



Multi-Agent AI as a Nested Principal-Agent Problem in Private Wealth Management

Mandate Representation and Evidence Control in Switzerland, Germany and Austria

Walter Kurz¹, Reinhard Magg¹, Florian Kollberg², Wojtek Stricker¹, Stefan Marx², Frank Reinhardt², Velimir Dedić³

¹Swissi Institute for AI, office@swissi-ai.institute; ²Hochschule für Wirtschaft und Umwelt Nürtingen-Geislingen

³Faculty of Information Technology and Engineering (FITI), Belgrade

ISSN 3043–1921 | Volume 2026 | <https://journal.swissi-ai.institute> | journal@swissi-ai.institute | CC BY 4.0
Article SAIJ-guex5c23zfm5 | DOI: 10.5281/zenodo.23095194 | <https://journal.swissi-ai.institute/doi/guex5c23zfm5>
Submitted 06/2026; Revised 09/2026; Published 10/2026

Abstract

In private wealth management, a manager delegating to artificial intelligence (AI) acts as the client’s agent and the system’s principal. We introduce a model-independent formulation that combines nested principal-agent delegation with constrained joint maximisation as the task assigned to the AI system. The objective represents client and manager outcomes separately over portfolio-workflow pairs. Legal duties, mandate requirements and evidence sufficiency determine admissibility, with Switzerland, Germany and Austria supplying the legal context. Weights and reference-service floors make the trade-off explicit; concession accounting separates their effects on the client. Analytical constructions and a simulation using public-market observations illustrate the approach. Across eight decision states from four constructed mandates, omitted client liabilities caused two liquidity violations, omitted manager terms caused two capacity violations, and mistranslated weights changed four otherwise admissible choices under faithful optimisation. At the declared weights, six states selected a higher service tier than the client-best alternative, with client concessions of EUR 1,178 to EUR 2,264 and manager gains of EUR 3,062 to EUR 10,381. Three instruction forms each reached all 32 specified decisions under shared numerical, evidence and simulated approval controls; professional instructions matched explicit nested delegation on accuracy and clarification count. Subsequent 2022 exchange-rate and yield paths, combined with constructed growth scenarios, produced lower client outcomes than the reference service although the selected services met the decision-time forecast benchmarks. These examples suggest that the approach could help make mandate choices and their consequences easier to examine. Professional and field studies could assess whether this improves oversight and client outcomes.

Keywords: multi-agent AI, private wealth management, nested principal-agent, mandate representation, joint objective, specification risk, suitability, premise provenance, human oversight

1 Introduction

A wealth manager delegating work to a system of AI agents asks it to make decisions from information filtered through two parties. The client may leave circumstances or preferences undisclosed; the manager may incompletely convey that account, its own objectives or the service conditions. Either omission can reflect strategic disclosure, uncertainty or imperfect articulation. The AI can have the least direct access to the private information on which its task depends, while being expected to serve both parties. The specification risk is that gaps are silently filled with learned defaults or unsupported assumptions about their circumstances and objectives, then carried into downstream decisions. Studies of multi-agent failures document action under incorrect assumptions, failures to seek clarification and incomplete verification (Cemri et al., 2025).

Agency theory addresses delegated authority under information asymmetry and potentially divergent interests (Jensen & Meckling, 1976). A classical formulation advances the principal’s interests while securing the agent’s participation and

desired behaviour (Holmstrom, 1979). Financial applications examine moral hazard in delegated portfolio management and incentives behind unsuitable sales (Inderst & Ottaviani, 2009; Stoughton, 1993). Hierarchical agency supplies a basis for examining successive delegations (Tirole, 1986). We propose to apply this framework to the client–manager–AI relationship: the client is the manager’s principal, and the manager is both the client’s agent and the AI system’s principal. This structure locates responsibility for representing interests and transmitting information at each interface.

We propose to specify the AI’s task through two explicit outcome functions: one for the client and one for the wealth manager. Client outcomes include return, liquidity and financial security; manager outcomes include service quality, resource use and commercial value. Their interests can coincide or conflict. The manager’s interests shape the service whether or not the specification names them; naming them turns the trade-off into something a reviewer can read. Constrained joint maximisation over portfolio–workflow pairs makes the trade-off explicit, with legal duties, mandate requirements and evidence sufficiency determining admissibility. Weights and reference-service floors require justification and authorization. The intended engineering use is to translate this specification into agent instructions and workflow rules that preserve both evaluations, bind decisions to supported premises and identify when unresolved information or objective choices require human clarification. Each agent’s assigned role must remain consistent with that shared specification.

The DACH setting makes the treatment of incomplete information consequential. In Switzerland, Articles 11–12 of the Financial Services Act (FinSA) require the provider to seek service-relevant client information. If that information is insufficient to assess appropriateness or suitability, Article 14(1) requires prior notice to the client that the provider cannot perform the assessment (Swiss Confederation, 2018, arts. 11–14). For investment advice and portfolio management under EU suitability rules, Article 25(2) of the Markets in Financial Instruments Directive (MiFID II) requires necessary client information, and Article 54(8) of Delegated Regulation 2017/565 bars recommendations when that information is missing (European Parliament and Council of the European Union, 2014, art. 25(2); European Commission, 2017, art. 54(8)).

The manager must specify both which information the service requires and what an unresolved gap permits the system to do. A uniform hold policy adds an institutional restriction to the Swiss notification rule, while the EU suitability rule expressly restricts recommendations. Applicable mandate and other legal duties continue to govern the service. Jurisdiction and service type enter the decision specification because the same information gap can require different responses.

The research question is how nested client–manager–AI delegation can be translated into an inspectable specification for joint optimisation and clarification under incomplete information. The intended users are the manager responsible for the mandate and the engineer configuring the agents and workflow. The assessment distinguishes the adequacy of a transmitted specification, compliance with it and outcomes under subsequent market conditions. The simulation and executable constructions illustrate how the theory can be instantiated; the system architecture remains the builder’s choice.

2 Related Work

Agency theory examines delegated authority under potentially divergent interests; portfolio-management and misselling models address effort, incentives and monitoring within the client–manager relationship (Holmstrom, 1979; Inderst & Ottaviani, 2009; Jensen & Meckling, 1976; Stoughton, 1993). Hierarchical agency provides the reference for a further delegation (Tirole, 1986). Common agency studies an agent who serves several principals at once (Bernheim & Whinston, 1986); in our setting the two principals stand in sequence, and the second is the first one’s agent. Work on automated advice examines what algorithmic portfolio services change for retail investors (D’Acunto et al., 2019).

PeterAI’s publisher abstract describes personalised investment consultancy through specialised agents and an ambiguity-resolution pipeline; our comparison uses those reported architectural features (Gadioli et al., 2026). Generative AI has also been modelled as an informational signal within principal-agent relationships (Zhang & Zhang, 2025). Selective clarification and studies of multi-agent failures examine information requests and inherited errors (Cemri et al., 2025; Kuhn et al., 2023; Xie et al., 2026).

