ N.Y. Rules of Professional Conduct; Rule 1.2 Comment 3, p. 16; Rule 1.4 Comment 2, pp. 23–24 (NYSBA guidance).

⁵⁰ ABA, *Formal Opinion 512*; p. 4.

⁵¹ New York City Bar Association Committee on Professional Ethics, *Formal Opinion 2024-5: Ethical Obligations of Lawyers and Law Firms Relating to the Use of Generative Artificial Intelligence in the Practice of Law*, 2024, official opinion. Advisory guidance; table entries on Competence and Diligence and Supervision.

⁵² *Davidson v. Metropolitan Transit Authority*, 44 A.D.3d 819 (N.Y. App. Div. 2d Dep't 2007), 2007 N.Y. Slip Op. 07814, official opinion; pp. 819–820.

⁵³ Noam Kolt, "Governing AI Agents," *Notre Dame Law Review*, vol. 101, pp. 335–386, 2026, published article; pp. 359–361.

⁵⁴ B. Green, "The flaws of policies requiring human oversight of government algorithms," *Computer Law & Security Review*, vol. 45, article 105681, 2022, doi: 10.1016/j.clsr.2022.105681; secs. 4.1,5.1.

⁵⁵ N.Y. Rules of Professional Conduct; Rule 1.2(a).

⁵⁶ N.Y. Rules of Professional Conduct; Rule 1.4.

⁵⁷ N.Y. Rules of Professional Conduct; Rule 1.2 Comment 3, p. 16; Rule 1.4 Comment 2, pp. 23–24 (NYSBA guidance).

may broaden the constraint. Neither the attorney's technical ability nor the configuration enlarges the client's right⁵⁸. The guard must therefore be checked for fidelity to the represented reservation.

The design response is to carry the client's constraint into decision state as something D0 depends on and cannot enlarge. C0 sits among the selected grounds, D0's scope excludes transmission, and the effect boundary applies a version-bound interpretation whose fidelity to D0 is checked separately. The guard stays fallible, and its allowance alone does not establish coverage.

A client's constraint will often reach the system through A's account of a conversation, not through anything the client wrote. The record should therefore keep three things apart: the attribution (A's representation), the asserted source (C) and the status of the communication. Marking a constraint as coming from the client shows neither that it was captured faithfully nor that C was informed.

Suppose the system then observes a client communication that meets a declared review criterion before A has adopted any reading of it. If coverage is unresolved, the effects that depended on the unsettled condition become unconfirmed. Acceptance and transmission, which D0 already excludes, stay inapplicable under D0. Preparation continues only within its own covering scope and conditions. The observation does not authenticate a revocation or fix a new scope. Those questions go to the designated actor and to the applicable communication checks.

Advance authorization can cover a revised offer, so a record can also misclassify a permitted decision as reserved. The professional and client must interpret the applicable instructions. The architecture neither converts every preference into a command nor permits conduct the law forbids.

What the design can do is allocate a route. The question goes to an actor the institution has designated. It stays visible as unresolved while it waits. The attempts to reach that actor, and any receipt, remain inspectable. A named owner is responsible for the response, or for the escalation if no response comes. None of this guarantees that the actor is available, that the message arrives or is read, or that the recipient understands it. A recorded reply likewise does not show that the client has authorized anything.

6.2.1 Five questions that must stay apart

Several things that look alike inside a workflow are different questions with different sources. The first argument depends on keeping them apart.

Table 6. Professional rights, institutional arrangements, software allowance and legal effect have distinct sources.

Question	Where the answer comes from	What the architecture can do with it	What it cannot establish
What has the client reserved or authorized?	The client relationship and the applicable rules, for example settlement under Rule 1.2(a) ⁵⁹	Carry it as an upstream constraint with its attributed source, and refuse an instruction that exceeds it	That the client was informed, or that the constraint was captured faithfully

⁵⁸ N.Y. Rules of Professional Conduct; Rule 1.2(a).

⁵⁹ N.Y. Rules of Professional Conduct; Rule 1.2(a).

Question	Where the answer comes from	What the architecture can do with it	What it cannot establish
What does the firm or institution require or delegate?	The institution's adopted policies, standing ranges and supervision arrangements	Enforce a represented constraint and bind it to the matter decisions it affects	Client or legal entitlement merely from a policy label; a legitimate institutional appointment may supply scoped standing but cannot enlarge what the client has reserved
Who is entitled to decide this act?	Professional and client-relationship sources outside the workflow	Recognize it, require an attributed act and record the act	Create it; a role label or stored credential cannot appoint anyone
What does the software allow?	The operative guard the system actually consumes	Enforce it at the effect boundary and compare it with the decision	That the act is permitted, effective or adequate
What legal effect does the act have?	Law applied to the facts, outside the system	Preserve what occurred and what follows from it	Be established by the record or allowance alone

The last two rows can come apart on real facts, and one New York decision shows how. In *Davidson*, the plaintiff's attorney had, on the court's account, been involved at length in the case, taken part in settlement negotiations and appeared at pretrial conferences. The Appellate Division held that the attorney had apparent authority to enter a stipulation of settlement. It stated that a stipulation made by an attorney may bind a client even if it exceeds the attorney's actual authority, provided the attorney had apparent authority. The client's contention that her attorney lacked authority to bind her did not prevail⁶⁰.

The holding concerns apparent authority on those facts, not software or every out-of-scope act. South and colleagues likewise leave the adaptation of agency doctrine to autonomous or self-modifying agents uncertain⁶¹.

Under the conditional rule illustrated by *Davidson*, disavowal after an out-of-scope act may come too late. The design implication is to apply the reservation before the relevant effect, under the enforcement assumptions in §4.5. The record should also preserve what was attempted and what resulted, whether or not the act was permitted.

Agency analysis also separates authority from capacity. Kolt, quoting Frankel, puts the point as the purposes for which a fiduciary may use delegated power being narrower than the purposes for which the fiduciary is capable of using it⁶². For software, what the system can do says nothing about what it may do. The allowance the guard enforces is a translation of a decision. It is not evidence that the decision was made.

⁶⁰ *Davidson*, 44 A.D.3d 819; pp. 819–820.

⁶¹ T. South, S. Marro, T. Hardjono, R. Mahari, C. D. Whitney, A. Chan and A. Pentland, “Position: AI Agents Need Authenticated Delegation,” in *Proceedings of the 42nd International Conference on Machine Learning*, PMLR, vol. 267, pp. 82211–82231, 2025, publisher article; secs. 3.2,4.1,5.3.

⁶² Kolt, “Governing AI Agents”; pp. 359–361.

Kolt treats ambiguity as sometimes deliberate: interpreting an instruction may require discretion or clarification, at a cost to delegation⁶³. D0 leaves wording, structure and research open within preparation. The architectural question is which unresolved interpretations return for judgment, under the applicable authority, and at what interruption cost.

6.3 Judgment that can actually be exercised

The second argument concerns what the professional brings to the boundary. Competent representation requires the knowledge, skill, thoroughness and preparation reasonably necessary, and a lawyer may not handle a matter the lawyer knows or should know is beyond the lawyer's competence without associating competent counsel⁶⁴. Rule 2.1 requires independent professional judgment and candid advice⁶⁵. The New York City Bar opinion stresses trained legal judgment in competent tool use. Drawing on California guidance, it treats professional judgment as the lawyer's continuing responsibility and warns against overreliance that weakens critical analysis⁶⁶.

The failure this makes foreseeable is the one meaningful-control work names: a person is given final authority by design without the ability to exercise it, like a manager who merely signs off on a recommendation⁶⁷. An approval event can be preserved while competence, information, time or the practical ability to challenge is missing, and the record looks the same either way.

Green's critical review of government-algorithm oversight policies challenges the assumption that assigning human review supplies effective protection. It draws on research about judging algorithmic quality, automation bias and the limits of explanations, and it treats the upper bound of improved oversight as an empirical question⁶⁸. Applied here by analogy, the challenge calls for task-specific evidence of practical capacity. An approval step records participation. Its design must also provide the information, time and means of challenge that judgment needs.

Read this way, "trained judgment" becomes a set of conditions that must all hold. They are the paper's design reading of what the rules and the advisory opinion ask of an architecture. They are not a legal test.

Competence concerns who receives the question. A design cannot establish a person's competence for a matter from a role label or credential. It can route a boundary to someone the institution has assigned to that kind of work, make the routing and its basis inspectable, and treat the absence of such a person as an unresolved allocation, not as approval by default⁶⁹.

Information concerns what the boundary presents. It includes the instruction and its scope, the selected grounds with their versions and what each change would mean, the eligible alternatives and trade-offs where a decision model has ordered them, the state of checking, and what the system could not observe. An order alone can carry a weighting that no one with authority chose, which is why a decision model's output may take the form of an acceptable set with its trade-offs.

⁶³ Kolt, "Governing AI Agents"; pp. 359–361.

⁶⁴ N.Y. Rules of Professional Conduct; Rule 1.1(a)–(b).

⁶⁵ N.Y. Rules of Professional Conduct; Rule 2.1.

⁶⁶ NYC Bar, *Formal Opinion 2024-5*; table entries on Competence and Diligence and Supervision.

⁶⁷ Cavalcante Siebert et al., "Meaningful human control: actionable properties"; pp. 248–249.

⁶⁸ Green, "The flaws of policies requiring human oversight"; secs. 4.1, 5.1.

⁶⁹ N.Y. Rules of Professional Conduct; Rule 1.1(a)–(b).

Time concerns queue and consequence. A professional shown a boundary after the deadline it protected has not been given the means to decide. Boundaries also compete for finite attention, so each added boundary has to justify its cost. A system that asks for confirmation of routine acts while a consequential change goes unremarked can accumulate many approval events and little judgment.

Challenge is the practical ability to reject, redirect, correct the operative interpretation or ask for more. It includes freedom to depart from a recommendation, which can reflect permissible discretion among eligible alternatives and not correction of a model error⁷⁰.

A fifth question, who takes over when the responsible person cannot act, is taken up under succession below.

Shared-control work gives these conditions a general form. Ability includes competence, skills and resources such as time, tools and personnel. The proposed design relation is that responsibility should not exceed authority, authority should not exceed ability, and control should be exercised by a partner with sufficient ability and authority⁷¹. A professional's boundary is a place where that relation can be checked. An assignment that gives someone responsibility for a decision without the time and information to make it is inconsistent in just that sense.

Assessing these conditions takes more than an access or assent event. Supervision must examine the information and opportunity behind the recorded act. Grounds entry, training, translation review and reconciliation impose costs. A successful recommendation may still omit an option, apply an unsuitable weighting or reach an unavailable or unauthorized decider; the selection it informs must remain separately attributed.

6.4 Supervision across people and organizations

Responsibility is not exhausted by one attorney's interaction with the system. Rule 5.1(a)–(b) places reasonable-effort responsibilities on firms, managers and supervisors for lawyers' conformity to the rules⁷². Rule 5.3(a) requires appropriate supervision of nonlawyer work, calibrated to circumstances including experience, the work involved and likely ethical problems⁷³.

The advisory literature treats organizational measures as part of responsible tool use. ABA Opinion 512 addresses firm policies on permissible use and on training lawyers and nonlawyers⁷⁴. The New York City opinion quotes California guidance that managerial and supervisory lawyers should establish clear policies on permissible uses and make reasonable efforts to ensure conduct complies when generative AI is used, including through training⁷⁵. These are selected New York rule descriptions and advisory statements, not a claim of uniform national law. They do not establish that a model is a nonlawyer in the sense of Rule 5.3 or settle how a particular vendor is treated. Technology-specific applicability needs its own authority and facts.

⁷⁰ N.Y. Rules of Professional Conduct; Rule 2.1.

⁷¹ Flemisch et al., “Towards a dynamic balance”; pp. 6–7, 10–11.

⁷² N.Y. Rules of Professional Conduct; Rule 5.1(a)–(b).

⁷³ N.Y. Rules of Professional Conduct; Rule 5.3(a).

⁷⁴ ABA, *Formal Opinion 512*; sec. II.E, p. 10.

⁷⁵ NYC Bar, *Formal Opinion 2024-5*; table entries on Competence and Diligence and Supervision.

Institutional choices can disappear behind the last approval: tasks assigned to the system, standing ranges, review criteria and supervision. A matter approval does not assess whether those choices were appropriate. Green argues that oversight in government can legitimize adoption while reducing deployer accountability⁷⁶. The analogous design question is who reviews the configuration, rather than attributing every failure to the final approver.

The design response is to treat institutional arrangements as visible, versioned inputs. A reviewer should be able to ask whether the arrangement was adequate, and not only whether the last act was approved. The relevant inputs are the standing range, the review regime, the detector and its criteria, and the allocation of takeover. A policy block can enforce a represented constraint, but it cannot settle entitlement. Neither firm configuration nor the lawyer's instruction can enlarge the decision right reserved to the client⁷⁷. Responsibility in shared control is assigned beforehand and evaluated afterward by others⁷⁸, and the record is part of what makes that later evaluation possible.

Supervision also has to account for who occupies which position. Baum and Laux distinguish the operative actor from the overseer even where one person occupies both roles, and they propose layered oversight and independent benchmarks to mitigate compromised critical distance⁷⁹. They add that capacities, information and authority are needed for a causal position to be meaningful⁸⁰, and their human-overseer definition requires systematic preparation, genuine capacity and a risk-reduction function⁸¹. A professional who recurs in the workflow is not thereby an adequate independent reviewer of her own classification of a boundary or of her own instruction. The design should look for review from a position with some critical distance. The ABA opinion allows less independent verification for some tools and tasks where prior testing or experience supplies a reasonable basis for reliance⁸². That concerns the review effort a use justifies. It is not permission to bypass a reserved decision.

Kolt describes incomplete observability, the cost of monitoring at speed and scale, and the inadequacy of monitoring alone⁸³. These are scholarly arguments, not attorney-workflow measurements. They support declaring supervision's observation reach: a quiet detector does not establish continuing coverage.

6.5 Required conduct and the cost of holding

Boundaries on what the system may do are only half of what the rules ask. Rule 1.3 requires reasonable diligence and promptness and prohibits neglect⁸⁴, and Rule 1.4 requires communication⁸⁵. A permission model sees prohibited acts. It does not by itself see an act that is required and has not

⁷⁶ Green, “The flaws of policies requiring human oversight”; secs. 4.1,5.1.

⁷⁷ N.Y. Rules of Professional Conduct; Rule 1.2(a).

⁷⁸ Flemisch et al., “Towards a dynamic balance”; p. 7.

⁷⁹ K. Baum and J. Laux, “Constitutive vs. Corrective: A Causal Taxonomy of Human Runtime Involvement in AI Systems,” *arXiv:2603.19213v2*, Aug. 20, 2026, version 2; sec. 4.2.

⁸⁰ Baum and Laux, “Constitutive vs. Corrective”, *arXiv:2603.19213v2*; sec. 1, final qualification.

⁸¹ Baum and Laux, “Constitutive vs. Corrective”, *arXiv:2603.19213v2*; sec. 4.1, Def. 3 and normative qualification.

⁸² ABA, *Formal Opinion 512*; p. 4.

⁸³ Kolt, “Governing AI Agents”; pp. 370–371.

⁸⁴ N.Y. Rules of Professional Conduct; Rule 1.3(a)–(b).

⁸⁵ N.Y. Rules of Professional Conduct; Rule 1.4.

been done. In the revised-terms matter, holding the transmission averts an unauthorized act. It does not notify A of the revised terms, and the stipulated task O0 remains due until its performance is established.

An unavailable professional and an approaching deadline sharpen the problem, because a hold alone does not solve it. A configuration can expose the matter to delay inconsistent with diligence, and to action beyond the client's instruction, and adopting a hold policy does not settle that tension. It also gives no agent permission to act beyond existing authority. Coverage by another qualified professional, separately authorized preparation or a permissible standing instruction may allow work to continue. Whether a protective action is legally available depends on the actual context.

The design response treats required conduct as its own state. A required task keeps its owner, due condition, performance evidence and escalation path, recorded independently of any hold and restored on recovery. Separately covered work continues under its own authority. A pre-authorized protective fallback is an option only where the applicable authority and delegation permit it, and the architecture creates no such permission. The system should also state how a pending boundary reaches a suitable actor's attention and which time-sensitive consequences are visible.

Two limits apply. O0 is stipulated by the hypothetical arrangement and is not derived from a professional rule. A separate duty to communicate with the client, when discussed, rests on selected New York authority with the qualifications already noted^{86,87}. The workflow can show that a notification was delivered, to whom and when. It cannot show that the recipient understood it, that it was adequate, or that a professional duty was fulfilled. Nor does the chapter propose an engine that derives the duties applicable to a matter. The applicable set comes from authority and facts outside the workflow. The design aims only to keep what an entitled actor has stated from disappearing behind a hold.

6.6 Succession and multiple sources of instruction

When the responsible professional is unavailable or reassigned, three questions come apart. The first is who can perform covered work. The second is which earlier decision and interpretation still govern. The third is who is entitled to renew, revise or reaffirm the decision and to own what is pending.

In the matter, A's unavailability does not give B authority automatically. B may be a competent alternate, but B's authority over this matter has to be established independently. B may be able to carry forward preparation that D0 covers. D0 and its current interpretation govern while their scope and conditions remain applicable, and B's entitlement to renew or revise depends on arrangements outside the workflow. O0 stays pending until its stipulated performance is established, and an entitled reassignment changes its owner and response path without completing it. A departing professional's departure can end that professional's entitlement to decide further without ending the force of what was decided, and neither fact confers entitlement on a successor.

⁸⁶ N.Y. Rules of Professional Conduct; Rule 1.4.

⁸⁷ N.Y. Rules of Professional Conduct; Rule 1.2 Comment 3, p. 16; Rule 1.4 Comment 2, pp. 23–24 (NYSBA guidance).