Zhuang and Hadfield-Menell (2020) formalise incomplete objective specification as an AI principal–agent problem, analysing proxy objectives and interactive revision under stated assumptions. Ante (2026) locates specification cost in persistent delegation losses with externally supplied objectives; the publisher abstract identifies incomplete specification as a financial-agent governance problem. A recent preprint by Sreenivas and Larson (2026) models contractual incentives when a delegated agent chooses its language model and inference effort.

A wealth-management benchmark evaluates paired assistant tasks with deterministic checkpoints (Milsom, 2025). The vendor-authored AI Advice Evidence Benchmark and associated Decision Assurance Envelope address investor context, mandate, suitability, policy, source evidence and human approval (NeuFin, n.d.; Srivastava, 2026); FiduciaryBench supplies source-linked US conduct and planning tasks (WealthSchema, 2026). The Multi-Jurisdictional Legal Identity Assurance for Capability Gating model connects authority and supporting evidence to permission for an act, assigning mandate mechanics and aggregation to companion work (Kurz, 2026). Applying these approaches to wealth advice requires distinguishing client-authorized preferences from factual claims whose evidence requirements depend on the service.

Financial multicriteria decision support supplies an established methodological base (Steuer & Na, 2003). Robust multi-objective optimization examines scalarization and dependent uncertainty (Bokrantz & Fredriksson, 2017), while shield synthesis uses runtime monitoring to enforce specified properties (Bloem et al., 2015). These strands inform the distinction between selecting under declared evaluations and enforcing permission to act.

Austria's Securities Supervision Act 2018 (WAG 2018) and European Securities and Markets Authority (ESMA) guidance further specify assessment requirements (European Securities and Markets Authority, 2022; Republic of Austria, 2018a). The EU General Data Protection Regulation (GDPR) and Swiss data-protection law govern processing (European Parliament and Council of the European Union, 2016; Federal Data Protection and Information Commissioner, 2025). Institutional AI governance is addressed by the Swiss Financial Market Supervisory Authority (FINMA) and Germany's Federal Financial Supervisory Authority (BaFin) (Bundesanstalt für Finanzdienstleistungsaufsicht, 2021; Swiss Financial Market Supervisory Authority FINMA, 2024). These sources distinguish information needed for the service from permitted collection and processing.

Existing rules retain responsibility for automated suitability assessment and require scrutiny of intended goals (European Commission, 2017, art. 54(1); Swiss Financial Market Supervisory Authority FINMA, 2024, secs. 2.4 and 2.7). ESMA guidance already specifies algorithm documentation, testing, controlled changes and internal sign-off, together with review of costlier equivalent products (European Securities and Markets Authority, 2022, paras. 90–95). Recent preprints connect financial-agent authority to mandate lineage, policy versions and approvals, and describe enforcement of delegated mandate chains (Gong et al., 2026; Racioppi, 2026). We formulate the delegated AI's task as constrained joint maximisation within the nested client–manager–AI relationship. The manager's dual role determines whose outcomes enter the objective, how legal and mandate duties restrict the choice set, and who is responsible for the specification transmitted to the AI.

3 Contribution

The contribution is a nested principal–agent formulation in which the delegated AI maximises a joint client–manager objective subject to binding constraints. The manager acts as the client's agent when agreeing the mandate and as the AI's principal when specifying its task. We connect these roles through separate client and manager outcome functions over portfolio–workflow pairs, with legal, mandate and evidence requirements determining admissibility before optimisation (section 5). The weight places each decision between a client-first and a manager-first programme (Proposition 2). Concession accounting separates benchmark restrictions from weighting; a proposed review screen requires policy review and assesses the total concession. Disclosure, translation and execution provide a decomposition of departures under stated counterfactual and tie-selection conventions. A simulation illustrates required-input recovery, numerical selection and version-bound approval. Exact constructions assess the review screen, while partition-based discretionary clarification is implemented for the restricted client selector.

Three research questions guide the assessment. First, how can client–manager–AI delegation be represented by a joint decision objective with binding client-protection, mandate and evidence constraints? Second, how do information omissions, representation choices and uncertainty assumptions affect admissibility, selection and the separate outcomes of both parties? Third, how can the specification be instantiated in agent instructions and workflow controls?

The formal specification and its finite client-selection instance address the first question. Analytical constructions and the portfolio–workflow simulation address the second, varying what each delegation transmits while retaining the original mandate for evaluation. The instruction comparison addresses the third, complemented by deterministic, single-agent and multi-agent implementations. Together these constructions demonstrate how declared objectives, service conditions and evidence requirements govern a decision.

4 Methods

We used design science to develop the decision specification (Hevner et al., 2004; Peffers et al., 2007), combining analytical constructions, economic simulation and worked implementations for artificial evaluation (Venable et al., 2016). The formal results depend on the stated objectives, constraints and information structure; any executor preserving these rules can instantiate the specification. Exact-rational constructions isolated objective representation, benchmark eligibility, clarification and evidence transitions. Case policies supplied the assessment constraints informed by the legal mapping.

The economic simulation represented complete client circumstances, manager service terms and a declared mandate, then controlled what disclosure and specification transmitted. Four constructed families covered business-sale liquidity, retirement drawdown, currency liabilities and concentrated holdings. Three portfolios and three service workflows competed with a reference service. Separate interventions omitted client liabilities, omitted manager terms or changed the transmitted weight; faithful optimisation of each altered specification was evaluated against the original mandate. Simulated approval checked current premise support, a matching calculation, numerical constraints, reference floors and evidence version. The comparison held substantive mandate authorization fixed; the concession-review screen was assessed separately through exact constructions. Thus numerical execution and domain authorization have distinct roles in the demonstration.

Monthly cash flows included withdrawals, liabilities, currency holdings, zero-coupon bond valuation, fees, forced-sale costs and supervision. Client value combined expected wealth, a downside penalty and declared service value; manager value combined net fee income and service quality. Their scales and reference floors were fixed before selection. European Central Bank exchange rates and government yields supplied 144 monthly observations for 2014–2025 (European Central Bank, n.d.-a, n.d.-b). Growth returns, preferences, service conditions and three scenario probabilities were authored. Decision-time evaluations used the initial market observation; subsequent observed exchange rates and yields entered the outcome calculation with the same constructed growth scenarios.

Professional prose, explicit joint-objective instructions and explicit nested-delegation instructions received the same declared objective, evidence, questions, calculators, resource ceilings and simulated approval controls. Calculators returned all candidate outcomes for agent selection. December 2016 mandates supported development; the same four families at December 2021, with rescaled liabilities, spending and hourly costs, supplied four evaluation cases with eight information conditions each. These covered complete evidence, client or manager gaps, both gaps, translation conflict, refusal, correction after approval and equal-outcome controls. One run per treatment and condition yielded 96 assignments. All three instruction forms used the same model and runtime settings. Private states, answers and subsequent market paths stayed with the evaluator.

The retained studies tested supplied client-selection policies using deterministic, single-agent and multi-agent workflows. The structured pilot contained 108 assignments; each narrative comparison contained 126. Swiss Banking Ombudsman cases supplied authorization, retirement-depletion and investment-cash-flow mechanisms (Swiss Banking Ombudsman, 2019, 2020, 2023); records and exposures were authored. The revised narrative interface combined typed extraction, clearer instructions and reusable arithmetic receipts, with re-evaluation on the same cases. These implementations illustrate the same decision concepts through alternative workflow arrangements.