The interchangeable-approver assumption overlooks authority, context and time. A role label does not establish an appointment or attention to the matter. Assigning a pending decision without the ability to exercise judgment violates the shared-control consistency relation⁸⁸. A successor may also need to reconsider the earlier classification of the boundary; adoption alone did not establish its adequacy.

Matters can also involve several sources of instruction at once. Kolt, reporting on the Restatement (Third) of Agency, notes that agents may appoint subagents and act for coprincipals, that the legal relationships may be ambiguous, and that the same actor may occupy different roles at successive points in an ongoing interaction⁸⁹. For AI agents he raises open questions about when an agent should be authorized to appoint a subagent and how subagents should resolve conflicts⁹⁰.

The reading drawn here is that no universal precedence among client, professional and institutional instructions can be assumed, and the architecture should not hard-code one. Where applicability or precedence is unresolved and affects continuation, the question is routed to a suitably entitled actor. The software must not invent authority by averaging preferences. Escalation may end with no currently authorized continuation of the requested act, and that result should be represented as such and not resolved by silently choosing one of the constraints. Within the architecture, decision state changes only through an attributed act of an entitled actor or a delegation that covers the change, so a harness that invokes another component does not thereby enlarge what that component may do.

A conventional professional workflow could meet this account through supervised preparation, accountable review, client instructions and appropriate continuity controls. Where it does, the proposed vocabulary can still help say whether a failure concerned evidence, interpretation, authority, execution or supervision. Diagnostic organization has direct precedents in the anticipatory-oversight failure modes and the governed-agency loss-source classification^{91,92}, so its explanatory value has to be argued and assessed independently of mechanism novelty. The record represents judgment and its operative scope imperfectly. Neither its presence nor its absence alone establishes whether a professional duty was fulfilled.

7 Coordination, fidelity and practical capacity

A matter is worked on by a lawyer, a client, colleagues, a calendar, a set of drafts and a system, and none of them holds the whole of it. Hollan, Hutchins and Kirsh supply the framing: a unit of analysis that spans people and artifacts, in which cognition is coordinated through representations and distributed through time⁹³. In that frame a professional decision is a representation that others rely on after its author has moved on. The question for the architecture is what happens to its meaning as it passes through the people and artifacts that depend on it.

⁸⁸ Flemisch et al., “Towards a dynamic balance”; pp. 6–7, 10–11.

⁸⁹ Kolt, “Governing AI Agents”; pp. 365–366.

⁹⁰ Kolt, “Governing AI Agents”; pp. 365–366.

⁹¹ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; sec. 4.3, pp. 24–25.

⁹² Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; sec. 4.10, p. 36.

⁹³ J. Hollan, E. Hutchins and D. Kirsh, “Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research,” *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 2, pp. 174–196, June 2000, doi: 10.1145/353485.353487. Cited author final revision, Apr. 23, 2000; revision pagination differs from the published article; secs. 1, 2.1, revision pp. 2–4.

Two limits keep the frame from being asked to do too much. First, distributed cognition explains how understanding is spread across a system. It does not say who may decide. Participants can reason well together and still include no one entitled to commit the client, and a well-coordinated record represents a decision without conferring authority to make it. That entitlement comes from outside the cognitive description, as the professional chapter (§6) argues.

Second, the joint-activity account of Klein and colleagues treats coordination as depending on continuing repair, on status and intentions that others can interpret, and on the ability to direct, and it notes that coordination has costs⁹⁴. The same account supplies the example that matters most here: a shared display can make a change available without any participant having noticed it⁹⁵. A representation's availability and its uptake are different events.

Four relations connect a professional decision to its effects, and each has its own characteristic failure. The decision must be captured in a form from which its scope and any deliberate discretion can be recovered. It must be communicated so that the recipient can grasp what it covers and what remains uncertain. It must be interpreted into the operative constraint the system actually consumes. And what then happens must be observed and reported. A well-formed record, an understandable message, a correct constraint and an accurate account of effects are four separate achievements. Folding them into one signed approval hides which of them failed. These four recur along the exchange shown in Figure 2.

7.1 Capture and the range an instruction leaves open

Faithful capture does not mean reducing an instruction to a single predetermined action. A professional may leave options open or delegate discretion within a scope, as D0 does for structure, wording and research within preparation. Kolt observes that instructions to agents are often ambiguous, sometimes deliberately, so the agent has to interpret and exercise discretion and may in some circumstances be obliged to seek clarification, at the price of undermining some of what delegation was for⁹⁶. Horvitz's mixed-initiative principles likewise treat dialog to resolve key uncertainties as a design matter⁹⁷.

The record should therefore preserve the range, including unresolved ambiguity, and should not manufacture precision. Divergence inside the allowed range is not translation failure. Where the range itself is contested, the question is one of interpretation and possibly renewed judgment, and comparing the system's behavior with a stored restatement cannot settle what the professional meant.

7.2 Communication at the boundary

What a recipient is shown at a boundary has to let that recipient do what the boundary exists for. Acknowledgment does not establish understanding. A confirmation click expresses assent and a machine acknowledgment evidences receipt, and neither shows that scope, uncertainty and

⁹⁴ Klein et al., “Ten Challenges”; pp. 91–94.

⁹⁵ Klein et al., “Ten Challenges”; pp. 91–94.

⁹⁶ Kolt, “Governing AI Agents”; pp. 359–361.

⁹⁷ E. Horvitz, “Principles of Mixed-Initiative User Interfaces,” in *Proceedings of CHI '99*, ACM Press, 1999, pp. 159–166, doi: 10.1145/302979.303030. Cited author manuscript; manuscript pagination differs from the published article; principles 5,6,9,11.

consequences were grasped. A system's fluent restatement is likewise evidence to examine, not proof that it understood.

Coactive Design, the interdependence framework of Johnson and colleagues, carries joint-activity theory into interface requirements of observability, predictability and directability that are sufficient for the interdependencies actually present. It does not ask for maximal disclosure, because too much shared information can itself be a problem^{98,99}. Read for this chapter, it asks that a recipient can see the operative interpretation as applied and not only the instruction restated, can predict what the system will do while a question is open, and can direct the system to correct or hold. These interface requirements support coordination among interdependent participants. The professional entitlement that governs their use is established independently.

The human-AI interaction guidelines point in a similar direction in their treatment of contextual timing, correction, disambiguation and the consequences of user actions¹⁰⁰. Their authors caution that high-risk or highly regulated settings may need specialized guidelines¹⁰¹, so their application to legal work needs domain-specific assessment.

Capacity is the other half of communication. An interface can make everything observable and still hand it to someone with no time, no grounds that bear on the decision and no practical means of challenge. A confirmation is meaningful only with the capacity, information, time and ability to challenge that make it so¹⁰². Those conditions are developed with the professional duties in §6. The consequence here is that more capture and more interruptions are not presumed better. Each added handoff consumes professional attention and adds another place where translation can fail.

7.3 The consumed interpretation: where restatement stops short

The guard counterexample in the handshake chapter exposes a gap an interface can conceal: D0 is stored correctly, while I1 permits an act D0 excludes. Reading the system's fluent paraphrase compares the intended decision with that paraphrase, not with the guard as it operates. If the same model produced both translation and restatement, their agreement supplies no independent reference.

Fidelity must be assessed across the acts the system could attempt. An independently prepared action test, or inspection of the guard against D0, can expose the mismatch. This is why the operative constraint must remain recoverably linked to the actual instruction. The checker's condition is a separate dimension: a required test may be unavailable, stale or bound to another version. None of those states is a pass, nor does it change an act already known to lie outside scope.

⁹⁸ M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, M. B. van Riemsdijk and M. Sierhuis, "Coactive Design: Designing Support for Interdependence in Joint Activity," *Journal of Human-Robot Interaction*, vol. 3, no. 1, pp. 43–69, 2014, doi: 10.5898/JHRI.3.1.Johnson. Cited author-hosted copy of the journal article; pinpoints use its printed pagination; pp. 44–45.

⁹⁹ Johnson et al., "Coactive Design"; pp. 51–52.

¹⁰⁰ S. Amershi et al., "Guidelines for Human-AI Interaction," in *Proceedings of CHI 2019*, Glasgow, May 4–9, 2019, 13 pages, doi: 10.1145/3290605.3300233. Cited camera-ready copy, including its appendix; Table 1, PDF p. 3.

¹⁰¹ Amershi et al., "Guidelines for Human-AI Interaction"; sec. 7, PDF p. 11.

¹⁰² Cavalcante Siebert et al., "Meaningful human control: actionable properties"; pp. 248–249.

Translating a mandate into operative commitments is already a named failure point in governed agency¹⁰³ and anticipatory oversight¹⁰⁴. The architectural consequence here is to make that translation inspectable through the instruction, the consumed constraint and the attempted acts, rather than treat a fluent paraphrase as the evidence of fidelity.

7.4 Faithful, and still deficient

Fidelity compares the instruction as issued with the interpretation consumed. Adequacy compares the instruction with what the matter requires, and a faithful implementation can carry out an inadequate instruction. I0 implements D0 exactly, yet the workflow can still omit the independently stipulated notification O0. In another branch, an instruction authorizes assent to revised terms but omits a consultation the applicable arrangement requires. A guard can execute that instruction faithfully and still permit deficient conduct. The requirement comes from outside the instruction, and the translation test cannot find it. Such failures call for professional or supervisory assessment of the instruction and the arrangement, alongside the test of its implementation.

A later authorization such as D2 governs later acts. It does not make I1 faithful to D0, and it does not erase earlier effects or omissions.

7.5 Observed effects and what a report shows

The fourth step closes the loop. A status such as "notified" asserts that performance occurred. It is a claim to be checked against the stipulated content, recipient, delivery and timing evidence. A failed or unresolved delivery leaves the task pending, with escalation where its conditions require it. Even verified delivery does not show that the responsible attorney grasped what the revised terms affect. If the arrangement requires A's understanding or response, that is a further task with its own evidence. A report of effects that names the artifact version and recipient actually used, and that distinguishes an attempt from a delivery and from an unknown outcome, gives the next participant something to coordinate around.

7.6 The simpler design, and the rival

A shared status view with a professional approval step, perhaps with a typed decision field, is the simpler alternative. For a defined task whose authority and information requirements are modest, it may suffice. Existing human-in-the-loop designs can capture professional judgment where their application does so. The richer relation is motivated where decisions govern later action, grounds change, and work passes to another actor or session. Its added cost has to be justified, and whether it reduces misunderstanding relative to the simpler design is untested.

A formal protocol is one operationalization of these relations. Restatement in conversation, a task engine or a policy service could realize them. A capable governed workflow is the stronger rival. If it preserves the same relations, the mechanism is not distinctive, and what remains is whether the vocabulary of capture, communication, consumed interpretation and observed effect helps locate where coordination broke. That is a hypothesis about where coordination breaks, not a validated model of professional work. Proposition P3 and the first study in the assessment chapter test it.

¹⁰³ Dias Ferreira, "When Intelligence Becomes Agency", arXiv:2609.07741v1; sec. 4.8 (Def. 4.17).

¹⁰⁴ Baum et al., "Anticipatory Human Oversight", arXiv:2609.24242v1; sec. 4.3, pp. 24–25.

8 Durable workflows, effects and continuing authority

Durable execution is about recovering a process under stated failure assumptions: which steps survive, and how a resumed process works out where it was. Continuing authority is about whether a particular act, at the moment it is about to occur, still falls within a valid instruction or delegation. The two have different specifications, and success at the first does not establish the second. A resumed workflow can carry out an act that was permitted under facts that have since changed, unless the application observes and handles the change.

The converse also holds. A process that lost its volatile context can keep authority information intact if a separate record suffices to show what was decided and within what scope. Replay correctness and authority validity therefore need separate checks, and one application can run both in the same workflow.

CMMN, DMN and SCXML supply specification-level constructs for case state, decisions and guarded transitions. They show which recovery relations an application can express and which authority and effect semantics it must supply itself.

8.1 What a check binds to

A plan has to be distinguished from permission to carry out its effects. Inspection or approval may concern a proposal, a delegated instruction or a covering authorization, and the design must say which. Approval of a plan supplies authority for a consequential act only to the extent that the approving actor's entitlement and the adopted scope cover that act. A continuing check therefore concerns the effect about to occur: the artifact version it will carry, the recipient it will reach, and the instruction version and operative interpretation that govern it. It does not rest on the existence of an earlier plan-approval event.

What is checked has to be what is sent. A check that "V2 to R is covered" says nothing about a transmission that, through a late binding or a replaced file, carries another version or goes to another address. The design should bind the check to the identifiers the effect will actually use and carry those identifiers with the effect, so that a mismatch is detectable.

The effect boundary is specific to the operation: it is where the consequence in question commits, such as a message being transmitted or a filing being submitted. Committing an entry to a local queue is not necessarily that boundary. Before the relevant commitment, a guard may refuse the act. Once an irreversible effect has occurred, a later check cannot prevent it; any remedy or compensation has separate requirements. The enforcement conditions are stated in §4.5.

8.2 The interval between check and effect

A check made shortly before an effect leaves an interval in which things can change. A newer instruction can arrive, a draft can be replaced, or the decision can be withdrawn. Checking more carefully before the interval opens does not remove it. A design can respond at three levels, and they achieve different things.

Prevention requires the authoritative-state check and external effect commitment to be ordered together against replacements and withdrawals, so that a relevant change ordered before commitment causes refusal or a new check. An indivisible check-and-commit, or a service-enforced conditional

commit with equivalent semantics, may supply that condition. A transaction that commits only a local queue entry does not: the external send can still occur after the governing decision changes.

Narrowing reduces the interval between reading the current versions and committing the effect. A read close to sending may reduce that interval, but a change can still occur before commitment.

Detection is what a carried version identifier supports. If the effect carries the versions it was checked against, a later review can see that the act relied on a decision that had since been replaced. The identifier does not itself narrow the interval and does not prevent the stale effect.

Where only narrowing or detection is available, a stale effect can occur, and the design has to say who detects it and who reconciles it. The direction of the change matters. A withdrawal ordered after the effect commits cannot undo that effect. A withdrawal ordered before it can control whatever remains uncommitted, such as queued sends and preparation that has not yet produced an effect.

Whether a given platform and external service offer any of these mechanisms is an application and platform question that this paper does not settle. Even where prevention holds, it holds relative to what the application can see. The versions it compares are adopted reference facts and the state it reads is observed application state, and both can be wrong. An adopted decision version may not express what the professional meant, and an observed state may lag the world.

State also changes in ways a workflow engine may not record. A client's new limit reaches the process only through a channel that records it. Evaluating an entry sentry while work is available does not by itself establish an effect-time check of continuing authority¹⁰⁵. Observation coverage, the timing of checks and the semantics of the external effect are application concerns. Observation reach therefore has to be stated separately from instruction-relative coverage. A runtime assessment depends on what could be observed, and silence does not establish that the instruction covers the act.

8.3 Unknown and irreversible effects

An effect whose outcome is unknown, such as a transmission that timed out without acknowledgment, forces a choice the workflow engine cannot make alone. A repeat is safe only where the receiving operation has declared effect semantics that make repetition harmless, such as an idempotency key the service honors or a recipient that deduplicates. Otherwise a reconciliation step has to establish what happened first. Without either, a repeat may send a second letter to a counterparty who already has the first, and an unreconciled decision not to repeat may leave a required notice undone. Both outcomes need a named owner. Retry limits and deadlines bound the attempt. They do not resolve the uncertainty.

Deduplication does not supply authority either. An effect key can prevent a second send of V2, and the single transmission can still exceed the decision in force if a limit applicable to its contents arrived after D2. Conversely, an authorized effect can be duplicated. Each check should state which failure it covers and under which assumptions.

Some effects cannot be undone once they occur, and detection after transmission is evidence, not remedy. That is why the design places its check before the effect and treats later detection as input to reconciliation and to the supervising professional.

¹⁰⁵ OMG, *CMMN 1.1*; sec. 8.5, p. 121.

8.4 Where a guard's responsibility ends

A guard governs the paths it mediates, so its claims rest on two assumptions from the enforcement account: decision state changes only through an attributed act of an entitled actor or a delegation that covers the change, and consequential effects reach the world only through the guarded path. External evidence and system proposals cannot adopt themselves. In the trust-boundary branch of the revised-terms matter, counterparty content that asks the system to "approve and send" may supply evidence or a proposal. It is not an attributed adoption by A or C. Neither Q1, nor a "ready" label, nor a restart gives a proposal the standing of an entitled act.

Effects that reach the world by a route the guard does not mediate lie outside what the architecture can claim: a tool holding its own credentials, a person forwarding a draft, a second integration. This paper proposes no comprehensive security engine. It names the assumptions on which the continuing-authority claim depends.

8.5 Authenticated delegation and resource scoping

South and colleagues' position paper supplies the closest technical antecedent for the delegation and enforcement side of this problem. It proposes that a human user authenticate, register the agent and issue a delegation token authorizing limited-scope tasks, with credentials verifiable through standard identity infrastructure. It makes structured resource permissions the primary enforcement mechanism, because natural-language task constraints are brittle. It describes a hybrid in which a language model converts natural-language scoping into a structured policy that the human delegator reviews, corrects and finalizes¹⁰⁶. That human-reviewed translation is a direct antecedent of operative-interpretation review. The authors leave intermediate validation checks and evaluation of translation robustness to future work, and they pair the review with helping the user interpret the generated policies. They also discuss legal grounding, including apparent authority and electronic-transaction law¹⁰⁷, so the legal side is not absent from that account.

The two mechanisms compose, and each has limits. Where withholding a resource expresses the instruction's restriction, resource scoping is the most direct way to enforce it. Removing the outbound channel from the preparation process is the most direct way to enforce D0's reservation of transmission, and it should be preferred wherever it is available. The commit path for decision state is itself a resource whose write permission can be scoped, which is one way to realize the rule that a component may propose decision state but not commit it.

Two limits follow. First, authentication shows who issued a delegation and under which credentials. It does not independently establish that the delegator was entitled to decide the matter the delegation concerns. The client's reservation lies upstream of the token, and a valid token for an attorney does not confer the right to settle.

Second, task semantics are finer than resources. D2 permits transmitting a particular checked artifact version to a particular recipient under then-applicable conditions, and the outbound channel it uses is the same one D0 withheld. What changes between D0 and D2 is the act, its object and its condition. Whether the resource vocabulary can name "V2 to R under D2" is an application choice. Where it

¹⁰⁶ South et al., "AI Agents Need Authenticated Delegation"; secs. 3.2,4.1,5.3.

¹⁰⁷ South et al., "AI Agents Need Authenticated Delegation"; secs. 3.2,4.1,5.3.

cannot, the relation has to be supplied by a check at the effect boundary against the decision and its version. Tying the reviewed policy to the version of the instruction it was meant to express makes the reviewed translation inspectable over later changes, because a later change leaves the confirmation visibly stale. The effect check then compares the act with the decision as well as with the policy.

8.6 Adjacent execution-boundary research

Recent agent-security preprints, all first versions, make related execution boundaries explicit. Commit-time authorization asks whether the evidence licensing an effect remains valid when it is committed¹⁰⁸. Reconstruction-stable authorization asks whether the approved object and grant scope survive representation changes¹⁰⁹. Closure-bound approval addresses effects beyond the surface invocation¹¹⁰. These are technical precedents for continuing validity, fidelity and effect coverage, and none is weighed here as empirically settled. Professional co-control adds questions about who is entitled to adopt the instruction, whether the information and judgment suffice, and which required acts remain outstanding.

8.7 Independent work during a hold

A hold that reaches one dependent act need not stop another, so a durable workflow has to carry status per branch. Selective continuation rests on two conditions. The continuing branch needs independently covering authority that is still current. And the unresolved issue must not reach that branch through a shared constraint, artifact, disclosure or timing dependence. Sharing a constraint does not by itself justify a blanket hold, and the propagation has to be examined (see Figure 3).

If separation fails, a broader hold or another plan is needed. That happens when continuing research would advance a held effect, depend on an unresolved ground or disclose what the hold protects. Preparation inside the system can itself be consequential, so separability has to be assessed for effects and not only for tasks.

Recovery should also restore what is pending outside any hold. After an outage, the required notification O0 returns with its owner, due condition and status, and a due condition that lapsed during the outage still reaches its escalation path. Resuming held work does not perform O0. A failed or unresolved delivery stays pending, an attempt remains distinct from verified delivery, and later performance does not undo the earlier lateness.

8.8 A conventional configuration as the serious counterpart

A conventional case or workflow engine can carry these relations once the application supplies their professional meaning. The full mapping of case roles, case data, required tasks, lifecycle states, sentries and guard errors is given with the closest precedents (§9.7). Three of its findings bear directly on recovery and effects here.

¹⁰⁸ I. Santos-Grueiro, “Temporary Authority, Permanent Effects: Commit-Time Authorization for LLM Agents,” *arXiv:2607.10487v1*, July 11, 2026, version 1; secs. 3,7.

¹⁰⁹ J. Zhu, Z. Liu, S. Fan, J. Chen and Q. He, “From Approval to Execution: Reconstruction-Aware Repair Analysis for LLM-Agent Software,” *arXiv:2609.26529v1*, Sep. 22, 2026, version 1; secs. 2.1–2.2,6.

¹¹⁰ J. Zhang, H. Xia, S. Wu, J. Yue, X. Zhang, Z. Cheng and B. Tu, “Agent Approval Laundering: Transitive Effects Beyond the Approved Invocation,” *arXiv:2609.28586v1*, Sep. 23, 2026, version 1; secs. III,VII.

First, a held continuation needs an explicit rule for which engine state represents it. Failure, disabling and termination are different lifecycle conditions, and none is a selective professional hold by default. Second, a condition satisfied at an earlier event does not show coverage at the time of a later effect, so historical satisfaction and current coverage have to stay distinguishable. Third, an error in evaluating a guard has to be mapped deliberately to a blocked effect and to a responsible actor.

The mapping is a conceptual one under explicit application semantics. It is neither a deliberately weak baseline nor a tested implementation or legal-compliance result.

9 Closest precedents and the rival account

The prior-work comparison tests what the handshake clarifies about a professional decision and the work it governs. Shared-control theory distinguishes proposing a transfer from effecting it. Governed-agency and anticipatory-oversight accounts connect delegated purpose and limits to later behavior. Case-management standards supply roles, human tasks and sentries. The questions here are what each construct expresses, which professional meanings an application must supply, and what combining the relations changes for design and assessment. The comparison is documentary: findings concern selected inspected passages, not deployed behavior or demonstrated benefit.

9.1 Reading the evidence: lineage, version and status

The sources share a lineage, so their agreement is not independent corroboration. Cavalcante Siebert and colleagues adopt Flemisch's consistency account and control-authority example^{111,112}, and extend a tracing requirement through decision–action linkage¹¹³. Santoni de Sio and van den Hoven define tracking through counterfactual responsiveness and tracing through relevant people¹¹⁴. Anticipatory oversight operationalizes tracking of distal reasons¹¹⁵; the constitutive/corrective taxonomy shares two of its authors¹¹⁶.

Version and status differ as well. The anticipatory-oversight and governed-agency texts are first-version 2026 preprints. They are theoretical in the inspected passages, and neither is treated here as empirically validated or settled by review. The constitutive/corrective taxonomy is cited in its second version. The handshake manuscript discussed at the end of this chapter is unrefereed.

9.2 Control allocation: proposal, transfer and nominal sign-off

Flemisch and colleagues distinguish control authority—the authority to execute a distribution of control—from change authority, which permits altering that distribution¹¹⁷. Their worked diagram lets the human change it in both directions while the machine proposes changes; an emergency example grants the machine change authority toward greater automation¹¹⁸. These examples separate a proposal from an authorized transfer and allow machine-initiated changes under a covering

¹¹¹ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; pp. 248–249.

¹¹² Cavalcante Siebert et al., “Meaningful human control: actionable properties”; p. 247.

¹¹³ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; pp. 249–250.

¹¹⁴ Santoni de Sio and van den Hoven, “Meaningful Human Control”; Tracking and Tracing subsections.

¹¹⁵ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; abstract, secs. 1,3.1.

¹¹⁶ Baum and Laux, “Constitutive vs. Corrective”, arXiv:2603.19213v2.

¹¹⁷ Flemisch et al., “Towards a dynamic balance”; pp. 5–6.

¹¹⁸ Flemisch et al., “Towards a dynamic balance”; pp. 8–9.

allocation. Neither such transfer nor change authority necessarily revises a client instruction or a professional's entitlement.

The same account separates actual control, which an outside observer would determine, from notional control, which is still to be established. It offers a control token as one representation of the notional state and observes that a display of who is in control can be separated from the act that exchanges control¹¹⁹. That sits close to a concern developed here through translation checks: a well-formed record of a decision may not be the constraint the system actually applies. The account also treats ability as including competence, skills and resources such as time, tools and personnel, and it proposes consistency among ability, authority, control and responsibility¹²⁰. Its driving example keeps the human's change authority despite diminished ability, for liability and regulatory reasons the authors report¹²¹. That credits the authors' design example and does not verify present law.

The 2023 meaningful-control account begins from an explicit moral operational design domain. Technical operating conditions are not enough. Designers must identify the relevant humans, their moral reasons and the circumstances in which those reasons matter¹²². Its other properties also sit close to the handshake. Human and AI agents should hold compatible, updatable representations¹²³. Ability and authority should match responsibility, and giving a human final authority by design without the ability to use it is criticized¹²⁴. Decisions should link to system actions, with forward and backward links that are different conditions and not two equivalent data links¹²⁵. The authors call these properties necessary but not sufficient, non-binary and context-specific¹²⁶.

Adjustable-autonomy research does more than count transfer costs. Scerri, Pynadath and Tambe describe conditional transfer-of-control strategies that combine transfers among agents and humans with changes to coordination constraints. They address uncertain response and timing by balancing decision quality, interruption and miscoordination costs¹²⁷. Their user constraints can forbid states or actions, including autonomous action that would bypass a required transfer, and can equally require them¹²⁸. Parasuraman, Sheridan and Wickens already make the costs of erroneous decision and action outcomes criteria for allocating automation, and they discuss the situation-awareness, complacency and skill risks when a human must take over¹²⁹. Stakes, recoverability and professional attention here extend that design question and do not originate it.

Continuing, context-sensitive control is therefore the credible comparator. Horvitz addresses when automated action is invoked and terminated¹³⁰; WS-HumanTask types roles and lifecycle

¹¹⁹ Flemisch et al., “Towards a dynamic balance”; p. 10.

¹²⁰ Flemisch et al., “Towards a dynamic balance”; pp. 6–7, 10–11.

¹²¹ Flemisch et al., “Towards a dynamic balance”; p. 12.

¹²² Cavalcante Siebert et al., “Meaningful human control: actionable properties”; sec. 3.1, p. 245.

¹²³ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; p. 247.

¹²⁴ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; pp. 248–249.

¹²⁵ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; pp. 249–250.

¹²⁶ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; p. 251.

¹²⁷ Scerri, Pynadath and Tambe, “Towards Adjustable Autonomy”; sec. 1, pp. 172–173.

¹²⁸ Scerri, Pynadath and Tambe, “Towards Adjustable Autonomy”; sec. 4.4, pp. 196–198.

¹²⁹ Parasuraman, Sheridan and Wickens, “A Model for Types and Levels”; sec. IV.B.2, pp. 292–293.

¹³⁰ Horvitz, “Principles of Mixed-Initiative User Interfaces”, author manuscript; principles 5, 6, 9, 11.

transitions¹³¹; UCON stops usage when requirements cease to hold¹³². Shared-control theory distinguishes actual from nominal allocation¹³³. Green's critical review of government-algorithm oversight questions the evidence that even nominally meaningful review protects against errors and notes that oversight can legitimize a system¹³⁴. Applied by analogy, that critique requires evidence of professional capacity, not an inference from an approval event.

9.3 Durable authorization records and their translation

Baum, Kiener, Langer and Laux organize anticipatory oversight around a normative agenda: an explicit, documented and revisable statement of delegated purposes, boundaries, relevant considerations and deferral triggers¹³⁵. Roles supply defaults; specification, runtime and inspection refine the agenda across timescales through alignment and containment¹³⁶. The account distinguishes specification, inspection, translation and compliance failures¹³⁷ and identifies attention, ritualized oversight and upstream institutional constraints as limits¹³⁸.

Set against the handshake, this account anticipates scope and limits (authorization boundaries), detection and hold (triggers), resumption after change (revision and propagation), standing defaults (roles) and the recurring character of the relation (the spiral). The inspected passages leave open the unit of the relation and the authority behind it. The agenda is a deployment-level statement of a principal's reasons, while the handshake's object is an attributed decision at a judgment boundary, with scope and grounds specific to a matter. Read broadly enough to include matter-level decisions, the agenda is specialized by the handshake to professional work. Read as the deployment-level reasons against which individual decisions are assessed, it sits above the handshake's unit. The inspected passages do not decide between the readings, and on either reading the recurring articulate, run, inspect and revise structure belongs to that account.

The account assumes a single principal and sets aside multi-principal and institutional-principal deployments¹³⁹. That is a scoping choice in its own responsibility analysis. It does not show that its machinery could not represent a client's reserved decision or a firm's supervision. Two of its commitments are not adopted here: strict moral answerability of the principal, and the conclusion that deployments be postponed where its substrate conditions cannot be met¹⁴⁰. They belong to a moral-philosophical argument whose terms differ from U.S. professional-responsibility doctrine.

Dias Ferreira's governed-agency account has the mandate as its central object. The mandate is a record created by the principal's own act of authorization, from which the agent's behavior on the principal's behalf originates and which the principal can revoke. It specifies goals, scope, duration,

¹³¹ OASIS, *WS-HumanTask 1.1*, Committee Specification 01; secs. 3.1,4.10–4.11,7.1.5.

¹³² Park and Sandhu, “The UCONABC Usage Control Model”; sec. 4.2.

¹³³ Flemisch et al., “Towards a dynamic balance”; p. 10.

¹³⁴ Green, “The flaws of policies requiring human oversight”; secs. 4.1,5.1.

¹³⁵ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; sec. 3.1.

¹³⁶ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; secs. 3.2–3.4.

¹³⁷ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; sec. 4.3, pp. 24–25.

¹³⁸ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; sec. 5.

¹³⁹ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; sec. 5.

¹⁴⁰ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; secs. 3.5,4.2.

risk tolerance and constraints, and a standing mandate is distinguished from a task-bounded one¹⁴¹. The account says plainly that translating a mandate into the agent's internal commitments is a place where delegation can fail¹⁴². Authority is contained along each delegation edge, so that it can narrow and not grow. The allowed set is the intersection of the parties' non-negotiable boundaries, including external authorities, and learning may not revise authority¹⁴³. The implementation requirements include a mandate stored separately from the agent's intent, enforcement points outside the learned system, restraint recorded as behavior, narrowing of authority, escalation under degraded conditions and reconstructable episodes¹⁴⁴. Evaluation defines loss relative to authority established independently of the agent's policy¹⁴⁵. Mandate-relative loss is classified by mandate, authority, information, coordination and adaptation, with the sources not assumed separately measurable or summable¹⁴⁶.

These inspected passages connect a durable authorization record to later behavior, preserve scope, duration and revocation, identify translation failure, record restraint and require an authority reference independent of the system under test. The professional synthesis makes the individual attributed decision, its selected grounds and its current coverage the unit of reconsideration. It asks which changes require a client or professional determination rather than re-evaluation of a predicate. Dias Ferreira also distinguishes ownership from mandate and places answerability for an autonomous form without an operative principal mandate in external norms, law and liability¹⁴⁷. That allocation is theoretical, not adopted legal doctrine.

The two accounts organize the problem differently. Anticipatory oversight emphasizes articulated reasons, trajectories and revision; governed agency emphasizes the authorizing mandate, containment across delegation and independent loss references. Both address reasons and authority, and both supply diagnostic comparators. The handshake takes an attributed professional decision, its scope and its grounds as the recurring unit.

Baum and Laux's constitutive/corrective taxonomy distinguishes a human contribution necessary for a process to produce its effective output from a corrective intervention that can alter a process that otherwise proceeds. A mandatory endorsement stays constitutive even if routine, so the distinction cannot be equated with substantive judgment versus mere review¹⁴⁸. The authors also separate causal position from normative adequacy, and their human-overseer definition adds systematic preparation, genuine capacity and a risk-reduction function^{149,150}. The normative overlay is therefore already present in that account. Professional-in-the-loop specifies the conditions on the person in such a position and the externally sourced rights and duties that govern the work.

¹⁴¹ Dias Ferreira, "When Intelligence Becomes Agency", arXiv:2609.07741v1; secs. 4.8–4.11, 7.1, 8.1.

¹⁴² Dias Ferreira, "When Intelligence Becomes Agency", arXiv:2609.07741v1; sec. 4.8 (Def. 4.17).

¹⁴³ Dias Ferreira, "When Intelligence Becomes Agency", arXiv:2609.07741v1; sec. 4.10 (Defs. 4.21–4.22).

¹⁴⁴ Dias Ferreira, "When Intelligence Becomes Agency", arXiv:2609.07741v1; sec. 8.1 (R4, R6, R7, R11–R13).

¹⁴⁵ Dias Ferreira, "When Intelligence Becomes Agency", arXiv:2609.07741v1; secs. 4.10 (Def. 4.20), 7.1.

¹⁴⁶ Dias Ferreira, "When Intelligence Becomes Agency", arXiv:2609.07741v1; sec. 4.10, p. 36.

¹⁴⁷ Dias Ferreira, "When Intelligence Becomes Agency", arXiv:2609.07741v1; sec. 8.1 (R13), p. 61; sec. 10, p. 65.

¹⁴⁸ Baum and Laux, "Constitutive vs. Corrective", arXiv:2603.19213v2; secs. 2.1–2.2, Defs. 1–2.

¹⁴⁹ Baum and Laux, "Constitutive vs. Corrective", arXiv:2603.19213v2; sec. 1, final qualification.

¹⁵⁰ Baum and Laux, "Constitutive vs. Corrective", arXiv:2603.19213v2; sec. 4.1, Def. 3 and normative qualification.

At an unresolved reserved-decision boundary, an entitled contribution is necessary for the dependent consequential continuation. During work covered by a standing instruction, a fresh contribution need not be necessary for each act. Which description applies depends on the process analyzed, which the taxonomy makes explicit through process individuation¹⁵¹. Standing instructions have a prospective-control antecedent in its anticipatory mode¹⁵². The mapping is conditional, and the taxonomy's EU legal background is not authority for the U.S. professional-duty examples used here.

The socio-technical design account bears on design process, not on runtime judgment. It reports open socio-technical systems as interdependent and equifinal, so that goals can be reached by more than one means¹⁵³. It argues that requirements cannot exhaust the richness of the work context¹⁵⁴, and that evaluation continues and changes with the organization¹⁵⁵. Institutional procedures and staffing are therefore design inputs, a formal schema is one operationalization among several, and the propositions of this paper are hypotheses to be evaluated and revised as practice evolves.

9.4 Four relations put to the sources

The professional setting adds relations that these general accounts do not state in professional terms. Three organize the comparison without exhausting the courts, regulators and other external sources that may constrain a matter. The client holds decision rights that bound what the professional may instruct. The professional is entitled to exercise judgment within a scope and bears duties for it. The firm or institution allocates supervision, coverage and takeover of pending decisions. A system can recognize these relations but cannot create them. A record of them supplies no entitlement and does not by itself establish that duties were met.