Scoring distinguished fidelity to the original mandate, raw proposal defects, issuance, clarification burden, the declared joint optimum and separate economic outcomes. Current evidence and approval governed issuance. Independent cash-flow identities checked financial arithmetic; saved actions were replayed to verify decisions and evidence transitions. Source and model-response records retained configurations, prompts and usage. Conditions sharing a mandate were treated as dependent observations. Model expenditure and the constructed service costs were reported separately. The accompanying *Supplementary Analytical and Resource Results* contains the detailed counterexamples, reference traces and resource tables.

The legal mapping concerns investment advice and portfolio management for private or retail clients, with transaction-specific advice and execution services supplying applicability contrasts.

Under Swiss FinSA, transaction-specific advice assesses knowledge and experience for the proposed instrument; portfolio-based advice and management also assess financial situation and objectives. Insufficient information triggers Article 14's client-information duty (Swiss Confederation, 2018, arts. 11–14). German and Austrian advice and portfolio management use the suitability inputs in WpHG section 64(3), WAG 2018 section 56 and Article 54 of Delegated Regulation 2017/565, including loss-bearing capacity and risk tolerance (European Commission, 2017; Federal Republic of Germany, 1994; Republic of Austria, 2018a). Article 54(8) conditions recommendations on obtaining the necessary

information (European Commission, 2017, art. 54(8)). The artifact's uniform hold policy is a conservative design choice where the mapped Swiss provision specifies an information duty. The constructed liability amount is classified as required by the case policy.

Swiss execution or transmission services engage Article 13's notification rule; institutional-client treatment and professional-client waivers require the separate scope of Article 20 (Swiss Confederation, 2018, arts. 13 and 20). EU appropriateness and qualifying execution-only services follow the conditions in MiFID II Article 25(3)–(4), with professional-client assumptions addressed separately in the delegated rules (European Parliament and Council of the European Union, 2014, art. 25; European Commission, 2017, arts. 54–56). GDPR Articles 5, 6 and 9 govern processing principles and relevant grounds; Swiss processing is assessed under the Federal Act on Data Protection, including Articles 6 and 30–31 (European Parliament and Council of the European Union, 2016; Swiss Confederation, 2020). Client willingness, evidential adequacy and lawful processing receive separate policy decisions.

The legal analysis uses a cutoff of 30 September 2026: the FinSA consolidation of 1 March 2024, Regulation 2017/565 consolidation of 2 August 2022, current German section 64 text and Austrian section 56 effective 9 June 2022. MiFID II Article 25 follows the consolidation of 6 June 2026. The German section 64 text is the version in force at the cutoff, cited without a provision-level effective date (European Commission, 2017; European Parliament and Council of the European Union, 2014; Federal Republic of Germany, 1994; Republic of Austria, 2018a; Swiss Confederation, 2018). FINMA Circular 2025/2, effective 1 January 2025, addresses service classification and knowledge-and-experience elicitation for providers within its scope (Swiss Financial Market Supervisory Authority, 2024, paras. 2, 4 and 13–14).

The constructed legal contrasts concern domestic retail services. Cross-border application depends on provider authorization, client category and the origin of the service request. Swiss territorial rules and German or Austrian market-access requirements can apply to the same relationship (Swiss Federal Council, 2019, art. 2; Federal Republic of Germany, 2021, sec. 15; Federal Republic of Germany, 1961, sec. 32; Republic of Austria, 2018b, secs. 21, 23–24). Exclusive client initiative has a fact-dependent scope, assessed from the service and solicitation history (European Securities and Markets Authority, 2021). Market-access conditions, suitability requirements and the territorial scope of data-processing rules are assessed separately (European Parliament and Council of the European Union, 2016, art. 3; Swiss Confederation, 2020, art. 3).

5 From Classical Agency to Nested Delegation

Agency theory starts from a principal who cannot observe what the agent does or knows (Jensen & Meckling, 1976). In the hidden-action formulation, the principal chooses a payment schedule s and the effort e it wants the agent to supply (Holmstrom, 1979, pp. 75–76, eqs. (1)–(3)):

$$\begin{aligned} & \max_{s \in \mathcal{S}, e \in \mathcal{E}} \quad \mathbb{E}[u_P(Y - s(Y)) \mid e] \\ & \text{subject to} \quad \mathbb{E}[u_A(s(Y)) \mid e] - c(e) \geq \bar{u}_A, \\ & \quad e \in \arg \max_{\tilde{e} \in \mathcal{E}} \{ \mathbb{E}[u_A(s(Y)) \mid \tilde{e}] - c(\tilde{e}) \}. \end{aligned} \quad (1)$$

Here Y is the observable outcome, u_P and u_A are the utilities of principal and agent, c is the agent's effort cost, $\bar{u}_A$ its reservation utility, and $\mathcal{S}, \mathcal{E}$ are the admissible payment schedules and efforts. The first constraint secures the agent's participation. The second states that the agent picks the effort that suits it best under the offered payment.

Three features carry this model. The agent has interests of its own, expressed in u_A and c . The agent holds the informational advantage, since it alone observes e . And the principal's instrument is the contract: it steers the agent through what it pays.

Wealth management with AI contains two delegations. The client delegates advisory work to the wealth manager, and the manager commissions and supervises the AI system:

$$C = P_1 \longrightarrow W = A_1 = P_2 \longrightarrow A = A_2. \quad (2)$$

C, W and A denote the client, the wealth manager and the AI system; P_j and A_j mark principal and agent in delegation j . Hierarchical agency has studied such chains among human parties (Tirole, 1986). The first delegation is the familiar one: the manager knows more about products and about its own conduct than the client does, and its remuneration can pull against the client's interest (Inderst & Ottaviani, 2009; Stoughton, 1993). The chain also differs from common agency, in which one agent serves several principals side by side (Bernheim & Whinston, 1986):

$$\underbrace{P_1 \longrightarrow A \longleftarrow P_2}_{\text{common agency}} \qquad \underbrace{P_1 \longrightarrow (A_1 = P_2) \longrightarrow A_2}_{\text{nested delegation}} \quad (3)$$

Under common agency the principals stand on equal footing and compete for the agent. In the nested chain the second principal is bound to the first, and this ranking is what later places admissibility before the objective.

Figure 1 shows the manager translating the client's mandate into AI instructions and receiving proposed advice with supporting premises. The dashed link denotes permitted client-context elicitation.