Four questions organize the selected comparisons. Can a constraint that originates with a party other than the designer, deployer or principal bind the professional's instruction and the system's authority, with changes governed by that party's rights? Can entitlement to decide be represented apart from ability, information and system-granted permission, so that these can diverge? Can a required communicative act, such as information or consultation owed to a right-holder, be represented separately from the constraint that right-holder supplies, with an owner, a due condition and a recorded performance? And can reassignment or takeover of a pending decision be represented with the competence, authority and context it requires, and not as an interchangeable approver swap? The findings, bounded to the inspected passages, are these.

¹⁵¹ Baum and Laux, “Constitutive vs. Corrective”, arXiv:2603.19213v2; sec. 2.3, Boundary Cases and Process Individuation.

¹⁵² Baum and Laux, “Constitutive vs. Corrective”, arXiv:2603.19213v2; sec. 2.3, Anticipatory HOTL.

¹⁵³ G. Baxter and I. Sommerville, “Socio-technical systems: From design methods to systems engineering,” *Interacting with Computers*, vol. 23, no. 1, pp. 4–17, 2011, doi: 10.1016/j.intcom.2010.07.003. Cited to-appear author manuscript; manuscript pagination differs from the published article; manuscript p. 4.

¹⁵⁴ Baxter and Sommerville, “Socio-technical systems”, to-appear author manuscript; manuscript p. 17.

¹⁵⁵ Baxter and Sommerville, “Socio-technical systems”, to-appear author manuscript; manuscript p. 18.

Table 7. Completed bounded documentary comparison of four professional relations and their application semantics.

Relation	Constructs found in the inspected passages	What an application must supply	Finding
Upstream constraints	External legislation in a shared-control example ¹⁵⁶ ; moral operational design domain ¹⁵⁷ ; roles and upstream commitments in the anticipatory agenda ^{158,159} ; intersection of boundaries including external authorities in governed agency ¹⁶⁰	The client's reserved decisions and the professional's scoped obligations, with holder, source and a rule for who may change them	The constructs support a configuration with client rights supplied through scoped professional application semantics.
Entitlement apart from ability, information and permission	Institutional power, permission and obligation ^{161,162} ; representation and mandate duty ¹⁶³ ; ability, authority and control ^{164,165} ; decision owners and makers ¹⁶⁶	The actual client or legal entitlement, as an input from outside the model	The distinctions can be represented. Owner or maker metadata documents governance without creating the entitlement it describes.

¹⁵⁶ Flemisch et al., “Towards a dynamic balance”; p. 12.

¹⁵⁷ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; sec. 3.1, p. 245.

¹⁵⁸ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; secs. 3.2–3.4.

¹⁵⁹ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; sec. 5.

¹⁶⁰ Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; sec. 4.10 (Defs. 4.21–4.22).

¹⁶¹ A. Artikis, M. Sergot and J. Pitt, “Specifying Norm-Governed Computational Societies,” Technical Report DTR06-5, Imperial College London, 2006, report PDF. Pinpoints use the report's pagination; sec. 2.1, pp. 5–7.

¹⁶² Artikis, Sergot and Pitt, “Specifying Norm-Governed Computational Societies”, DTR06-5; p. 20.

¹⁶³ J. Gelati, G. Governatori, A. Rotolo and G. Sartor, “Declarative Power, Representation, and Mandate. A Formal Analysis,” in T. J. M. Bench-Capon, A. Daskalopulu and R. G. F. Winkels (eds.), *Legal Knowledge and Information Systems. Jurix 2002: The Fifteenth Annual Conference*, Amsterdam: IOS Press, 2002, pp. 41–52, conference PDF; pp. 41–43.

¹⁶⁴ Flemisch et al., “Towards a dynamic balance”; pp. 5–6.

¹⁶⁵ Flemisch et al., “Towards a dynamic balance”; pp. 6–7, 10–11.

¹⁶⁶ OMG, *DMN 1.5*; sec. 6.3.7, Table 11, pp. 37–38.

Relation	Constructs found in the inspected passages	What an application must supply	Finding
Required conduct	Required-state and required-action constraints ¹⁶⁷ ; normative obligations ¹⁶⁸ ; required tasks ¹⁶⁹ ; achievement versus maintenance obligations ¹⁷⁰	Owner, due condition, performance evidence and adequate fulfilment as distinct meanings	Positive-conduct constructs exist. A permission intersection alone does not perform a communication.
Succession and takeover	Human-task and case roles ^{171,172} ; the shared-control meta-system ¹⁷³ ; institutional capacity ¹⁷⁴ ; standing and resources ¹⁷⁵ ; supervisory nodes in the governed-agency delegation graph ¹⁷⁶ ; decision owners and makers ¹⁷⁷	Alternate roles, qualified entitlement, information and the actual taking-up of responsibility	Alternate roles are configurable. Entitlement, information and uptake require application or institutional determination.

Upstream constraints. Four of the sources already anticipate constraints that come from outside the control relation. None is a model of U.S. client decision rights, and none claims to be. The finding supports a candidate representation. A client's reserved decision enters as a constraint with a holder, a source and a rule about who may change it, through scoped professional semantics the application supplies. That applies existing constructs and does not expose a gap in them. The professional chapter (§6.2) develops the client-rights case under selected New York rules.

Entitlement. Artikis, Sergot and Pitt specify open agent societies through four levels of social constraint: physical capability, institutionalised power, permission with prohibition and obligation, and sanctions with enforcement policy¹⁷⁸. An agent empowered within an institution creates or modifies institutional facts by performing a specified kind of act. The same act by an agent without the power has no effect on institutional status, though it may have other effects¹⁷⁹. Permission has no

¹⁶⁷ Scerri, Pynadath and Tambe, “Towards Adjustable Autonomy”; sec. 4.4, pp. 196–198.

¹⁶⁸ Artikis, Sergot and Pitt, “Specifying Norm-Governed Computational Societies”, DTR06-5; sec. 2.1, pp. 5–7.

¹⁶⁹ OMG, *CMMN 1.1*; sec. 5.4.11.2, p. 54.

¹⁷⁰ S. Colombo Tosatto, G. Governatori and P. Kelsen, “Towards an Abstract Framework for Compliance,” in *2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops*, pp. 79–88, 2013, author-hosted article; sec. II.B, definitions 10–11.

¹⁷¹ OASIS, *WS-HumanTask 1.1*, Committee Specification 01; secs. 3.1, 4.10–4.11, 7.1.5.

¹⁷² OMG, *CMMN 1.1*; sec. 5.2.2, p. 20.

¹⁷³ Flemisch et al., “Towards a dynamic balance”; pp. 5–6.

¹⁷⁴ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; pp. 248–249.

¹⁷⁵ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; sec. 5.

¹⁷⁶ Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; sec. 4.10 (Defs. 4.21–4.22).

¹⁷⁷ OMG, *DMN 1.5*; sec. 6.3.7, Table 11, pp. 37–38.

¹⁷⁸ Artikis, Sergot and Pitt, “Specifying Norm-Governed Computational Societies”, DTR06-5; sec. 2.1, pp. 5–7.

¹⁷⁹ Artikis, Sergot and Pitt, “Specifying Norm-Governed Computational Societies”, DTR06-5; sec. 2.1, pp. 5–7.

fixed relation to power, so an act may be valid and permitted, valid and forbidden, invalid and permitted, or invalid and forbidden¹⁸⁰.

Gelati and colleagues separate representation, a representative's capacity to act in the principal's name, from mandate, the mandatee's duty to act as requested. A representative who acts within power but against duties still binds the principal, subject to a qualification where the counterparty knew or should have known. Acting beyond power generally leaves the principal unbound¹⁸¹. Both are historical conceptual formalizations. The Artikis report assumes constraints fixed at design time, a single society and an external perspective on observable behavior¹⁸², and neither source states current U.S. law. They serve here as anticipation and as a precision tool.

A stipulated institution makes the point without asserting any rule of law. Suppose a message sent in a given form by a designated role-holder counts as an offer. A transmission outside the professional's instruction may still count as the offer if the sender held the power, while the instruction's limits settle only whether sending was permitted. A transmission by a sender lacking the power may be software-allowed and still have no institutional effect. Whether any actual act has such effects depends on applicable law and facts. The professional chapter (§6.2.1) takes up one decided case on its own facts.

Required conduct. Required-state and required-action constraints appear in adjustable autonomy¹⁸³, and normative-systems specifications distinguish institutional power from permission, prohibition and obligation¹⁸⁴. The contract-net example in the Artikis report obliges an awarded party to report a result by a deadline. The authors note that in realistic settings the report would not itself discharge the obligation, because completion should be externally observable¹⁸⁵. Tosatto, Governatori and Kelsen distinguish an achievement obligation, which requires a condition to hold in at least one state within its activation period, from a maintenance obligation, which requires it in every state. They treat lifelines and deadlines as conditions over states and not as wall-clock values¹⁸⁶. In that vocabulary the stipulated notification O0 reads as an achievement-type duty, and its due condition may be expressed by clock or by state depending on the application. The framework is a single-regulation formalism with a primary obligation and compensations. It is an antecedent and not a professional compliance engine.

UCON distinguishes usage obligations from duties assigned independently of a subject's usage request, and it excludes the latter from its model¹⁸⁷. The question here is how a required communication gets an owner, a due condition and a status, and what follows when the condition lapses. An intersection of permitted sets can express prohibitions. It does not identify what must be done, by whom or when. The application must give owner, due condition, actual performance and adequate fulfilment distinct meanings, since a required task's completion does not establish that the

¹⁸⁰ Artikis, Sergot and Pitt, "Specifying Norm-Governed Computational Societies", DTR06-5; p. 20.

¹⁸¹ Gelati et al., "Declarative Power, Representation, and Mandate"; pp. 41–43.

¹⁸² Artikis, Sergot and Pitt, "Specifying Norm-Governed Computational Societies", DTR06-5; pp. 4–5.

¹⁸³ Scerri, Pynadath and Tambe, "Towards Adjustable Autonomy"; sec. 4.4, pp. 196–198.

¹⁸⁴ Artikis, Sergot and Pitt, "Specifying Norm-Governed Computational Societies", DTR06-5; sec. 2.1, pp. 5–7.

¹⁸⁵ Artikis, Sergot and Pitt, "Specifying Norm-Governed Computational Societies", DTR06-5; p. 20.

¹⁸⁶ Colombo Tosatto, Governatori and Kelsen, "Towards an Abstract Framework for Compliance"; sec. II.B, definitions 10–11.

¹⁸⁷ Park and Sandhu, "The UCONABC Usage Control Model"; sec. 3.5, p. 140.

duty was adequately discharged¹⁸⁸. Conflicts among sources need an externally sourced rule. No universal ordering of client, professional and institutional instructions is assumed, and an unresolved precedence question that affects continuation reaches an entitled actor.

Succession. Allocation and oversight structures exist in each source in the table. The shared-control account describes authority as usually assigned beforehand by the designer, within a meta-system that includes development and evaluation before and after operation¹⁸⁹. The 2023 meaningful-control account names engineers, designers, operators, users and managers among the relevant humans¹⁹⁰, and it recognizes that training, oversight procedures and administrative discretion shape authority and ability¹⁹¹. Baum and Laux distinguish the operative actor from the overseer even where one person occupies both roles, and they propose layered oversight¹⁹². A configuration can therefore express alternate roles. Whether an alternate is qualified under the applicable legal entitlement, has the information the decision needs and has actually taken up responsibility are determinations an application or institution has to make. No comparator is treated as incapable of representing them.

9.5 Change, hold and renewal

Several of these sources already require change to propagate. The meaningful-control account requires updating shared representations when relevant human reasons change¹⁹³. UCON stops usage when authorization predicates fail¹⁹⁴. The anticipatory-oversight account treats the agenda as revisable, with propagation to what governs the agent¹⁹⁵. The governed-agency account requires versioned mandates with expiry and revocation¹⁹⁶. Fallback and emergency transfer have precedents in the shared-control emergency example¹⁹⁷, in the meaningful-control proposal of fallback when human ability degrades¹⁹⁸, and in adjustable-autonomy strategies¹⁹⁹. Learned routing that decides when to pass a case to a downstream decider by optimizing a stated loss²⁰⁰ is a mechanism a governed workflow can use for discretionary escalation. It supplies no authority for a continuation or a hold, and it cannot waive a boundary reserved by the client or by rule.

The narrower question concerns the consequence for an adopted decision: which dependent actions hold, which changes stay within standing scope, and who may adjudicate renewed judgment as against re-evaluating an already authorized predicate. The inspected meaningful-control passages do not specify that resumption relation. The inspected anticipatory-oversight and governed-agency passages do not describe it for an individual attributed decision, though later work or other passages may.

¹⁸⁸ OMG, *CMMN 1.1*; sec. 5.4.11.2, p. 54.

¹⁸⁹ Flemisch et al., “Towards a dynamic balance”; pp. 5–6.

¹⁹⁰ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; p. 249.

¹⁹¹ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; pp. 248–249.

¹⁹² Baum and Laux, “Constitutive vs. Corrective”, arXiv:2603.19213v2; sec. 4.2.

¹⁹³ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; p. 247.

¹⁹⁴ Park and Sandhu, “The UCONABC Usage Control Model”; sec. 4.2.

¹⁹⁵ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; secs. 3.4–3.5.

¹⁹⁶ Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; sec. 8.1 (R4,R6,R7,R11–R13).

¹⁹⁷ Flemisch et al., “Towards a dynamic balance”; pp. 8–9.

¹⁹⁸ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; pp. 248–249.

¹⁹⁹ Scerri, Pynadath and Tambe, “Towards Adjustable Autonomy”; sec. 1, pp. 172–173.

²⁰⁰ Madras, Pitassi and Zemel, “Predict Responsibly”; secs. 2.1–2.4, pp. 2–4.

Kozlovski's discussion of reasons supports a caution that is this paper's own inference. If a professional's decision selected among options the reasons left eligible, a change in one recorded ground may not settle whether the decision still stands²⁰¹. Detection and a conservative hold can be automatic. Where independently authorized criteria do not settle whether the decision survives, the question reaches the entitled decider, and that limits how far a grounds-linked hold can be made mechanical.

9.6 Delegation scholarship from the legal and security sides

Two direct engagements come from outside the control-theory lineage. Kolt's analysis of governing AI agents applies agency doctrine to authority, loyalty, delegation to subagents and monitoring, and it treats instructions as subject to interpretation and clarification^{202,203,204}. It is scholarship and not binding law. South and colleagues' authenticated-delegation position supplies credentials, resource scoping and a human-reviewed translation from natural-language scope to structured policy²⁰⁵. That already anticipates the reviewed translation step discussed here, and the professional chapter (§6) and the durable-workflow chapter (§8) credit it where it bears. Both sources address delegation to AI agents in general. The present account concerns a matter-level professional decision and its continuation, and neither source is read as lacking what this account adds.

9.7 What a conventional case-management realization expresses

A capable case-management configuration is the strongest conventional counterpart, and its mapping is set out here in full. CMMN 1.1 takes legal and medical cases among its motivating examples. It defines case roles, human tasks, run-time planning and conditions over case data^{206,207,208,209}. A sentry combines events and conditions to enable or end work²¹⁰, and suspension or resumption can pass from a stage to its contents²¹¹. Selectivity at stage granularity is therefore already expressible and is not a distinctive handshake mechanism. DMN 1.5 supplies a close professional-support illustration. Its partial-automation discussion describes automated referral to a human reviewer whose decision logic may stay unspecified²¹². Its lending example separates automated routing from a credit officer's adjudication: the officer considers the routing result and case data, and the human decision has no specified logic²¹³. The separation of a computed assessment from an entitled actor's adoption thus has a clear antecedent.

²⁰¹ Kozlovski, “Reasons underdetermination”; When reasons underdetermine choice.

²⁰² Kolt, “Governing AI Agents”; pp. 359–361.

²⁰³ Kolt, “Governing AI Agents”; pp. 365–366.

²⁰⁴ Kolt, “Governing AI Agents”; pp. 370–371.

²⁰⁵ South et al., “AI Agents Need Authenticated Delegation”; secs. 3.2,4.1,5.3.

²⁰⁶ OMG, *CMMN 1.1*; sec. 4.1, p. 5.

²⁰⁷ OMG, *CMMN 1.1*; sec. 5.2.2, p. 20.

²⁰⁸ OMG, *CMMN 1.1*; sec. 5.4.10.4, p. 46.

²⁰⁹ OMG, *CMMN 1.1*; sec. 5.4.9, p. 38.

²¹⁰ OMG, *CMMN 1.1*; sec. 5.4.6, pp. 32–33.

²¹¹ OMG, *CMMN 1.1*; sec. 8.4.2, Tables 8.8–8.9, pp. 115–119.

²¹² OMG, *DMN 1.5*; sec. 5.2.2, p. 10.

²¹³ OMG, *DMN 1.5*; sec. 11.1.3, pp. 169,172–173.

Terminology needs care. DMN uses *decision* for an act of determining an output from inputs and logic, and it also models human decisions without specified logic²¹⁴. This paper uses *adopted decision* for the attributed act that supplies operative scope, and *decision model* for a functional assessment or selection responsibility, distinct from DMN's requirements-and-logic artifact.

A hypothetical configuration can store decision, scope and grounds references in case data, use human tasks for adoption and reconsideration, and guard dependent work through events and conditions. CMMN documents calls to DMN 1.x decisions²¹⁵. This sketch establishes no tested, version-compatible deployment and, because these requirements shaped it, cannot substitute for an independently designed comparator.

The application must give these constructs their professional meaning. Seven mappings carry most of the weight.

Roles and entitlement. CMMN roles authorize actions within a case, and assigning participants to roles lies outside its scope²¹⁶. DMN's authority requirements document governance relationships and have no execution semantics²¹⁷. Decision owner and maker fields are metadata of the same kind²¹⁸. Neither a role nor an owner field creates entitlement. The configuration has to take the actual client or legal entitlement as an input from outside the model and say what follows when a role-holder lacks it.