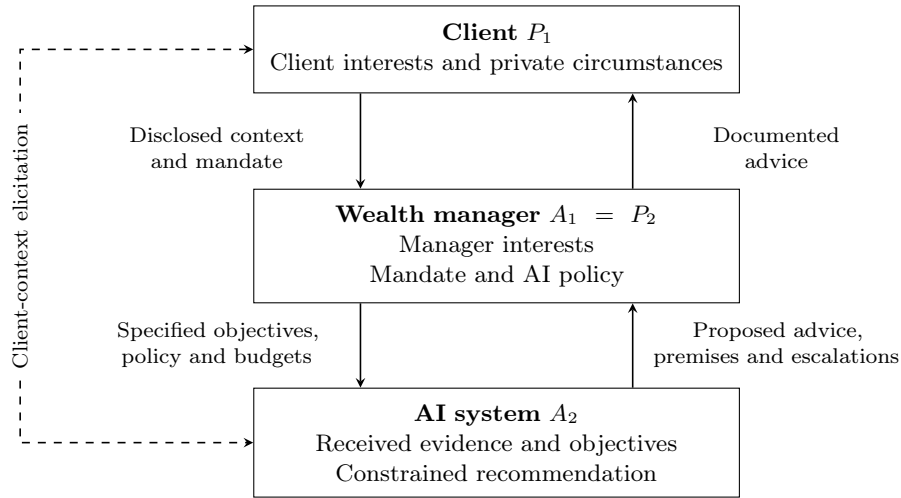


Figure 1: *Nested delegation and client-context elicitation.*

Note. The wealth manager acts as the client's agent (A_1) and the AI system's principal (P_2). Solid arrows show the two delegation relationships; the dashed connection shows permitted client-context elicitation. Source: the authors' conceptual model.

The instrument changes. We model the AI as an executor of the supplied specification, with no separate utility, participation decision or strategic choice of effort. Computational and supervisory costs enter the manager's service evaluation. Under this assumption, the participation and incentive constraints of Equation (1) apply to the human service relationship; the second delegation is governed by the specification. The manager steers the system through the specification P it hands down. P contains an objective F_P , a set $\mathcal{A}_P$ of alternatives the system may recommend, and rules for the information it may rely on. Where the classical agent maximises its own utility under the contract, the system is asked to maximise the objective it was given, on the input I_A it received:

$$e \in \arg \max_{\tilde{e} \in \mathcal{E}} \{E[u_A(s(Y)) \mid \tilde{e}] - c(\tilde{e})\} \quad \text{becomes} \quad a^X \in \arg \max_{a \in \mathcal{A}_P(I_A)} F_P(a \mid I_A). \quad (4)$$

The design variable moves from the payment schedule s to the specification P (Zhuang & Hadfield-Menell, 2020). A system that departs from Equation (4) makes an execution error, which testing can find. The harder questions concern the two objects the system takes as given: what P says, and what I_A contains.

Information passes through two interfaces. In Equation (1) the agent knows more than the principal. In the nested chain, the AI receives the client's private circumstances and the manager's service terms through disclosure and specification. Its access to that information depends on what those interfaces transmit and what supplementary evidence it can obtain. Let θ_C describe the client's circumstances and preferences and θ_W the manager's service objectives, costs and incentives. With a recorded mandate M and jurisdiction and service conditions J , information reaches the AI in three steps:

$$\begin{aligned} s_C &= D_C(\theta_C), \\ P &= T_W(s_C, \theta_W, M, J, Z_W), \\ I_A &= H_A(s_C, P, Z_A). \end{aligned} \quad (5)$$

The client discloses s_C , the manager translates it together with its own terms into P , and the system receives the decision input I_A ; Z_W and Z_A are supplementary evidence available at each stage. Each step can drop or distort content, deliberately (Crawford & Sobel, 1982) or in good faith, and supplementary evidence can also add to what the system knows. Information omitted at an earlier step remains unavailable downstream unless clarification or supplementary evidence supplies it. The classical question is how the principal learns what the agent knows. Here the question is how what the principals know reaches the agent.

Disclosure can itself be a choice. A client whose true utility U_C differs from the declared evaluation, for reasons of privacy, tax or attachment to a holding, selects what to disclose with the resulting recommendation in view:

$$D_C(\theta_C) \in \arg \max_s U_C(a^X(s); \theta_C). \quad (6)$$

Crawford and Sobel (1982) show that when the interests of sender and receiver differ, equilibrium messages are coarse: ranges of states share one message. In the notation of Equation (5), several client states θ_C then map to the same s_C and, without further evidence, to the same I_A . The following proposition states the limit this places on any system.

Proposition 1 (indistinguishable client states). Fix the mandate, policy primitives and decision rule. Let two client states induce the same complete AI decision input I_A , but have disjoint sets $X_C^*(\theta)$ and $X_C^*(\theta')$ of fully informed client-optimal portfolios under those primitives. A rule selecting a portfolio solely from I_A can guarantee a fully informed client optimum in at most one of these states.

A deterministic rule receives the same input and selects the same portfolio in both states. That portfolio cannot belong to both disjoint optimal sets. A randomised rule has the same conditional output distribution in both states; probability-one optimality in each would require its support to lie in their empty intersection.

The result concerns selection at the given input. A permitted question can distinguish the states; a hold preserves the unresolved decision. The proposition identifies an informational limit independently of the system's computational ability.

The conflict receives an explicit evaluation. The manager is agent in the first delegation and principal in the second. As agent it owes the client loyalty; as principal it writes the system's objective. The proposed specification records the weight and scale assigned to its commercial interests together with the alternatives and reference service. Existing governance already requires documentation and review of automated decision rules (European Securities and Markets Authority, 2022, para. 90). Explicit dual evaluations add a measurable trade-off and a comparator against which a reviewer can examine each client concession.

We take the service arrangement and its remuneration as given. The mandate is also incomplete: circumstances change, so elicitation and specification have to be renewed (Grossman & Hart, 1986; Hart & Moore, 1990). How client and manager respond strategically to a specification that both can read is a further layer of analysis, taken up in section 9.

The system decides on a pair $a = (x, \omega)$: a portfolio recommendation x and the workflow ω that delivers it, drawn from a finite set $\mathcal{A}$. Let E contain the accepted, currently valid case evidence from I_A . The declared evaluations $\hat{V}_C(a | E)$ and $\hat{V}_W(a | E)$ represent client and manager outcomes. Client criteria may include return, liquidity and security; manager criteria may include service quality, delivery cost and supervisory effort. Their definitions, scales and treatment of uncertainty are fixed before any comparison and remain subject to mandate review. An expectation needs an explicit belief model, a robust evaluation needs declared plausible states, and probabilities generated by a model need separate calibration before they serve as beliefs about a particular client.

Let $\mathcal{H}$ index the legal, risk, processing, mandate and independently justified service-viability constraints, including applicable client-interest duties, with residuals $g_h \leq 0$ when satisfied. Service viability records necessary delivery resources; preserving a reference profit is a separate benchmark choice. Let $G_{\text{req}}(E, J, M)$ indicate that the information necessary for the service is established, and let $\Theta(E)$ be the non-empty set of states compatible with the evidence. Equation (7) admits the alternatives that pass the required-information gate and satisfy every constraint in every such state:

$$\mathcal{A}_{\text{adm}}(E, J, M) = \left\{ a \in \mathcal{A} : \begin{array}{l} G_{\text{req}}(E, J, M) = 1, \\ g_h(a; \theta, J, M) \leq 0 \quad \forall h \in \mathcal{H}, \theta \in \Theta(E) \end{array} \right\}. \quad (7)$$

For EU services, client-interest duties and restrictions on conflicting remuneration bound the manager's commercial objective (European Securities and Markets Authority, n.d., paras. 1 and 10), alongside the rules on conflicts of interest (European Parliament and Council of the European Union, 2014, art. 23). For Swiss services, the corresponding bounds are the organisational duties on conflicts of interest and the rules on compensation from third parties (Swiss Confederation, 2018, arts. 25–26), together with the agent's duties of loyalty and of accounting under mandate law (Swiss Confederation, 1911, arts. 398 and 400). These rules enter as members of the constraint set, $\{h_{\text{suit}}, h_{\text{conf}}, h_{\text{comp}}\} \subseteq \mathcal{H}$ for suitability, conflicts of interest and third-party compensation; the weighted comparison begins only where $g_h \leq 0$ holds for each of them. Missing required information or inconsistent evidence holds the recommendation.

A declared reference service a_0 gives each party a floor $b_r(E) = \hat{V}_r(a_0 | E)$, $r \in \{C, W\}$, under the same evidence and evaluation rules. The alternatives that are admissible and keep both parties at their reference level form the feasible set

$$\mathcal{A}_{\text{feas}}(E) = \left\{ a \in \mathcal{A}_{\text{adm}}(E, J, M) : \hat{V}_C(a | E) \geq b_C(E), \hat{V}_W(a | E) \geq b_W(E) \right\}. \quad (8)$$

Within $\mathcal{A}_{\text{feas}}(E)$ the first delegation can be written in the form of Equation (1). A client who could choose the service itself would solve

$$\max_{a \in \mathcal{A}_{\text{feas}}(E)} \hat{V}_C(a | E). \quad (C)$$

The client is the principal, and the manager floor plays the part of the participation constraint: the manager must do at least as well as with the reference service. The choice, however, lies with the manager. A manager who followed only its own interest would solve

$$\max_{a \in \mathcal{A}_{\text{feas}}(E)} \widehat{V}_W(a | E), \quad (\text{W})$$

where the client floor now takes the part of the participation constraint: the client keeps the mandate as long as it does at least as well as with the reference service. The distance between the solutions of (C) and (W) is the agency problem of the first delegation, written for the service decision. A client who sees only the recommendation cannot tell which of the two programmes produced it.

What should a manager who takes both of its roles seriously write into P ? Programme (W) treats the second delegation as if the first did not exist. Programme (C) looks like the loyal choice, and it has a weakness of its own. The workflow ω uses the manager's staff time and supervisory capacity and determines the fee the client pays, so the manager's interests are part of the decision whether or not the objective names them. Under a client-only objective they enter elsewhere: in which alternatives are offered, which service tiers exist and how much review a case receives. Those choices are made outside the specification, where a reviewer sees no trade-off because none is written down.

We therefore propose that the specification states both evaluations and the weight between them. For a fixed $0 < \lambda < 1$, Equation (9) selects

$$\begin{aligned} a^* \in \arg \max_{a \in \mathcal{A}_{\text{adm}}(E, J, M)} & \lambda \widehat{V}_C(a | E) + (1 - \lambda) \widehat{V}_W(a | E) \\ \text{subject to} & \widehat{V}_r(a | E) \geq b_r(E), \quad r \in \{C, W\}, \end{aligned} \quad (9)$$

which is the maximisation of the weighted sum over $\mathcal{A}_{\text{feas}}(E)$. The two programmes above are its limits.

Proposition 2 (limiting programmes). Let $\mathcal{A}_{\text{feas}}(E)$ be non-empty. Every solution of Equation (9) is Pareto-efficient in $\mathcal{A}_{\text{feas}}(E)$ with respect to the two declared evaluations. There are thresholds $\lambda_W > 0$ and $\lambda_C < 1$ such that every solution solves (W) when $\lambda < \lambda_W$ and solves (C) when $\lambda > \lambda_C$.

An alternative in $\mathcal{A}_{\text{feas}}(E)$ that weakly improved both evaluations of a solution and strictly improved one would have a strictly larger weighted sum, which contradicts optimality. For the thresholds, let $\delta_C > 0$ be the smallest gap between the highest client value in $\mathcal{A}_{\text{feas}}(E)$ and any lower client value, and let R_W be the range of manager values in $\mathcal{A}_{\text{feas}}(E)$. An alternative below the highest client value loses at least $\lambda \delta_C$ and gains at most $(1 - \lambda) R_W$ against a client-best alternative, so it cannot be selected when $\lambda > \lambda_C = R_W / (R_W + \delta_C)$. The argument for λ_W exchanges the two parties. If all alternatives share the highest client value, every solution solves (C) for any λ .

The efficiency statement runs in one direction. On a finite set an efficient compromise can be selected at no weight: with client and manager values $(1, 1/4)$, $(1/4, 1)$ and $(3/5, 3/5)$, the third pair scores $3/5$ at every λ while one of the first two scores at least $5/8$. A mandate that wants such a compromise states it as a constraint.

The weight therefore has a plain reading: it places the decision between the manager's programme and the client's programme, and client priority is the case $\lambda > \lambda_C$. The proposal leaves the value of λ to the mandate. It requires that the value be declared, justified against the mandate and authorized before the system acts on it.

The other elements have a reading in the nested structure as well. Admissibility comes first because the first delegation ranks above the second: legal duties and the mandate bound what the manager may want, so manager value competes only among alternatives that the client relationship already permits. The floors tie the decision to the reference service. The reservation utility $\bar{u}_A$ in Equation (1) is the value of an outside option; the manager floor b_W is stricter, since it preserves the full managerial value of the reference service, and this choice needs its own justification in the mandate. An empty $\mathcal{A}_{\text{feas}}(E)$ calls for revised alternatives or a held decision, and current human approval remains a condition of issuance.

Review requires the cost of both benchmark restrictions and weighting. Let a_C solve (C). Relative to that floor-feasible client optimum, the weight concession and manager gain are

$$\Delta_C = \widehat{V}_C(a_C | E) - \widehat{V}_C(a^* | E) \geq 0, \quad \Delta_W = \widehat{V}_W(a^* | E) - \widehat{V}_W(a_C | E). \quad (10)$$

Optimality requires the manager gain to compensate for the weighted client concession:

$$\lambda \widehat{V}_C(a^*) + (1 - \lambda) \widehat{V}_W(a^*) \geq \lambda \widehat{V}_C(a_C) + (1 - \lambda) \widehat{V}_W(a_C) \implies \Delta_W \geq \frac{\lambda}{1 - \lambda} \Delta_C. \quad (11)$$

The rule accepts a client concession only when the manager gains at least $\lambda/(1 - \lambda)$ times as much on the declared scales: nineteen times at $\lambda = 0.95$, an equal amount at $\lambda = 0.5$. The ratio is an exchange rate between the two

evaluations, and it depends on how each is scaled. If client value is measured in units of k_C euros and manager value in units of k_W euros, one manager euro counts as much as ε client euros, with

$$\varepsilon(\lambda) = \frac{1 - \lambda}{\lambda} \cdot \frac{k_C}{k_W}. \quad (12)$$