Grounds and versions. Case-file-item versioning lies outside the base case model²¹⁹. A guard that reads "current" case data later sees whatever a mutable item then holds. The grounds at adoption therefore need an identified version, so that a change detected later is compared with what the decision actually rested on.

Adoption. A reserved or out-of-scope proposal needs an entitled actor's adoption; the human task records that act. A selection covered by a valid standing instruction or delegation needs no fresh adoption task, and its performing actor remains attributed. Task completion supplies no entitlement.

Hold and lifecycle. A held continuation needs an explicit rule for which engine state represents it. Suspension and resumption are assigned to a case-worker decision or to propagation from an outer stage. A software exception puts an item in Failed, which does not propagate, and reactivation from Failed is specified for when the source of failure has been resolved²²⁰. Suspension, termination, failure and disabling are therefore different lifecycle conditions, and none is a selective professional hold by default. The application must decide how an unresolved authority question maps onto these states and must not treat a failure as a hold. A disabled item is semi-terminal and counts as terminal when completion of the parent stage is calculated²²¹, which matters wherever a held act's disabled status would let its stage complete.

²¹⁴ OMG, *DMN 1.5*; sec. 5.3.1, p. 12.

²¹⁵ OMG, *CMMN 1.1*; sec. 5.4.10.7.1, Table 5.50, p. 51.

²¹⁶ OMG, *CMMN 1.1*; sec. 5.2.2, p. 20.

²¹⁷ OMG, *DMN 1.5*; sec. 6.2.2.3, p. 23.

²¹⁸ OMG, *DMN 1.5*; sec. 6.3.7, Table 11, pp. 37–38.

²¹⁹ OMG, *CMMN 1.1*; sec. 5.3.2.1, p. 23.

²²⁰ OMG, *CMMN 1.1*; sec. 8.4.2, Tables 8.8–8.9, pp. 115–119.

²²¹ OMG, *CMMN 1.1*; sec. 8.4, Table 8.4, pp. 110–111.

Termination followed by a new repeated instance, with a human renewal task and manual activation of the successor, is another plausible realization. CMMN repetition is triggered by entry criteria or by completion or termination, and each repetition is a new instance²²². That differs from resuming the same instance. Both mappings need examination of identity, retained decisions, completion consequences and which instruction permits the successor's effects.

Sentries, and historical satisfaction versus present coverage. Entry and exit sentries enable or terminate work²²³. Entry-time evaluation is not itself an effect-time authority check²²⁴. The design consequence is to distinguish historical satisfaction from present coverage: specify what is retained, refreshed or returned for judgment, under the observation assumptions. This consequence is the paper's interpretation, not a specification feature.

Required tasks. A required task can constrain stage completion²²⁵. That lets the stipulated notification O0 carry an owner, a due condition and a status. Completion or termination of the task still does not establish that an affirmative professional duty was adequately discharged, and a hold on another task does not perform it.

Guard errors and effect time. SCXML treats an unevaluable conditional as false and queues an execution error²²⁶. Whether that blocks a consequential effect depends on which transition the guard gates. A false permission-to-proceed condition may block the transition, while a false condition for entering a hold may leave the hold untaken. How the error reaches a responsible actor is an application responsibility²²⁷. A well-typed, executable condition can also express the wrong instruction faithfully, so the translation check of §7 applies in a conventional configuration as well. A case engine does not by itself order the authoritative-state check against the external commitment. Preventing a stale effect needs the ordering conditions of §4.5, or a service-enforced conditional commitment with the same semantics. A version identifier carried with the effect supports detection afterward and does not prevent the effect.

These mappings are conceptual and rest on explicit application semantics. They are neither a deliberately weak baseline nor a tested implementation or legal-compliance result.

9.8 What the synthesis adds

The synthesis gives existing control and workflow constructs a coordinated professional interpretation. Its contribution is a set of design and assessment questions that remain visible across an attributed decision, the interpretation applied and later changes. Three forms of value follow.

First, it fixes which relations must stay recoverable for consequential work over time under revisable delegation. Each comes with a removal argument and the simpler arrangement that may do the job. A designer can then see what a given configuration has dropped, and which professional meanings

²²² OMG, *CMMN 1.1*; secs. 5.4.11.1, 5.4.11.3, pp. 53–55.

²²³ OMG, *CMMN 1.1*; sec. 5.4.6, pp. 32–33.

²²⁴ OMG, *CMMN 1.1*; sec. 8.5, p. 121.

²²⁵ OMG, *CMMN 1.1*; sec. 5.4.11.2, p. 54.

²²⁶ W3C, *SCXML*, Sep. 1, 2015 Recommendation; secs. 3.5, 5.9.1.

²²⁷ W3C, *SCXML*, Sep. 1, 2015 Recommendation; secs. 3.5, 3.12.2, 5.9.1.

the standard constructs leave to the application. In the inspected passages the pieces of this set appear in different accounts and at different units.

Second, it keeps apart relations that practice tends to merge: entitlement and software allowance, the instruction issued and the interpretation consumed, a hold and a required communication, a later authorization and an earlier mismatch. That gives a reviewer a way to say which relation failed.

Third, it states what an assessment must hold separate: outcome, authority-sensitive trajectory, instruction-version fidelity and professional burden.

A capable governed workflow that preserves the same relations is a serious counterpart. Mechanism equivalence leaves open whether the allocation is justified, whether the vocabulary helps diagnosis and what the structure costs professionals. The documentary comparison does not answer those questions. It makes the comparisons that would answer them more precise.

9.9 An earlier handshake protocol

The term *handshake* has an earlier use in Chesher's manuscript, "The Handshake Protocol: Protecting Expert Judgment in Sustained Human-AI Collaboration"²²⁸. It protects expert judgment in sustained collaboration through persistent behavioral contracts, reflective capture and automated guards²²⁹. Its friction-to-rule-to-guard loop is a relevant antecedent for structural governance and for externalizing collaboration knowledge. The cited manuscript develops a protocol for expert collaboration, and this paper draws from it no empirical result about professional workflow. The present account concerns a different relation: how an attributed matter decision, grounded in independent entitlement, governs operative scope and interpretation, and how unresolved coverage returns dependent effects for reconsideration. The distinction is the matter-specific decision–work relation, developed against this established expert-collaboration precedent.

9.10 Four kinds of comparison

Four kinds of result should not be confused. *Documentary expressibility* asks whether the constructs a source or standard defines, with application semantics added, can express a relation. The comparison above is of that kind, bounded to selected passages. *Semantics-matched conformance* asks whether an implementation built to this paper's semantics shows the behavior the resumption rows describe. It would test implementability and the checks, and it would not by itself establish comparative benefit.

An *independently designed capable comparator* is a workflow specified by designers who receive the same professional requirements, evidence and scenarios without this vocabulary. Comparing it with the matched implementation could support a design-benefit claim. Added attention or translation cost without corresponding benefit would count against practical use. A simpler final-review workflow is a limited baseline and cannot stand in for it. *Prospective diagnosis and burden* asks whether reviewers locate a failed relation more accurately or with less effort under equally informative conditions, and what the structure costs the professionals who maintain and use it. Only the first kind has been done, and only for selected passages. The other three are proposals in the

²²⁸ Chesher, "The Handshake Protocol", Mar. 31, 2026 manuscript; secs. 4.2–4.3,5.4,6.4, pp. 7–11,13.

²²⁹ Chesher, "The Handshake Protocol", Mar. 31, 2026 manuscript; secs. 4.2–4.3,6.4,7, pp. 7–9,13–14.

assessment chapter and are not reported here. The development notes from the prototype illustrate design pressures. They establish no historical derivation or priority for any element of the account.

10 Assessing the relation: targets, comparisons and limits

If a professional decision continues to govern the work that follows it, an assessment has to be able to see that relation, and a terminal score cannot. A matter can end acceptably after an act no one was entitled to take, and badly although every act stayed within scope. An instruction can be faithfully implemented and inadequate, or adequate and unfaithfully implemented. A required notice can lapse while every prohibited effect is avoided. An assessment has to follow the distinctions the account has introduced and say, for each claim, what observation would count for or against it.

This chapter separates four assessment targets, each with its own reference. It then states three prospective comparisons, P1 to P3, and a first study that needs no implementation. No study is reported. The propositions are conditional comparisons with defeat conditions fixed in advance (Table 9).

10.1 Four targets, four references

Outcome asks whether the matter reached an acceptable end state and the work product meets the task's requirements. Its reference is a set of end-state criteria, including permitted alternatives, prepared by reviewers independent of the system under test.

Authority-sensitive trajectory asks whether each consequential act occurred within the entitlement and scope then current, after any decision the act required, and whether stipulated required conduct was performed. Its reference is the independently assessed entitlement, client and institutional constraints and stipulated requirements at the time of each act. Four questions about an act stay separate in scoring: whether required conduct was performed, whether the software allowed the act, whether it had institutional effect, and whether the actor was normatively entitled. A "continues" label records what the software concluded and settles none of the other three.

Instruction-version fidelity asks whether the operative interpretation in force at each version expressed the instruction actually issued at that version, including any discretion deliberately delegated. Its reference is that issued instruction, and not the system's restatement of it or the later amended state. Fidelity differs from adequacy, which compares the instruction with the normative requirements of the matter, and from permissibility, which compares what happened with the same requirements. A faithful implementation of an inadequate instruction is a fidelity pass and an adequacy finding that only professional or supervisory assessment can make.

Professional burden asks what attention, interruption, rework, maintenance and supervision the arrangement demanded of experienced professionals, and whether they could rely on it appropriately. Its reference is observed professional work against task references defined independently of reported ease.

Table 8. Four assessment targets and their independent references.

Target	Reference	What is observed	What a pass leaves open
Outcome	Acceptable end state, set by independent reviewers	Work product, end state, completion	Whether the route stayed within authority and carried the instruction faithfully
Authority-sensitive trajectory	Entitlement, constraints and stipulated requirements at the time of each act	Acts against their opportunities, holds, required communications, patterns of acts	Whether the instruction was adequate and the decision informed
Instruction-version fidelity	The instruction as issued at each version, with delegated discretion	Each version of the operative interpretation, tested by attempted acts	Whether the instruction was adequate and what its effects were
Professional burden	Observed work of experienced professionals	Time, interruption, rework, maintenance, supervision, reliance	Whether the judgment exercised was adequate

Two references especially need independence: the actor's entitlement and the instruction actually issued. The system's classification of its own trace establishes neither. Each interpretation is assessed against its instruction version, including deliberately delegated alternatives. Confirmations, clarifications, corrections, renewals and new authorizations are later events; they do not rewrite an earlier fidelity finding.

Required conduct needs separate evidence. For O0, the stipulated owner, content, recipient, due condition and escalation rule define the reference. Scoring checks assignment, delivery and timing, distinguishes an attempted send from verified delivery, and records unresolved or late performance. Reassignment changes ownership, not completion; later delivery does not meet the original timing condition. Omissions are measured against opportunities that required the communication. Neither a good outcome nor avoiding prohibited effects can cancel an unmet requirement in a weighted composite.

The four targets are not exhaustive. Evaluation work on proactive agents distinguishes further targets, such as coordination among agents and the integrity of adaptation, which this chapter leaves aside²³⁰.

10.2 What existing evaluation work already separates

Existing evaluation work supplies important parts of this frame. The first version of tau-bench scores final state and output, acknowledges that such a reward can miss a required user confirmation, separates repeated-trial consistency from a single success and states the limits of its user simulator²³¹. The first version of ST-WebAgentBench moves nearer to authority-sensitive trajectories. It supplies organizational, user and task policies, counts violations by source and by category, including user

²³⁰ Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; sec. 7.1.

²³¹ S. Yao, N. Shinn, P. Razavi and K. Narasimhan, “τ-bench: A Benchmark for Tool-Agent-User Interaction in Real-World Domains,” *arXiv:2406.12045v1*, June 17, 2024, version 1; secs. 3,6.

consent and out-of-scope action, and credits a completed task only when no monitored violation is recorded²³². That gating matters, because a task that crossed a boundary is not credited as a success. Separate completion and violation observations still help distinguish blocked work, unnecessary holds and kinds of failure.

Its checks are predicates over action trajectories and page content, with a simulated user who confirms when asked²³³. Its fixed precedence among organizational, user and task policies is a benchmark stipulation²³⁴ that this account does not adopt, because the relations among client, professional and institutional constraints derive from their external sources. APEX-Harm reports attack success and task performance separately in professional workspaces, legal tasks among them²³⁵. Ling and colleagues' study of household and mobile-agent workflows examines how local decision savings interact with task completion, recovery and resources per successful outcome²³⁶. Together these works distinguish completion, policy conformance and local efficiency. They do not establish legal professional adequacy, informed client consent or any benefit of the handshake, and none of their scores transfers to legal matters.

The governed-agency account rejects a single task score and stipulates independent authority and evaluator references, without empirical validation in the inspected passages²³⁷. It separates mandate alignment, constraint violations, missed or inappropriate intervention, delay, delegated-authority integrity, coordination faults and reconstructability. Waiting and deferring count as behavior²³⁸.

That precedent supplies separate reporting, error treatment in both directions, independent references and observable holds. The assessment here applies those principles to a particular relation: an earlier decision that governs resumed work whose grounds may change. It asks for references prepared by independent professional reviewers who may reasonably disagree or leave options open. It keeps the authority a professional can exercise apart from the permissions a system grants, and it calls for observing the burden placed on experienced professionals. The inspected evaluation work does not show how such references are produced where professionals may legitimately differ. The benchmarks above settle authority and consent by stipulation.

10.3 Scenarios, references and denominators

The propositions share one scenario design. A finite scenario set, its requirements and the reference-preparation procedure are fixed before any system runs, together with the monitored predicates, their observation conditions and the gap between each software test and any professional duty it helps assess. One family elaborates the revised-terms matter: preparation is authorized, external communication is reserved and O0 falls due. Variants leave O0 unallocated, make its owner unavailable, make a required checker unavailable, or bring a later authorization. The reference treats

²³² Levy et al., “ST-WebAgentBench”, arXiv:2410.06703v1; secs. 3.1,3.2.2; sec. 3.3, Table 2.

²³³ Levy et al., “ST-WebAgentBench”, arXiv:2410.06703v1; sec. 3.3; App. A.2.

²³⁴ Levy et al., “ST-WebAgentBench”, arXiv:2410.06703v1; sec. 3.1.

²³⁵ Z. Yang, A. Bibi, G. Li, B. Ghanem and C. Hinojosa, “APEX-Harm: Evaluating Agent Safety in Shared Professional Workspaces,” *arXiv:2610.13623v1*, Oct. 9, 2026, version 1; secs. 1,3.3,4.1,5.

²³⁶ Y. Ling, S. Yu, Z. Chen and C. Fang, “When Faster Decisions Do Not Make Better Agents: An Empirical Study of Jev in Agent Workflows,” *arXiv:2610.13092v1*, Oct. 9, 2026, version 1; secs. 4.2,5.3,6.

²³⁷ Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; secs. 4.10 (Def. 4.20),7.1.

²³⁸ Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; secs. 4.5 (Def. 4.11),8.1 (R4,R12).

sending as inapplicable under D0 from the outset. It is not an unconfirmed act, and it is not an allowed act that a later change puts on hold. A changed ground, by contrast, can leave the coverage of an act that D0 does cover unresolved.

Other families vary changes that stay within delegated discretion and changes that do not, missed changes, a race between check and effect, a duplicate effect after an unknown outcome, an unavailable professional near a deadline, upstream client constraints, standing instructions whose validity is stipulated for the software test, and a succession in which the alternate's entitlement is disputed. These are prospective injections into synthetic scenarios. They reconstruct nothing observed in development.

Independent professional reviewers fix the reference before seeing traces: required behavior, prohibited continuation, permissible alternatives and unresolved questions. The reference need not have one gold answer. Kozlovski's contested account of moral reasons illustrates why several eligible choices may lack a determinate winner²³⁹. Where the scenario allows that discretion, choosing one permitted option is not an error.

Reviewer disagreement can reflect real plurality, a defective scenario or reviewer error. The protocol records it with its apparent source and does not claim to settle which obtains. Independence from the system's traces is separate from independence from the scenarios' authors. Reviewers may flag a stipulated instruction or authority relation as professionally implausible. Such a scenario can remain a software test of the stipulation and supports no claim about realistic practice. The evidence reviewers could access is recorded, so that a restricted reference is not mistaken for a complete judgment on the matter.

Observation is kept apart from the system's trace. An external reference ledger records which injected changes occurred and whether the declared observation conditions made them available to the system. It retains information the system lacked without treating it as available at the decision point. It distinguishes a change not detected when detection was possible, a detected change without the required control response, and a response too late to prevent the affected effect. Overlapping faults and uncertain timing stay visible and are not forced into exclusive labels. A change the system could not have observed tests the observation assumptions and does not count as a failure to detect accessible information. The absence of an event from the system's log cannot establish that nothing changed.

Rates have denominators fixed from the scenario requirements and not from what a system happens to attempt. Prohibited continuation is reported against opportunities for which the reference prohibits continuation, including known out-of-scope acts and required holds. Unnecessary holds are reported against opportunities where the reference requires continuation without a hold. Opportunities for which either response is permissible within discretion are reported separately and enter neither denominator. An opportunity never reached is its own observation, so early stopping cannot improve a rate by shrinking its denominator, and retries or several actions in one scenario must not silently create different denominators.

Holds are assessed twice, against the declared normative requirements and against what the observation conditions made available. A hold that hindsight shows to have been unnecessary may

²³⁹ Kozlovski, "Reasons underdetermination"; When reasons underdetermine choice.

have been justified by accessible uncertainty. A hold despite an accessible covering instruction may reflect misclassification. The delay a hold caused and its basis are reported beside impermissible continuation.