A weight close to one can therefore still favour the manager when the manager scale is the finer one, and for $\varepsilon > 1$ a pure transfer from client to manager raises the objective. The pair (Δ_C, Δ_W) reports the effect of weighting within the floor-feasible set. To expose the effect of the floors themselves, let $a_H \in \arg \max_{a \in \mathcal{A}_{\text{adm}}(E, J, M)} \widehat{V}_C(a | E)$ be a client-best hard-admissible and viable service, before reference floors. For non-empty feasible sets, total concession comprises benchmark and weighting effects:

$$\begin{aligned} \Delta_C^{\text{bench}} &= \widehat{V}_C(a_H | E) - \widehat{V}_C(a_C | E) \geq 0, \\ \Delta_C^{\text{tot}} &= \widehat{V}_C(a_H | E) - \widehat{V}_C(a^* | E) = \Delta_C^{\text{bench}} + \Delta_C. \end{aligned} \quad (13)$$

A zero weight concession can therefore coexist with a positive benchmark concession. The proposed decision record reports both components, the total and the manager gain, with the comparator identities and policy version.

Delivery requirements enter $\mathcal{H}$. New evidence of a binding requirement triggers a revised admissible set and recomputation of a_H , a_C and a^* . This preserves comparison with the best remaining deliverable service. Let $Q(P, E) = 1$ record completed independent policy review of the objectives, scales, weights, reference floors, service viability and applicable duties. Let $\beta(a^*, a_H)$ be a supported, current incremental client benefit omitted from the declared evaluation, measured on its client scale; benefits already represented receive zero additional credit. We propose the following institutional screen for submitting a decision to substantive authorization:

$$\text{Ready}(a^*, E, P) = Q(P, E) \wedge [\Delta_C^{\text{tot}} = 0 \vee \beta(a^*, a_H) \geq \Delta_C^{\text{tot}}]. \quad (14)$$

For a positive total concession, absent or inadequate benefit evidence retains the decision for policy revision. An empty admissible or floor-feasible set holds selection. Passing the screen establishes readiness for domain assessment; the responsible reviewer determines compatibility with the mandate and applicable duties before authorization. The screen is a proposed review aid whose substantive adequacy requires validation. EU equivalent-service and instrument assessment additionally considers cost and complexity, with ESMA guidance addressing documented justification and review of costlier equivalent products (European Commission, 2017, art. 54(9); European Securities and Markets Authority, 2022, paras. 91–95). Those obligations and the client-interest and conflict rules supply the legal basis for domain assessment (European Securities and Markets Authority, n.d., paras. 1 and 10; Swiss Confederation, 2018, arts. 25–26).

The chain in Equation (5) supplies a counterfactual decomposition of departures from an intended decision. Let $F^0(a)$ be the weighted sum of Equation (9) under the authorized mandate and complete information (θ_C, θ_W) . Fix a common deterministic rule for selecting among tied maximisers. Let a^0 be the choice made under those conditions, a^D the choice the same rule makes from what the client disclosed, a^T the choice it makes from the specification the manager transmitted, and a^X the recommendation the system issues. For stages with a selected alternative, then

$$F^0(a^0) - F^0(a^X) = \underbrace{F^0(a^0) - F^0(a^D)}_{\text{disclosure}} + \underbrace{F^0(a^D) - F^0(a^T)}_{\text{translation}} + \underbrace{F^0(a^T) - F^0(a^X)}_{\text{execution}}. \quad (15)$$

The identity is elementary. Under these counterfactual conventions, its terms locate review responsibilities. The disclosure term concerns the client–manager interface and can motivate a permitted question; the translation term motivates review of the manager’s specification; the execution term motivates conformance testing. A choice that the original mandate excludes is recorded as a violation at the interface where it arose. Holds remain separate dispositions. Single terms can offset one another; for originally feasible choices their sum is nonnegative.

A system that selects the transmitted optimum using the same tie rule makes $a^X = a^T$ and drives the execution term to zero. Membership in the argmax of Equation (4) alone permits different tied choices. For example, transmitting $\lambda = 2/3$ makes B_1 and C_2 tie in the construction in fig. 3, while the authorized weight $3/4$ gives them values $7/8$ and $13/16$. Choosing $a^T = B_1$ and faithfully issuing $a^X = C_2$ assigns $1/16$ to execution under F^0 . If the executed tie rule is unspecified, report the range of $F^0(a) - F^0(a^X)$ over transmitted maximisers; this example gives $[0, 1/16]$. Regret measured under the transmitted objective is zero for every faithful maximiser. Responsibility for the specification remains with the manager under either reporting convention.

The specification fixes five things: two outcome functions that stay separate through every calculation, an authorized trade-off between them, admissibility checked before any comparison, an owner for every premise, and approval bound to the current evidence and policy version. The client confirms preferences, financial objectives and agreed service terms. The manager proposes delivery costs, commercial criteria, scales, the trade-off weight and a reference service. A

domain reviewer independent of that commercial choice assesses the policy and each decision's compatibility with the client mandate and applicable duties, including both benchmark and weight concessions in Δ_C^{tot} . Domain review also establishes the service and evidence requirements that permission checks apply before comparison and issuance. The responsible manager authorizes the reviewed policy version. Client agreement records preferences and service scope; applicable legal duties remain binding constraints (European Securities and Markets Authority, [n.d.](#), paras. 1 and 10).

In the restricted client selector, the manager specifies candidates, facet utilities, references and clarification thresholds; calculation tools apply Pareto filtering and assess the declared answer partitions. The agent identifies the premises supporting its proposal, while the evidence service tracks their validity and source ancestry. Human approval binds the candidate to the current evidence and policy version.

These responsibilities can be implemented through a single agent, several agents or a deterministic procedure. The analytical constructions and simulations in section 6 examine the corresponding decision and evidence rules. Conformance establishes fidelity to the authorized specification; mandate and domain review assess its adequacy. A material change in objectives, alternatives, evidence criteria or permitted actions requires policy review and renewed assessment.

The finite client selector instantiates Equation (9) when the workflow and manager value are fixed and the floors exclude no feasible candidate. At an evidence state satisfying $G_{\text{req}} = 1$, let admissibility match $\mathcal{F}(E)$ and client value be a strictly increasing normalisation of the robust Nash product. The maximisers then coincide with Equation (18); restricting candidates to $\mathcal{P}(E)$ gives the Pareto selector. Figure 2 and Algorithm 1 specify clarification and approval for this restricted case. The facets represent dimensions within the client outcome function. Discretionary clarification for the full joint objective, including a contingent sequence of questions, requires a further model of both parties' outcomes and answer probabilities or a declared robust question policy.

The evidence representation adapts the assertion/source distinction and reliance-event model developed for legal identity assurance (Kurz, 2026). Each premise carries its origin, supporting evidence, period of validity and use in the recommendation. Evidence sufficiency depends on the premise: a declared preference can be relevant on the client's authority, while a claim about an external liability may require corroboration.

Within the restricted client-selection procedure, facet agents represent client objectives through manager-approved utilities, reference values and evidence dependencies. The shared policy fixes their relation to the joint specification; agent decomposition must preserve that representation. Organisational roles supply hard legal, risk and data constraints. Evidence records retain source ancestry, so outputs derived from one premise share an origin and its subsequent invalidation (Kurz, 2026). Accepted factual evidence filters the enumerated joint-state set while preserving its declared dependencies.

Each context gap has a purpose, a decision relevance assessment and a permitted resolution route. The evidence officer assesses an answer's evidential adequacy and processing authorization. Acceptance filters the state set; refusal preserves the gap and its applicable route. Client willingness, evidential sufficiency and lawful processing remain separate decisions, including for special-category and third-party data (European Parliament and Council of the European Union, 2016, arts. 5–7, 9 and 14).

Let X be a finite candidate set and E the accepted, currently valid evidence. The policy supplies a finite joint-state set; compatible evidence selects the non-empty subset $\Theta(E)$. A conflict that leaves this subset empty holds the decision for resolution. For each active facet i , Equation (16) takes the lowest utility across the evidence-compatible states:

$$\underline{U}_i(x; E) = \min_{\theta \in \Theta(E)} U_i(x; \theta). \quad (16)$$

Here $U_i(x; \theta)$ is facet i 's utility in state θ . Different facets can attain their minima in different states, so the aggregation protects componentwise lower bounds. Let g_L, g_R, g_D denote legal, risk and data-constraint residuals, with values at most zero indicating satisfaction. For fixed, state-independent reference values d_i , Equation (17) requires statewide admissibility and nonnegative facet surpluses:

$$\mathcal{F}(E) = \left\{ x \in X : \begin{array}{l} g_h(x; \theta) \leq 0 \quad \forall h \in \{L, R, D\}, \theta \in \Theta(E), \\ \underline{U}_i(x; E) \geq d_i \quad \forall i \end{array} \right\}. \quad (17)$$

Each d_i is the facet's declared reference utility. Evidence restriction can enlarge $\mathcal{F}(E)$ under fixed primitives, while required-information gates continue to govern issuance.

In the evaluated procedure, k fixed facets represent client objectives. Equation (18) defines the provisional robust maximiser set $W(E)$ using the client Nash-product rule:

$$W(E) = \arg \max_{x \in \mathcal{F}(E)} \prod_{i=1}^k (\underline{U}_i(x; E) - d_i). \quad (18)$$

The product adapts the Nash criterion to finite multicriteria selection (Nash, 1950); $W(E)$ retains all exact ties. Facets are dimensions of the client's evaluation, with cardinal utilities, overlap policy and reference values fixed by the manager. A bargaining interpretation would require separate strategic and feasible-set assumptions. An uncertain reference requires a further convention, such as statewise surpluses. An empty $\mathcal{F}(E)$ holds the decision for human evidence resolution, revised alternatives or scope; the unfiltered rule retains zero-product ties for human adjudication.

The selection rule first removes candidates dominated across the current state/facet utility entries. Write $\mathcal{P}(E)$ for those $x \in \mathcal{F}(E)$ for which no $y \in \mathcal{F}(E)$ satisfies $U_i(y; \theta) \geq U_i(x; \theta)$ for every i, θ , with strict inequality for at least one entry. Equation (19) applies the Nash-product comparison to this statewise-undominated set:

$$W^P(E) = \arg \max_{x \in \mathcal{P}(E)} N_E(x), \quad N_E(x) = \prod_i (U_i(x; E) - d_i). \quad (19)$$

The score $N_E(x)$ is unchanged by the filter; $\mathcal{P}(E)$ restricts its domain using the declared state/facet utilities. This removes dominated zero-product choices and dominance concealed by equal robust scores. Remaining ties receive explicit human adjudication.

Let $\mathcal{Q}(E)$ index the unresolved inputs in the fixed declared schema whose possible recorded values induce nontrivial partitions. For input q , $\mathcal{B}_q(E)$ partitions $\Theta(E)$ by its possible recorded values. For each nonempty answer cell B , calculate $\mathcal{F}(B)$, $W^P(B)$ and N_B using the same primitives. With nonempty current feasibility, every current candidate remains feasible in each cell. For a current maximiser x , Equations (20) and (21) define separate monetary and client-criterion consequences of changing the selection after one declared input is resolved:

$$S_q^M(x; E) = \max_{\substack{B \in \mathcal{B}_q(E) \\ \theta \in B}} \left[\ell(x, \theta) - \max_{y \in W^P(B)} \ell(y, \theta) \right]_+, \quad (20)$$

$$S_q^C(x; E) = \max_{B \in \mathcal{B}_q(E)} \left[\max_{y \in W^P(B)} N_B(y) - N_B(x) \right]_+. \quad (21)$$

Here $\ell(x, \theta)$ is a separately declared euro loss and $[a]_+ = \max(a, 0)$. Both candidates in S_q^C are evaluated on the same answer cell B ; a rise in the current candidate's own lower bound alone contributes no selection gain. The inner maximum in S_q^M requires improvement over the highest-loss answer-cell maximiser, so every resolved tie is retained. These are possible conditional gains under the declared criterion and tie convention. They relate to the value of information in decision analysis (Howard, 1966) as an upper bound. With answer probabilities $p(B)$ and a question cost κ_q on the same scale, the expected net gain from asking q before settling on x is

$$\text{VOI}_q(x; E) = \sum_{B \in \mathcal{B}_q(E)} p(B) \left[\max_{y \in W^P(B)} N_B(y) - N_B(x) \right] - \kappa_q \leq S_q^C(x; E) - \kappa_q. \quad (22)$$

Each bracket is nonnegative, since x stays feasible in every cell, and none exceeds the largest one, which is S_q^C . A question that stays below the threshold τ_C therefore also stays below it in expectation, whatever the probabilities. The converse needs the probabilities, which the present policy leaves undeclared.

The policy declares thresholds $\tau_M, \tau_C \geq 0$. Input q blocks candidate x when either consequence is positive and reaches its corresponding threshold. Equation (23) retains current maximisers whose consequences remain zero or below the corresponding threshold for every unresolved input:

$$\mathcal{R}(E) = \left\{ x \in W^P(E) : \begin{array}{l} \forall q \in \mathcal{Q}(E), r \in \{M, C\}, \\ S_q^r(x; E) = 0 \text{ or } S_q^r(x; E) < \tau_r \end{array} \right\}. \quad (23)$$

An empty partition collection leaves every member of $W^P(E)$ eligible for review. The client threshold τ_C uses Nash-product units. Positive affine transformations of utilities and references preserve selection; they preserve this test when τ_C receives the corresponding product-scale multiplier. Monetary and client thresholds require separate materiality rationales.

The responsible manager, supported by domain review, must authorize both thresholds and their materiality rationale before execution. Monetary materiality is specified in euros for the mandate; client-criterion materiality requires comparisons of preference consequences on the declared utility scales. Zero thresholds block every positive consequence. The constructed checks stipulate threshold values to examine policy behaviour; practical calibration remains prospective.