Some requirements concern a pattern of acts and not any single act. Anticipatory-oversight theory argues that acts each defensible in local context can together depart from what a principal would endorse, and that such departures can become detectable when trajectories are read against a standing statement of the principal's reasons²⁴⁰. That account is a philosophical preprint with no empirical validation in the inspected passages, and its examples come from a delegated-secretary setting. As a design consequence, scenario families could pass every per-act check while the sequence violates a separately specified standing requirement. Examples are a stated cumulative bound on disclosures, or a pattern of routing one category of request away from timely attention, whose significance needs aggregate professional interpretation. Such cases would be described at case level and reported apart from the opportunity-based rates, which are defined over individual acts. No cumulative-drift metric is specified or validated here. The record available to reviewers must preserve enough of the sequence to read the pattern, which pulls against minimizing the record to limit exposure of confidential material.

Automated checks assess named representation or enforcement properties; incomplete logs and passing checks cannot establish professional adequacy. Tracking concerns counterfactual responsiveness²⁴¹, which recorded behavior alone does not show. Changed-ground scenarios test only a stipulated slice, and reviewers may dispute the relevant reasons²⁴². A stated reason may also omit why an option was preferred to eligible alternatives²⁴³. Such reasons are evidence to assess, not proof of comparative warrant.

Incomplete evaluation output needs an explicit status as well. An unpublished diagnostic from the experimental prototype reports a complete preliminary pass for a candidate reviewed under a reduced-context projection and an incomplete rejection at the output limit for a negative control, with the full-context comparisons not run²⁴⁴. It establishes neither a benefit from projection nor a detection rate. Its narrower lesson is that completion, rejection and an unresolved evaluation are different states, and an incomplete response cannot stand in for an accepted verdict.

10.4 P1: separating proposal from adoption

P1 concerns a package: keeping a proposed assessment distinct from an adopted decision while retaining the adopted scope and grounds. Continuation outside covering authority means a consequential act for which no applicable current decision or covering delegation supplies permission. Reviewers establish the applicable requirements independently of the system, since recorded permission alone does not establish legally adequate authority.

²⁴⁰ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; secs. 3.1,3.4–3.5.

²⁴¹ Santoni de Sio and van den Hoven, “Meaningful Human Control”; Tracking and Tracing subsections.

²⁴² Santoni de Sio and van den Hoven, “Meaningful Human Control”; Tracking subsection.

²⁴³ Kozlovski, “Reasons underdetermination”; Reasons underdetermination in tracing.

²⁴⁴ Author's development record, “Closed review-view diagnostic,” Sep. 23, 2026. Unpublished synthetic-matter engineering note, on file with the author.

In a scenario built on the revised-terms matter, one arm lets the system's assessment Q1 serve both as advice and as the basis for permission, so that a "ready" label carries it past a restart. The other arm keeps Q1 as a proposal until an entitled actor adopts it, and carries the adopted scope and grounds forward. Q1 is a constructed proposal on a reserved decision. The scenario set also needs automatic selections that a valid standing instruction already covers. The separated arm has to consume those under the instruction without a fresh adoption. P1 then rewards neither arm for demanding approval where none is needed, and unnecessary approval requests count as burden.

Both arms receive the same evidence, decision rights, faults, models, tools, resource limits and review time. Where either arm uses a learned or rule-based policy to decide when to pause for review, that policy and the conditions under which it was tuned are declared, because it determines the opportunities for holds. The estimand is the between-arm difference in continuation outside covering authority, reported against the opportunities for which the independent reference prohibits continuation, including required holds, under declared faults and limited reviewer time. The hypothesis is that the package reduces it.

Even a clear result would support less than it seems to. The contrast combines separation with retained scope and grounds, so a difference cannot be attributed to separation alone unless the other features are fixed in both arms or varied by declared ablations one at a time. It would also show only that the package outperforms a deliberately limited baseline. That exposes a failure mode and does not locate the package against a capable alternative. Three comparisons answer three questions and should not be merged.

- A single-assessment or final-review baseline, with its actual capabilities stated, tests the failure mode.
- An implementation built to this paper's semantics tests *semantics-matched conformance*: whether the behavior described in the resumption rows appears and the checks work. Conformance differs from advantage, because a capable workflow built to the same semantics would be expected to match it, and matching specified semantics does not show equivalence under faults.
- A claim of design benefit needs an *independently designed capable comparator*. It is specified by designers who receive the same professional requirements, evidence and scenarios without this paper's vocabulary, and it is assessed on errors, translation mistakes, disagreement and attention cost.

Altering or removing the grounds-linked behavior of the matched implementation does not produce that comparator. It changes the semantics being compared unless another mechanism preserves them, so it is an ablation and is reported as one.

A conventional workflow can be configured to record the distinction. It can use roles, outcomes and attributed history in human-task specifications²⁴⁵, continuing authorization of the kind modeled in usage control²⁴⁶, and case-management manual activation and role-authorized work^{247, 248}. The

²⁴⁵ OASIS, *WS-HumanTask 1.1*, Committee Specification 01; secs. 3.1,4.4,4.11,7.1.5.

²⁴⁶ Park and Sandhu, "The UCONABC Usage Control Model"; sec. 4.2.

²⁴⁷ OMG, *CMN 1.1*; sec. 8.4.2, Tables 8.8–8.9, pp. 115–119.

²⁴⁸ OMG, *CMN 1.1*; sec. 5.2.2, p. 20.

proposal-versus-adoption distinction is an application of those mechanisms. Behavioral equivalence with such a workflow would limit a claim of distinctive control and would decide neither the normative nor the explanatory question. If the separation adds review effort without reducing consequential errors, the practical-advantage claim fails. A lower error count would not show legally adequate judgment or a discharged duty.

10.5 P2: selective holds

P2 concerns what selectivity is worth. A selective hold leaves unconfirmed the dependent consequential effects whose covering decision rests on a changed or unresolved ground, and it lets separately covered work continue. A blanket hold stops the matter. That a selective hold permits more activity is true by definition and would be a progress win by fiat. The proposition concerns what the additional activity is worth once classification errors, coordination and consequences are counted. The estimand is the between-arm difference in useful work retained or reusable once the hold is resolved, observed over a declared window, with discarded work, reconciliation effort and context reacquisition reported beside it. A net-advantage claim would need a declared valuation, since unlike measures are not subtracted from one another. The hypothesis is that selective holds increase retained or reusable work, subject to separately declared limits on impermissible action and on coordination burden.

Fairness depends on shared design. Both arms are drawn from a common scenario population with a common time origin and observation horizon. Holds still unresolved at that horizon, and reuse that cannot be assessed, stay in the population and are handled through reported unresolved outcomes or stated bounds, not by restricting the comparison to resolved cases. A summary conditional on resolution may be reported separately and labeled as such, but it cannot replace the headline comparison or give one arm extra calendar time.

Required judgments and authorized action opportunities are fixed independently of what either arm attempts. The reference distinguishes the actual stipulated dependencies, the selected grounds record and the operational review criteria, so omitted grounds, defective criteria and erroneous coverage classifications can be varied separately. A wrong independence judgment is the characteristic way a selective hold fails, and preparation that is itself consequential is the characteristic way separability is overstated. Response windows, escalation routes and permitted fallbacks are stated in both arms, because a faster return bought by silently waiving a reserved decision is no advantage. Never-reached effects, late responses and impermissible actions remain separate outcomes, as do scenario-level completion and omissions, including any stipulated communication left unallocated. Violation, delay and burden are not combined into one figure.

Selectivity as such has precedent, since case-management stage suspension already propagates to contained work²⁴⁹. The proposed test concerns dependency reliability and coordination burden. A blanket hold may be preferable where dependency information is unreliable, where preparation creates consequential effects, or where stopping separately owned activities is hard to coordinate. A deadline alone supplies no workflow permission, and any protective fallback needs independently applicable authority and not a model's recommendation. A capable conventional workflow with the same scoped instructions may behave identically and more simply, and equivalence would limit a

²⁴⁹ OMG, *CMMN 1.1*; sec. 8.4.2, Tables 8.8–8.9, pp. 115–119.

claim of distinctive mechanism. Routine work within an adequate standing instruction needs no fresh signature at each step, so neither arm should be built to demand one.

10.6 P3 and a first prospective study

P3 claims that the account helps people reason about faults. For equally informative fault cases, organizing the material around the proposed relations may help reviewers identify the affected relation, its version and an appropriate follow-up, with fewer errors or less effort. Those relations are the entitled act, its scope, the consumed interpretation, selected grounds with their reconsideration conditions, the allocated response, and effects with pending tasks. The proposition tests the synthesis as a way of reasoning about design. That is the claim that remains open if a capable workflow matches the mechanism.

The comparison is fair only if information, training and presentation demands are matched and the alternative is as rich as the account. Bare lifecycle terminology would be an artificially limited comparator. A case-management or other capable design framework qualifies if it carries professional authority, information and decision-right concepts of equal richness²⁵⁰. The anticipatory-oversight failure types²⁵¹ and the governed-agency loss-source classification²⁵² are direct diagnostic readings of similar faults. Several causal readings of one fault may be acceptable, and the proposed vocabulary is not the only acceptable wording. Semantic fidelity and detection of relevant change remain assessment targets with no separate directional proposition, and neither can be inferred from an improvement on any of the three.

One plausible first study would test P3 alone. It needs no implementation of the architecture and no integration with practitioners' workflows, and its answer bears on whether the vocabulary is worth the cost of building the rest. What follows is a proposed design. No case set has been written, no study has been run, and no registration is claimed.

Cases and their origin. Cases describe a matter segment in neutral factual terms: what was instructed, what the system applied, what changed and what happened. They supply the information needed for the diagnostic question, including absent or uncertain evidence. They need not instantiate every field of the handshake record.

Independent labeling and independent case construction are different protections. A neutral panel can label fairly cases that the account's author built to display the account's own distinctions, and the study would then mostly measure whether the vocabulary can name what it was made to name. The design therefore specifies two strata in advance and reports them apart.

The first stratum contains synthetic cases whose facts and inclusion and exclusion criteria are developed independently of this framework, without using handshake categories to design or select them. Case facts, inclusion decisions, origin rules and reference requirements are frozen before competing presentations are prepared. Inclusion and exclusion follow those independently set criteria, not either frame's performance. Cases informed by comparator frameworks' failure types are identified by that origin and their results reported separately within the case-origin analysis.

²⁵⁰ OMG, *CMMN 1.1*; sec. 8.4.2, Tables 8.8–8.9, pp. 115–119.

²⁵¹ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; sec. 4.3, pp. 24–25.

²⁵² Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; sec. 4.10, p. 36.

The second stratum holds framework-derived teaching and conformance cases, such as the branches of the revised-terms matter. They are labeled as such. They show that the vocabulary can be applied and taught and that its checks behave as described. They cannot carry a diagnostic-advantage claim, because their facts were built from the distinctions under test.

Both strata include a few cases with no fault and some whose fault admits several defensible readings, so diagnosis is not rewarded for always finding a fault or for matching one reading.

Case-origin analysis. Record authorship, source and any framework that informed each case; report results by stratum. An independent reader using both frames flags cases whose fault only one frame can express. Report such flags rather than remove inconvenient cases after observing results. The proposed analysis can be descriptive; it makes no claim of a completed corpus or a powered interaction.

Reference labels. Labels are fixed before any participant sees a case. A panel of professional reviewers, working independently of one another and of the account's authors, states for each case in plain case-fact terms which act or interpretation failed and against which instruction or requirement. It states to which version or point in time the failure attaches, and which follow-ups are appropriate: rejecting the act, correcting and reconfirming the interpretation, renewing the decision, notifying, reconciling before any retry, or escalating to a named role. It also records follow-ups that would themselves do harm, by permitting a prohibited continuation or imposing an unnecessary hold. The panel then maps both the labels and the follow-ups into each frame's vocabulary, so that a reviewer using either frame can state a finding and a follow-up without borrowing the other's terms.

Where panel members disagree, a predeclared adjudication procedure distinguishes justified alternatives within professional discretion from factual error, unclear requirements and unresolved disagreement. The accepted readings and their reasons are recorded before scoring, so neither a majority vote nor one member's assertion establishes correctness. Unresolved cases are analyzed separately and not silently dropped.

Participants and conditions. Recruit plausible diagnostic reviewers, including supervising attorneys and legal-technology or compliance reviewers. Record legal/technical expertise and frame familiarity before assignment, balance them across conditions and report subgroup results where feasible. Expertise effects can be described without claiming power for an interaction.

The conditions present the same case facts, at the same length, under different diagnostic organizations. The primary comparator is the governed-agency loss-source classification, which classifies loss relative to independently established authority by mandate, authority, information, coordination and adaptation²⁵³. It comes closest to the account in what it asks reviewers to examine. The anticipatory-oversight failure types, which separate specification, inspection, translation and compliance failures²⁵⁴, would serve as a second comparator. A presentation of the same facts with no organizing scheme would be a floor, and a gain over the floor alone cannot establish advantage. The primary comparator is fixed before data collection. Training for each frame is equal in time and worked examples and is prepared by someone other than the account's authors. A pilot checks that

²⁵³ Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; sec. 4.10 (Defs. 4.20–4.22), p. 36; sec. 7.1.

²⁵⁴ Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; sec. 4.3, pp. 24–25.

case descriptions do not borrow one frame's terms. Where the account's diagrams are part of its presentation, the comparator receives matched-information diagrams of its own. Cases and frames are counterbalanced across participants and order, and repeated observations from the same participant and case are modeled explicitly.

Measures. Four things are assessed separately, and none is folded into a composite.

- *Label accuracy*: whether the reviewer's identification of the failed relation, its location and its temporal reference lies within the acceptable set.
- *Follow-up appropriateness and harm*: whether the proposed follow-up is among the acceptable ones, and whether it would itself permit a prohibited continuation or impose an unnecessary hold.
- *Effort*: time to diagnosis, dependencies missed, false alarms on the no-fault cases, reading and training time, and reported workload if an instrument is used.
- *Practical diagnostic value*: a separately stated judgment of whether any accuracy or follow-up gain justifies the added reading and training cost, under a decision rule declared in advance.

Coding and unblinding. Coders are kept unaware of assigned condition where feasible. Remove frame labels only where meaning is preserved, independently double-code a sample and report how often responses reveal the condition. Compare evident-condition responses with the others; do not assume effective blinding from nominal masking.

What would count. No accuracy advantage over the primary comparator would fail to support the primary hypothesis. A gain only over the floor would not establish advantage against a capable frame. A gain confined to the framework-derived stratum would support teaching and conformance, and not diagnostic advantage. Effort savings at comparable accuracy would support a different, predeclared efficiency claim. Added reading or training time is reported separately, with a stated decision rule for whether that cost offsets the gain. An advantage that disappears with neutral case wording, depends on labels drawn from this account, or cannot be reproduced by reviewers applying the vocabulary consistently would undermine the corresponding diagnostic claim.

A positive result would support diagnostic use among reviewers of the sampled synthetic cases. It would not establish improved legal outcomes, adequate professional judgment or implementation benefit. The development environment contains no practitioner observations for this study. Sample size and meaningful effect thresholds require pilot evidence; the study remains proposed.

10.7 Burden, capacity and reliance

Grounds selection, updating, translation review, reconciliation, escalation and supervision impose professional work. Assessment must weigh that work against the relations the structure makes visible. Simulated operators cannot measure practitioner burden; sampling and instrument choices require pilot evidence from experienced professionals.

Candidate observations include time spent resolving a boundary, interruptions, rework, delay and workload reports, with denominators and missing observations stated. The human-AI interaction

guidelines motivate examining handoff intelligibility and interaction cost²⁵⁵. They do not validate these measures or establish an optimal interruption rate. NASA-TLX is a candidate workload procedure. Its Version 1.0 manual describes six subjective subscales and an original weighted procedure²⁵⁶. Any variant, altered wording or timing would be named and predeclared, since dropping the weights cannot silently retain the original scoring claim, and the subscales and missing responses stay visible and are not absorbed into a total. Policy load varied for an agent²⁵⁷ is a different variable from a professional's workload.

Burden is distinct from appropriate reliance. Automation-allocation research already identifies situation-awareness, complacency and skill risks where a person must take over²⁵⁸. Fewer approvals and lower reported workload do not establish preserved awareness. Candidate observations therefore include whether professionals identify the stipulated detector limits, notice an omitted or missed grounds change, question an unconfirmed or misleadingly continuing status, and retain the context needed to take over, each against an independently defined task reference. A confirmation counts as review only where the professional has the capacity, information, time and ability to challenge that make it meaningful²⁵⁹. The critique of nominal oversight in government algorithms is a reason to measure that capacity in the task and not to presume it²⁶⁰.

A hold is an intervention whose value depends on the professional who receives it. Learned-deferral work makes a parallel point in a different setting: whether a system should answer or pass a case depends on the downstream decider's modeled performance, so the combined system is what gets evaluated²⁶¹. Its experiments use simulated deciders in labeled prediction tasks with accuracy and fairness objectives, which differ from professional judgment under permissible discretion, and its loss-based framing is not adopted here. The transferable point is that the effect of a hold is observed in the professional's use of it and is not read from routing statistics. Distribution of routing and delay across declared matter or client categories could expose uneven access to timely judgment, provided case mix and reference uncertainty are reported. A difference alone does not establish unfairness. A participant study would use matched tasks in randomized or counterbalanced conditions, with order, learning and carryover treated explicitly and withdrawals and missing ratings reported. Lower experienced workload cannot by itself establish better judgment.

10.8 Where the claims stop

Each limit below bounds a particular claim.

Observation. A system can respond only to changes it can observe. A missing event in its own trace is not evidence that the world stood still, and an external ledger resolves that problem in a synthetic study without resolving it in practice. Any claim that a detector or hold caught what it should have caught is conditional on the declared observation reach. Some limits are properties of a task and can

²⁵⁵ Amershi et al., “Guidelines for Human-AI Interaction”; Table 1, PDF p. 3.

²⁵⁶ NASA Ames Research Center, *NASA Task Load Index (NASA-TLX), Version 1.0: Paper and Pencil Package*, n.d., official manual rehost, accessed Oct. 3, 2026. Pinpoints use the manual's printed pagination; secs. 2–3.

²⁵⁷ Levy et al., “ST-WebAgentBench”, arXiv:2410.06703v1; secs. 3.3,4.2.

²⁵⁸ Parasuraman, Sheridan and Wickens, “A Model for Types and Levels”; sec. IV.B.2, pp. 292–293.

²⁵⁹ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; pp. 248–249.

²⁶⁰ Green, “The flaws of policies requiring human oversight”; secs. 4.1,5.1.

²⁶¹ Madras, Pitassi and Zemel, “Predict Responsibly”; secs. 2.1–2.4, pp. 2–4.

be measured there: how much relevant change a detector can see, how often a translation departs from the instruction, how reliable dependency classification is, how long a hold waits for a response. The account says where to measure and supplies no value for any of them.

Enforcement. The claim that a hold or an effect check controls an act depends on stated assumptions about guarded effect paths, authorized commits, declared handling of unknown effects, and the ordering of the authoritative-state check against the effect commitment. Prevention needs that ordering, or a service-enforced conditional commitment with the same semantics, so that a relevant change ordered before commitment causes refusal or a recheck. A read shortly before sending narrows the interval and does not prevent a stale effect. A carried version identifier supports detection and nothing more. Where an assumption fails, the relation can still be stated and assessed, and enforced control cannot be claimed.

Technical enforcement does not supply the institutional arrangement it presupposes: who owns the review criteria, who is a competent alternate, and how a pending question reaches an entitled actor in time. Effects that have occurred cannot necessarily be undone by a later hold. Staffing and procedure can defeat timely adjudication even where the interface is adequate.

Representation. The grounds record is a fallible selection. An omitted dependency may matter, a recorded reason may be misunderstood, and a populated record can produce false assurance or ritualized compliance. Recording every reason in machine-readable form may expose sensitive material or impose disproportionate burden, so a grounds link may refer to protected material or to a category of condition, with its permitted detail depending on the setting. The account offers no categorical solution to privilege, confidentiality or legal delegation. Any use of real matters would need separate analysis of consent, confidentiality, privilege and evaluator access.

Instruction and fidelity. A stored instruction and a passing validator do not show that every permissible interpretation of a deliberately open instruction was preserved or that every harmful action was excluded. Fidelity to a misspecified instruction is not adequacy. A faithful implementation of an instruction that omits a required consultation, or is itself impermissible, remains deficient.

Capacity. Professional skill, availability, supervision arrangements and client decision rights lie outside a record's power to confer. An approval event can be preserved while competence, information, time or the means of challenge is missing.

Scope and architecture. The five computational functions with professional judgment are an analytical allocation, not an exhaustive taxonomy or a requirement for separate services, and any schema is one operationalization. The New York rules are selected illustrations and not uniform national doctrine or a compliance finding, and the advisory opinions remain advisory. Nothing here shows that this architecture, or any architecture, is necessary or sufficient for professional agentic work in general. A capable governed workflow that expresses the same relations limits a claim of distinctive mechanism and leaves in place the synthesis's explanatory purpose, which P3 and the first study test.

The case and method behind these claims, including what the development notes do and do not show, come next.

Table 9. Prospective propositions, capable comparisons and possible defeat conditions. Designs are compared with matched information, constraints and reference judgments. Diagnostic usefulness, practical benefit and conformity to a specified relation are separate questions; no result or effect size is supplied.

Proposition	Hypothesized effect	Fair comparison	Observations / defeat
P1: proposal and adoption package	Separation plus retained scope and grounds may reduce continuation outside covering authority under faults and limited review time; the estimand is the arm difference over opportunities where the reference prohibits continuation.	Failure-mode comparison: limited single-assessment baseline. Advantage comparison: equally capable governed workflow with the same evidence, decision semantics, models/tools and reference judgments; removing grounds-linked behavior changes the matched semantics and is an ablation.	Unsupported continuations, consequential errors, translation errors, disagreement, attention. Behavioral equivalence limits distinctive-control claims; added effort without fewer consequential errors defeats that advantage claim.
P2: selective holds	More useful work retained or reusable after resolution, over a declared window, with discarded work, reconciliation and context reacquisition reported separately, subject to authority and coordination limits.	Common scenario population, time origin and horizon; explicit treatment of unresolved holds and unassessable reuse. Predeclared change/fault/availability frequencies, mistaken coverage or reserved-decision classification; fixed required judgments and action opportunities. Compare selective and blanket holds without awarding a definitional progress win.	Retained and discarded work, reconciliation and context reacquisition; scenario-level completion and omissions with opportunity denominators; late responses, unrealized opportunities, impermissible acts, delay and burden separately. Unreliable dependencies, classification/coordination costs or consequential preparation may defeat net advantage. A capable workflow may match the same selective relations; equivalence limits distinctiveness.
P3: diagnostic organization	Responsibility/ authority/ dependency relations may help identify a fault and appropriate follow-up.	Equal information, training and presentation demands; equally rich causal readings, including anticipatory-oversight failure modes ²⁶² and governed-agency loss sources ²⁶³ .	Identification accuracy, follow-up quality, effort. No improvement or reading/training costs offsetting gains defeats the corresponding usefulness claim.

11 Method and generative case

The account rests on three kinds of ground, and their order matters. The first is theory: definitions, relations and consequences argued from the problem of professional co-control, each tested by whether it changes what a design would store, check or adjudicate. The second is prior work on

²⁶² Baum et al., “Anticipatory Human Oversight”, arXiv:2609.24242v1; sec. 4.3.

²⁶³ Dias Ferreira, “When Intelligence Becomes Agency”, arXiv:2609.07741v1; sec. 4.10.

automation allocation, workflow, usage control, shared control, oversight and governed agency, read at the scope its authors give it. It supplies mechanisms and credit, and it sets the standard any distinctive claim has to meet. The third is a synthetic generative environment, an experimental legal-work prototype whose development notes first made certain design questions concrete. That environment raises questions. It does not support the first two grounds, and the account's claims about entitlement, client rights and renewed judgment do not depend on it.

11.1 Kinds of contribution and kinds of warrant

The paper is a conceptual architectural synthesis with prescriptive design propositions. Gregor distinguishes theory types by purpose (analysis, explanation, prediction, explanation with prediction, and design and action) and ties the assessment of a contribution to its type²⁶⁴. Gregor and Jones list the components of a design theory: purpose and scope, constructs, principles of form and function, anticipated mutability, testable propositions, justificatory knowledge, implementation principles and expository instantiation^{265,266}.

Here, recurring professional co-control supplies purpose and scope; responsibilities, decisions, grounds, interpretations and coverage supply constructs. The minimum relation and its consequences supply design principles; revision supplies mutability. Contracts and the illustrative record suggest implementation principles, P1–P3 supply tests, prior scholarship supplies justification and the constructed matter supplies exposition. Gregor and Jones permit design theory without a working instance while recognizing the credibility an implementation may add. Classification itself establishes no validity.

Five kinds of warrant follow, and each claim in the paper should be read with the kind it needs.

- *Definitional and relational claims* hold that responsibilities differ in what warrants their outputs, or that a scoped decision stands in a stated relation to its grounds. They are justified by clarity and coherence, and tested by whether the distinctions change a design choice.
- *Conditional operational claims* hold that a specified relation obtains between an observed change, a decision and an affected action. They are tested by examining behavior under stated assumptions. Writing a rule does not show that it is enforced, and enforcement would not show practical advantage.
- *Comparative-advantage claims* hold that one design outperforms an equally capable alternative on a stated measure. They need comparative evidence, and none is offered.
- *Normative claims about professional authority* rest on scoped U.S. sources that explain why judgment, information and client authority matter at boundaries. They do not show that any record satisfies a duty.

²⁶⁴ Gregor, “The Nature of Theory”; pp. 619–620, 624–629.

²⁶⁵ Gregor and Jones, “The Anatomy of a Design Theory”, Mar. 22, 2007 author manuscript; manuscript pp. 26–28, 37–38, 41–44.

²⁶⁶ Gregor and Jones, “The Anatomy of a Design Theory”, Mar. 22, 2007 author manuscript; Table 2 and discussion, manuscript pp. 26–28.

- *Reported observation* has three separable parts: what a development note says occurred, what the account infers from it, and what architectural response it proposes. The text keeps the three apart.

11.2 The documentary comparison

The comparison with prior work is documentary and bounded. Selected passages of control-allocation, human-task, usage-control, clinical decision-support, institutional-power, case-management and governed-agency sources are read for what their constructs let a design express. Findings are limited to what the inspected passage supports. A construct a passage does not mention is not thereby absent from a system or a field, and the inspection does not reach deployed behavior.

The comparison follows a strongest-rival discipline. The rival is a governed workflow combining durable decisions, role-based permissions, exception handling, attributed history and continuing authorization^{267,268}, with its professional semantics represented at their strongest and not assumed absent. Where the rival can express a relation, the finding limits any claim that a mechanism is distinctive. It leaves open whether the synthesis explains a failure better, whether the allocation is justified and whether it is worth its cost. Each of those is argued and tested at its own level.

The revised-terms matter used throughout is constructed for exposition and is not a reconstruction of any prototype trace. Where a development note motivates a design pressure, the text says so. Where a branch is constructed, it is labeled as constructed.

11.3 The generative environment and its limits

The experimental prototype was built to prepare and review draft documents for synthetic U.S. matters. A simulated operator, implemented through a development agent, supplied the scope-confirmation and release events. No attorney or client exercised the judgments this paper discusses, and the prototype is not offered as an implementation of the handshake.

The evidence comprises four author-written engineering notes from September 22–24, 2026, read in full and retained unpublished on file. Raw traces and artifacts were not independently reconstructed. Reported acceptance criteria and checks are therefore authored observations, not an independent reconstruction or practitioner study. Shared authorship and setting limit corroboration and frequency claims, and private access limits inspection.

The four were selected purposively for architectural relevance, documentary specificity, contrast among competing explanations and the limits they report. The criteria were stated before full-text retrieval but after exposure to the notebook index and earlier summaries, so the selection is retrospective. The subset is neither preregistered nor saturated, and no inventory of excluded notes is established. Successful paths and failed continuations are both retained. The notes illustrate selected design pressures and do not reconstruct how the theory emerged or who first proposed any element of it.

²⁶⁷ OASIS, *WS-HumanTask 1.1*, Committee Specification 01; secs. 3.1,4.9–4.11,7.1.5.

²⁶⁸ Park and Sandhu, “The UCONABC Usage Control Model”; sec. 4.2.

11.4 Report, pressure and proposal

Three examples show the separation the method asks for. Each gives what a note reports, the design pressure the account draws from it, and the architectural response proposed.

Receipt versus semantic acceptance. A note of September 24, 2026 closes a batch of original synthetic matters with the summary that receipts passed and semantic classification failed²⁶⁹. In one run a source-stated hosting location stayed in the draft's prose but was classified as missing in the structured output. The model reviewer recognized the misclassification and still treated it as non-blocking, and a separately run project stage gate then rejected the ready file. The repair step was not invoked, so the run does not show whether repair would have succeeded, and the note does not isolate the effect of the competing instructions the components had been given. The fact was not lost everywhere.

The pressure the account draws is that a valid receipt, a reviewer's severity judgment and a gate's rejection answer different questions, and a later gate is not bound by an earlier judgment that a defect was minor. The proposal is that each check name the condition it establishes and the artifact version it was run against, so that no pass serves as a general warrant for continuation. The same separation reappears at the effect boundary, where a fluent restatement of an instruction can pass while the guard actually consumed is broader. The overbroad guard I1 extends this pressure as a constructed case and was not observed in any note.

Preserved history and unfinished work. A note of September 23, 2026 on a supervised qualification run reports an initially accepted draft package followed by an unaccepted continuation on a synthetic U.S. matter²⁷⁰. The first package used a protected scope interface and a host-built plan. Two documents were reviewed and accepted for preliminary preparation despite disclosed wording and provenance-label defects. Scope confirmation and release were simulated operator acts, and the reported send was simulated. After recovery the note reports no duplicate send. The follow-up exhausted its review output before an accepted verdict, and no follow-up package or delivery was created, while the earlier accepted output stayed readable. The note does not isolate why the follow-up failed, no second matter was run, and no conclusion about context size follows.

The pressure drawn is narrower. Accepted history has to be preserved, unfinished work has to stay distinct from accepted state, and preserved output supplies no authority for a later continuation. The proposal is to separate task status from adopted decision in matter state and to check resumption at the effect boundary, so that an incomplete follow-up is pending work and not approved work. The note shows nothing about legal advice of tested quality, actual delivery or an attorney's handshake.

A bounded decision and its verification. A note authored September 22, 2026 and filed the next day compares two routes for a bounded review-decision task²⁷¹. The task was whether to act on a supplied criticism of a draft, by upholding it, dismissing it or leaving it uncertain. It was not a task of finding all defects or approving an artifact. The comparison reports well-formed responses alongside

²⁶⁹ Author's development record, "Original-case batch closed: receipts pass, semantic classification fails," Sep. 24, 2026. Unpublished synthetic-matter engineering note, on file with the author.

²⁷⁰ Author's development record, "Closed supervised qualification: initial success and follow-up failure," Sep. 23, 2026. Unpublished synthetic-matter engineering note, on file with the author.

²⁷¹ Author's development record, "Closed Qwen/Jev review-decision comparison," authored Sep. 22, 2026, filed in the workspace notebook Sep. 23, 2026. Unpublished engineering note, on file with the author.

unsafe dismissals of supplied criticisms. It was a small development task with project-assessed labels, no external panel or uncertainty calibration, and no runtime adoption.

The inference is that structural validity and the soundness of an adjudicative choice need different warrants. Testing a supplied criticism is verification. Choosing among uphold, dismiss and uncertain is a bounded decision-model function that selects or routes, and a component can discharge it without acquiring professional authority. The account therefore proposes that such a recommendation be usable within an independently authorized branch, and that a valid format create no authority. No ranking of the routes and no accuracy estimate for legal production follows.

Taken together, the observations motivate continuity across interruption, differentiated checks and the separation of advice from authorized adoption. They do not show that the proposed allocation improves outcomes, and none reports judgment exercised by an entitled attorney or client. Entitlement, client constraints and recurring judgment are justified through prior work, scoped professional-responsibility analysis and architectural argument. A fourth note, a review-view diagnostic, is reported with the assessment chapter because it bears on the status of incomplete evaluation output.

12 Conclusion

Professional agentic work requires a decision to keep governing the work that follows it. An attorney may give an adequate instruction, yet software can apply a broader permission; an instruction may remain faithfully represented, yet changing circumstances can make its use inappropriate. The architectural question is how to preserve that relation while people, facts and systems change. The handshake answers it through recurring exchanges in which the system presents relevant state and uncertainty, the professional makes an attributed decision within their authority, and later work remains connected to its scope, interpretation and conditions for reconsideration.

The contribution is a synthesis with explicit consequences for design, review and assessment. It identifies which relations must remain recoverable, explains what fails when they are lost and allows simpler arrangements that preserve them. It separates errors that a single approval or completion label would conceal. It also defines comparisons that can test usefulness without mistaking conformity to the proposed architecture for an advantage over a capable alternative. These contributions concern how professional decisions govern continuing work; their practical benefits remain questions for the proposed studies.

12.1 What a design has to preserve

The allocation begins with the warrant for each output. Reasoning supplies reasons; computation supplies a result under stated inputs and assumptions; a decision model compares, selects or routes alternatives under stated objectives. Execution produces effects, while verification tests a named property. Treating these outputs as interchangeable hides what the next step can rely on. The lapse-date example shows the problem: two calculations can agree while sharing the wrong triggering event. The arithmetic may be sound and its use inappropriate. Similarly, a ranking may apply a weighting no entitled person adopted. The design must preserve the assumptions and criteria that determine what each result is fit for.

Bounded contracts make those distinctions inspectable at the points where outputs are consumed. A micro-harness states the task, dependencies, permitted operations, check and output warrant for a unit of work. Its value comes from clarifying an interface or directing attention to a consequential failure. Smaller units alone do not establish reliability, and extra handoffs can introduce translation errors. Where review of the final artifact detects the same faults at no greater cost, the simpler arrangement remains a serious option.

Above these interfaces sits the five-role minimum: an attributed act under independently grounded entitlement, applicable scope, a link to the interpretation actually consumed, conditions for reconsideration and a responsible response to unresolved coverage. The removal arguments explain why each matters for consequential work over time under revisable delegation. They do not prescribe five services or a universal record schema. A direct operation may make the relations evident in the act; an adequate standing instruction may cover many later choices. The selective hold is one response to unresolved coverage, alongside an entitled alternate, a covering rule, a bounded fallback or ending the task.

Durable matter state carries the relations through interruption and succession. Persistence preserves recorded content; semantic continuity concerns what still governs the work. A surviving proposal is still a proposal, a completed draft is not an adopted instruction, and an old check does not cover a revised artifact. Selected grounds connect a decision to facts, interpretations and conditions that may change. Those records are necessarily fallible: an omitted dependency can matter, and recording more detail can widen confidentiality exposure. The design needs enough information to identify relevant change without assuming that every reason should enter a model context or log.

The coordination analysis identifies another failure between the recorded instruction and its effects. Capture, communication, operative interpretation and observation are separate achievements. A fluent paraphrase can agree with the saved instruction while the executable guard permits something the professional withheld. An independent action test or inspection of that guard can reveal the mismatch. This is why fidelity must refer to the instruction actually issued, the interpretation applied and the acts it could permit. A faithful implementation can still carry out a deficient instruction. A later authorization can govern a new act without repairing an earlier translation or erasing its effects.

At resumption, the relevant check is tied to the actual artifact, recipient, instruction and effect boundary. Prevention requires stronger conditions than a version record: the authoritative-state check and external effect commitment must be ordered together so a relevant recorded change before commitment causes refusal or a new check, or an external service must enforce equivalent conditions. Reading shortly before sending may shorten the interval; carrying an identifier can support later detection. Neither alone prevents a stale effect. Under weaker arrangements, the design must allocate detection and reconciliation, without promising that every mismatch will be discovered. An unobserved change may remain unseen, and an irreversible effect cannot be prevented after it has occurred.

Recovery also has to preserve unknown outcomes. A timeout does not establish that nothing reached the recipient. Deduplication may prevent a repeated transmission without establishing that the first one was authorized. Reconciliation and retry therefore need their own stated conditions. The same discipline governs selective continuation: other work can proceed only under its own covering

authority and where the unresolved issue does not propagate through shared constraints, disclosure, artifacts or timing. Preparation itself may be consequential.

12.2 What professionals and institutions have to supply

Moving from human in the loop to professional in the loop means specifying who can exercise the judgment required at a particular boundary. Expertise, entitlement, information, time and a practical means of challenge have different sources. A credential cannot settle them all. A person may occupy a causal position without the capacity or standing that makes the position adequate^{272,273}. Meaningful confirmation requires an opportunity to understand and challenge what governs the work²⁷⁴. The critique of nominal oversight in government algorithms reinforces the need to assess that capacity in the actual task rather than infer it from participation²⁷⁵.

Selected professional-responsibility rules explain why these distinctions have architectural consequences. A client-held decision limits what the attorney may instruct; an attorney's technical ability to send does not enlarge that right. The institution determines which work the system performs, which standing delegations and review criteria apply, and who supervises them. These choices should remain visible for review instead of disappearing behind the final approver. Professional judgment enters through matter decisions and continuing supervision, while its duties remain attached to the responsible people and institutions.

Required conduct must remain visible alongside restrictions. Holding an unauthorized transmission does not perform a required notification. A pending communication needs an owner, a due condition and evidence of performance; recovery or reassignment must preserve its status. Late delivery remains late. A protective fallback can address urgency only where independent authority covers it. The architecture can represent these conditions and expose omissions, but a completed workflow step cannot establish comprehension, adequate judgment or the discharge of a professional duty.

Succession similarly separates the ability to perform work from the decision that governs it and the authority to renew it. An alternate may continue covered preparation while lacking authority to resolve a reserved decision. Departure can end a person's entitlement to make further decisions without ending the force of an earlier instruction. Where client, professional and institutional directions conflict, the applicable sources must determine precedence. Escalation can legitimately end with no authorized continuation of the requested act.

These arrangements impose work: selecting grounds, reviewing translations, maintaining criteria, answering questions, reconciling effects and supervising detectors. Fewer approval requests may reduce interruption while encouraging misplaced reliance on unflagged continuation. Institutions must consider that trade-off alongside staffing and confidentiality. The architecture succeeds as a design account only if those costs remain visible and the professional has a realistic opportunity to exercise the judgment assigned to them.

²⁷² Baum and Laux, “Constitutive vs. Corrective”, arXiv:2603.19213v2; secs. 2.1–2.2, Defs. 1–2.

²⁷³ Baum and Laux, “Constitutive vs. Corrective”, arXiv:2603.19213v2; sec. 1, final qualification.

²⁷⁴ Cavalcante Siebert et al., “Meaningful human control: actionable properties”; pp. 248–249.

²⁷⁵ Green, “The flaws of policies requiring human oversight”; secs. 4.1, 5.1.

12.3 What an assessment has to hold apart

The four assessment targets follow from the distinctions above. Outcomes concern the work product and end state. Authority-sensitive trajectories concern consequential acts and required conduct over time. Fidelity concerns the instruction as issued at each version. Professional burden concerns the attention and maintenance the arrangement demands. Their references must be specified independently of the system under test. A good end state cannot excuse an unauthorized route, and later corrected state cannot rewrite an earlier mismatch. Rates need scenario-based denominators that retain unresolved holds, omitted communications and opportunities never reached.

Prior control theories and workflow standards make the capable comparator concrete. Shared control, human tasks, ongoing authorization, clinical decision state, governed agency and anticipatory oversight supply substantial parts of the architecture. The synthesis brings their implications together around a matter-specific professional decision. CMMN, DMN and SCXML can carry many of the required relations when an application supplies entitlement, instruction meaning, version binding and effect semantics. A conventional realization that preserves the same relations should be assessed on those terms. It cannot be dismissed because it uses a different vocabulary.

That comparison separates the purposes of P1, P2 and P3. A deliberately limited baseline can expose a failure mode; an implementation with matched semantics can demonstrate conformance. Practical advantage requires an independently designed capable alternative. Selective holds should be judged by retained or reusable work, errors, delay and coordination burden over shared horizons, rather than by the additional activity they permit by definition. Diagnostic usefulness asks whether reviewers identify faults and appropriate follow-up more accurately or with less effort under equally informative presentations.

The proposed first study addresses that last question. Its cases need independent construction as well as independent labeling, with facts and selection fixed before competing presentations are prepared. Cases derived from the paper remain useful for teaching and conformance, with results reported separately. No-fault and ambiguous cases, multiple defensible readings and strong comparator frameworks prevent matching the paper's terminology from becoming the definition of success. Label accuracy, harmful follow-up, effort and practical diagnostic value remain separate findings. A gain in one does not establish a gain in the others.

The method keeps the evidentiary scope equally clear. The constructed matter tests the coherence and consequences of the argument. Selected unpublished notes from a synthetic development environment motivate questions about acceptance, continuity and computational warrant; they contain no trial of attorney judgment and do not validate the architecture. The documentary comparison identifies constructs and scoped relationships, not deployed-system performance. An empirical claim about professional outcomes or burden therefore requires evidence beyond the material presented here.

12.4 The enduring relation

The handshake makes a professional decision an object of continuing architectural attention. It asks whether the work remains governed by the decision that was made, whether the interpretation applied expresses it, whether relevant changes require reconsideration and who is responsible for

what remains unresolved. That account preserves room for delegated initiative while keeping the conditions of professional judgment visible.

An attorney returning to a matter should be able to understand what was decided, what it still covers, which interpretation is in force, what changed and what is waiting on whom. The system should present the information needed to challenge or redirect the work, and the institution must supply the time and authority to use it. Attorney $\leftrightarrow$ system $\leftrightarrow$ attorney $\leftrightarrow$ system describes a relation sustained throughout the matter: a decision that remains attributable to someone entitled to make it, and work that remains answerable to that decision.

References

1. K. Baum, M. Kiener, M. Langer and J. Laux, “Anticipatory Human Oversight of Agentic AI: A Philosophical Account,” *arXiv:2609.24242v1*, Sep. 21, 2026, version 1.
2. J. Dias Ferreira, “When Intelligence Becomes Agency: A Theory of Governed, Proactive Agency for Symbiotic AI Systems,” *arXiv:2609.07741v1*, Sep. 7, 2026, version 1.
3. W. C. Chesher, “The Handshake Protocol: Protecting Expert Judgment in Sustained Human-AI Collaboration,” manuscript dated Mar. 31, 2026, doi: 10.6084/m9.figshare.31914687.
4. S. Gregor, “The Nature of Theory in Information Systems,” *MIS Quarterly*, vol. 30, no. 3, pp. 611–642, Sep. 2006, journal article in the ANU repository.
5. S. Gregor and D. Jones, “The Anatomy of a Design Theory,” author manuscript, Mar. 22, 2007, ANU repository copy. Pinpoints refer to this manuscript's pagination.
6. Object Management Group, *Decision Model and Notation (DMN), Version 1.5*, formal/24-01-01, cover dated Jan. 2024, official specification.
7. 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, May 2000, doi: 10.1109/3468.844354.
8. D. Madras, T. Pitassi and R. Zemel, “Predict Responsibly: Improving Fairness and Accuracy by Learning to Defer,” in *Advances in Neural Information Processing Systems*, vol. 31, 2018, publisher PDF.
9. A. Kozlovski, “Reasons underdetermination in meaningful human control,” *Ethics and Information Technology*, vol. 27, article 59, Oct. 15, 2025, doi: 10.1007/s10676-025-09858-x.
10. American Bar Association Standing Committee on Ethics and Professional Responsibility, *Formal Opinion 512: Generative Artificial Intelligence Tools*, July 29, 2024, official opinion. Advisory interpretation of ABA Model Rules.
11. New York Rules of Professional Conduct, 22 N.Y.C.R.R. Part 1200, court-hosted NYSBA compilation, rules and comments stated as amended through Nov. 10, 2025; selected provisions and later amendments index checked Oct. 4, 2026. Accompanying NYSBA comments are not adopted rules. Cited as a New York illustration, not a comprehensive currentness certification.
12. F. Santoni de Sio and J. van den Hoven, “Meaningful Human Control over Autonomous Systems: A Philosophical Account,” *Frontiers in Robotics and AI*, vol. 5, article 15, Feb. 28, 2018, doi: 10.3389/frobt.2018.00015.
13. G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman and P. J. Feltovich, “Ten Challenges for Making Automation a ‘Team Player’ in Joint Human-Agent Activity,” *IEEE Intelligent Systems*, vol. 19, no. 6, pp. 91–95, Nov./Dec. 2004, doi: 10.1109/MIS.2004.74.

14. L. Cavalcante Siebert et al., “Meaningful human control: actionable properties for AI system development,” *AI and Ethics*, vol. 3, pp. 241–255, 2023, online May 18, 2022, doi: 10.1007/s43681-022-00167-3.
15. F. Flemisch, M. Heesen, T. Hesse, J. Kelsch, A. Schieben and J. Beller, “Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations,” *Cognition, Technology & Work*, vol. 14, pp. 3–18, 2012, online Nov. 18, 2011, doi: 10.1007/s10111-011-0191-6.
16. J. Park and R. Sandhu, “The UCONABC Usage Control Model,” *ACM Transactions on Information and System Security*, vol. 7, no. 1, pp. 128–174, Feb. 2004, doi: 10.1145/984334.984339.
17. P. Scerri, D. V. Pynadath and M. Tambe, “Towards Adjustable Autonomy for the Real World,” *Journal of Artificial Intelligence Research*, vol. 17, pp. 171–228, 2002, doi: 10.1613/jair.1037. Cited copy: arXiv:1106.4573v1, deposited June 22, 2011; pinpoints follow journal pagination.
18. I. Levy, B. Wiesel, S. Marreed, A. Oved, A. Yaeli and S. Shlomov, “ST-WebAgentBench: A Benchmark for Evaluating Safety and Trustworthiness in Web Agents,” *arXiv:2410.06703v1*, Oct. 9, 2024, version 1.
19. Object Management Group, *Case Management Model and Notation (CMMN), Version 1.1*, formal/16-12-01, Dec. 2016, official specification. Pinpoints follow specification pagination.
20. OASIS, *Web Services—Human Task (WS-HumanTask) Specification Version 1.1*, Committee Specification 01, Aug. 17, 2010, HTML rendition. The specification designates its .doc rendition as authoritative.
21. B. Meyer, “Applying ‘Design by Contract,’” *Computer*, vol. 25, no. 10, pp. 40–51, Oct. 1992, doi: 10.1109/2.161279.
22. L. de Alfaro and T. A. Henzinger, “Interface Automata,” in *Proceedings of ESEC/FSE 2001*, Vienna, 2001, pp. 109–120, doi: 10.1145/503209.503226.
23. W3C, *State Chart XML (SCXML): State Machine Notation for Control Abstraction*, Recommendation, Sep. 1, 2015, dated edition.
24. D. R. Sutton and J. Fox, “The Syntax and Semantics of the PROforma Guideline Modeling Language,” *Journal of the American Medical Informatics Association*, vol. 10, no. 5, pp. 433–443, Sep./Oct. 2003, doi: 10.1197/jamia.M1264.
25. S. Miksch, Y. Shahar and P. Johnson, “Asbru: A Task-Specific, Intention-Based, and Time-Oriented Language for Representing Skeletal Plans,” in *7th Workshop on Knowledge Engineering: Methods & Languages (KEML-97)*, Milton Keynes, 1997, author copy. Pinpoints use the cited copy's pagination.
26. New York City Bar Association Committee on Professional Ethics, *Formal Opinion 2024-5: Ethical Obligations of Lawyers and Law Firms Relating to the Use of Generative Artificial Intelligence in the Practice of Law*, 2024, official opinion. Advisory guidance.
27. *Davidson v. Metropolitan Transit Authority*, 44 A.D.3d 819 (N.Y. App. Div. 2d Dep’t 2007), 2007 N.Y. Slip Op. 07814, official opinion.
28. Noam Kolt, “Governing AI Agents,” *Notre Dame Law Review*, vol. 101, pp. 335–386, 2026, published article.
29. B. Green, “The flaws of policies requiring human oversight of government algorithms,” *Computer Law & Security Review*, vol. 45, article 105681, 2022, doi: 10.1016/j.clsr.2022.105681.
30. T. South, S. Marro, T. Hardjono, R. Mahari, C. D. Whitney, A. Chan and A. Pentland, “Position: AI Agents Need Authenticated Delegation,” in *Proceedings of the 42nd International Conference on Machine Learning*, PMLR, vol. 267, pp. 82211–82231, 2025, publisher article.

31. K. Baum and J. Laux, “Constitutive vs. Corrective: A Causal Taxonomy of Human Runtime Involvement in AI Systems,” *arXiv:2603.19213v2*, Aug. 20, 2026, version 2.
32. J. Hollan, E. Hutchins and D. Kirsh, “Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research,” *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 2, pp. 174–196, June 2000, doi: 10.1145/353485.353487. Cited author final revision, Apr. 23, 2000; revision pagination differs from the published article.
33. E. Horvitz, “Principles of Mixed-Initiative User Interfaces,” in *Proceedings of CHI '99*, ACM Press, 1999, pp. 159–166, doi: 10.1145/302979.303030. Cited author manuscript; manuscript pagination differs from the published article.
34. M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, M. B. van Riemsdijk and M. Sierhuis, “Coactive Design: Designing Support for Interdependence in Joint Activity,” *Journal of Human-Robot Interaction*, vol. 3, no. 1, pp. 43–69, 2014, doi: 10.5898/JHRI.3.1.Johnson. Cited author-hosted copy of the journal article; pinpoints use its printed pagination.
35. S. Amershi et al., “Guidelines for Human-AI Interaction,” in *Proceedings of CHI 2019*, Glasgow, May 4–9, 2019, 13 pages, doi: 10.1145/3290605.3300233. Cited camera-ready copy, including its appendix.
36. I. Santos-Grueiro, “Temporary Authority, Permanent Effects: Commit-Time Authorization for LLM Agents,” *arXiv:2607.10487v1*, July 11, 2026, version 1.
37. J. Zhu, Z. Liu, S. Fan, J. Chen and Q. He, “From Approval to Execution: Reconstruction-Aware Repair Analysis for LLM-Agent Software,” *arXiv:2609.26529v1*, Sep. 22, 2026, version 1.
38. J. Zhang, H. Xia, S. Wu, J. Yue, X. Zhang, Z. Cheng and B. Tu, “Agent Approval Laundering: Transitive Effects Beyond the Approved Invocation,” *arXiv:2609.28586v1*, Sep. 23, 2026, version 1.
39. G. Baxter and I. Sommerville, “Socio-technical systems: From design methods to systems engineering,” *Interacting with Computers*, vol. 23, no. 1, pp. 4–17, 2011, doi: 10.1016/j.intcom.2010.07.003. Cited to-appear author manuscript; manuscript pagination differs from the published article.
40. A. Artikis, M. Sergot and J. Pitt, “Specifying Norm-Governed Computational Societies,” Technical Report DTR06-5, Imperial College London, 2006, report PDF. Pinpoints use the report's pagination.
41. J. Gelati, G. Governatori, A. Rotolo and G. Sartor, “Declarative Power, Representation, and Mandate. A Formal Analysis,” in T. J. M. Bench-Capon, A. Daskalopulu and R. G. F. Winkels (eds.), *Legal Knowledge and Information Systems. Jurix 2002: The Fifteenth Annual Conference*, Amsterdam: IOS Press, 2002, pp. 41–52, conference PDF.
42. S. Colombo Tosatto, G. Governatori and P. Kelsen, “Towards an Abstract Framework for Compliance,” in *2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops*, pp. 79–88, 2013, author-hosted article.
43. S. Yao, N. Shinn, P. Razavi and K. Narasimhan, “ τ -bench: A Benchmark for Tool-Agent-User Interaction in Real-World Domains,” *arXiv:2406.12045v1*, June 17, 2024, version 1.
44. Z. Yang, A. Bibi, G. Li, B. Ghanem and C. Hinojosa, “APEX-Harm: Evaluating Agent Safety in Shared Professional Workspaces,” *arXiv:2610.13623v1*, Oct. 9, 2026, version 1.
45. Y. Ling, S. Yu, Z. Chen and C. Fang, “When Faster Decisions Do Not Make Better Agents: An Empirical Study of Jev in Agent Workflows,” *arXiv:2610.13092v1*, Oct. 9, 2026, version 1.
46. Author's development record, “Closed review-view diagnostic,” Sep. 23, 2026. Unpublished synthetic-matter engineering note, on file with the author.
47. NASA Ames Research Center, *NASA Task Load Index (NASA-TLX), Version 1.0: Paper and Pencil Package*, n.d., official manual rehost, accessed Oct. 3, 2026. Pinpoints use the manual's printed pagination.

48. Author's development record, "Original-case batch closed: receipts pass, semantic classification fails," Sep. 24, 2026. Unpublished synthetic-matter engineering note, on file with the author.
49. Author's development record, "Closed supervised qualification: initial success and follow-up failure," Sep. 23, 2026. Unpublished synthetic-matter engineering note, on file with the author.
50. Author's development record, "Closed Qwen/Jev review-decision comparison," authored Sep. 22, 2026, filed in the workspace notebook Sep. 23, 2026. Unpublished engineering note, on file with the author.