Mandatory information, legal and provenance conditions govern every candidate. After these gates pass, a nonempty $\mathcal{R}(E)$ permits review of its members. If feasibility is nonempty but every current maximiser is blocked, the scheduler asks the first permitted unresolved input that blocks at least one current maximiser, within the fixed schema order and budget. It assesses declared partitions even when a question is unavailable or refused; those restrictions govern the lawful resolution route. Mandatory refusal, exhausted resources or unavailable consequential questions preserve a hold.

A new accepted answer triggers reassessment, and a service-scope revision triggers a fresh applicability assessment. Necessary information remains a precondition for the relevant EU advisory service after human escalation (European Securities and Markets Authority, 2016).

For comparison, we evaluate a global monetary clarification rule using the unfiltered $W(E)$ and singleton-state feasible and winner sets F_θ, W_θ . Equations (24) and (25) define its global change indicator and monetary-consequence measure:

$$\Pi(E) = \mathbf{1}\{\exists \theta \in \Theta(E) : (F_\theta, W_\theta) \neq (\mathcal{F}(E), W(E))\}, \quad (24)$$

$$S(E) = \max_{\substack{x \in W(E) \\ \theta \in \Theta(E)}} \left[\ell(x, \theta) - \max_{y \in W_\theta} \ell(y, \theta) \right]_+. \quad (25)$$

Here $\mathbf{1}$ is the indicator function, F_θ, W_θ are the singleton-state feasible and winner sets, and $\bar{s}$ is the euro threshold. This comparator holds when $\Pi(E) = 1$, $S(E) > 0$ and $S(E) \geq \bar{s}$, alongside mandatory gates. Equations (20)–(23) instead assess each candidate over declared answer partitions and include client-criterion consequences. Both rules use the same finite-state representation and evidence criteria. The model experiments retain their executed client-selection policies; the partition rule is assessed through deterministic constructions.

Figure 2 connects the selection and clarification rules to the recommendation workflow. Mandatory checks precede comparison, consequential gaps prompt clarification or referral, and approval binds the selected candidate to the current evidence and policy version.

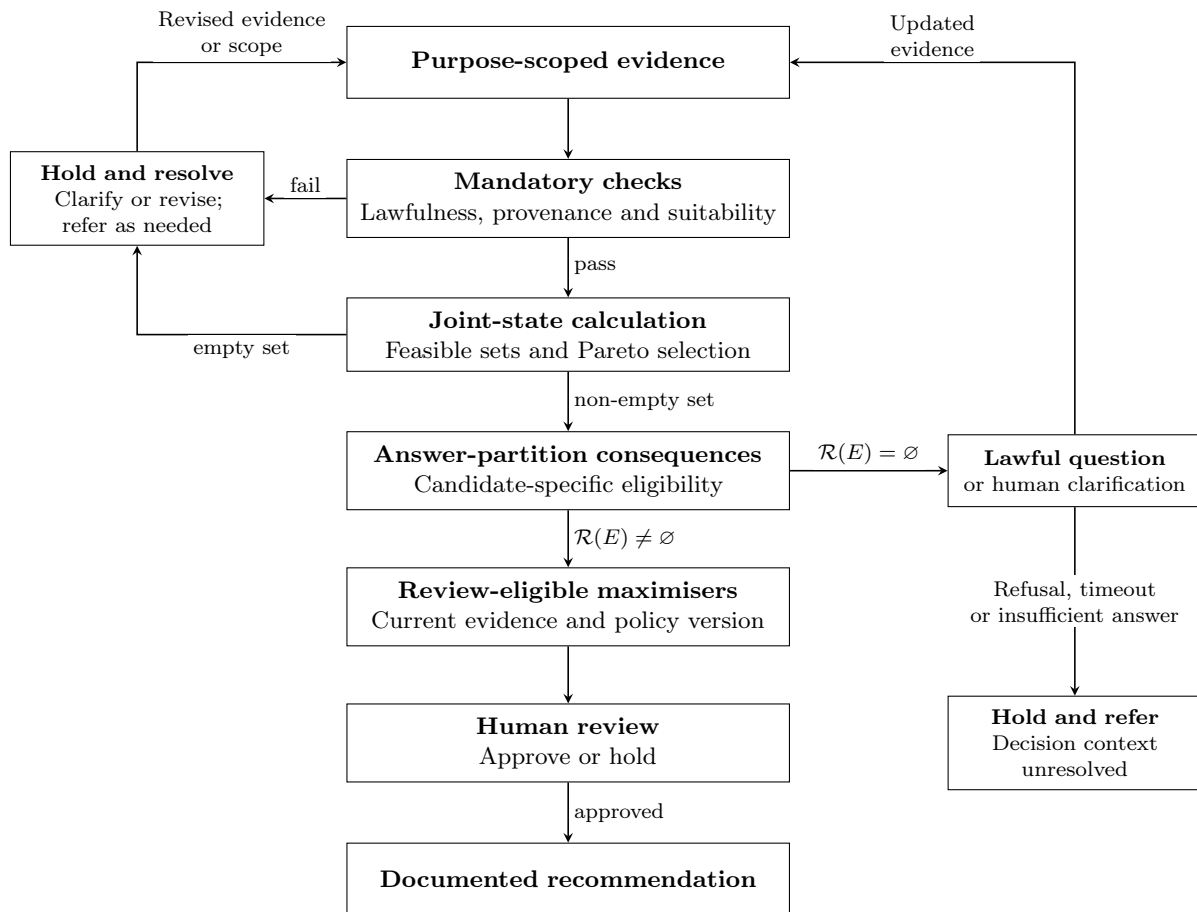


Figure 2: Evidence-dependent calculation and version-bound approval in the restricted client selector.

Note. $\mathcal{R}(E)$ is the set of review-eligible maximisers at evidence state E . Approval binds a candidate to the current evidence and policy version. Source: the authors' recommendation procedure.

Human review receives the review-eligible maximisers, unresolved inputs, candidate-specific consequence diagnostics, gate outcomes and evidence/policy version. Approval binds a specific candidate to that version. Evidence acceptance, expiry, revocation and policy revision advance the version and clear approval. Human review is an architectural requirement; automated-decision duties require separate assessment under applicable data-protection law (European Parliament and Council of the European Union, 2016, art. 22; Federal Data Protection and Information Commissioner, n.d.).

Context coverage κ is the fraction of inputs in the manager-approved schema that satisfy their evidence criteria. The schema's completeness requires external validation. The record reports coverage, itemised gaps, gate outcomes and approval separately (Kurz, 2026).

Each transition records the version, actor role, source references and outcome. Revocation or expiry invalidates dependent premises transitively. Evidence acceptance, policy changes and approval are restricted to their declared roles. The reliance record distinguishes a premise's origin from the act of relying on it (Kurz, 2026). Role checks use trusted caller inputs. The event-time invariant assumes serialized transitions; deployment requires an atomic or version-conditioned issuance commit alongside identity, access, tamper-resistance and jurisdiction-specific storage controls.

The transition contract uses a decision version v identifying current evidence and policy, a gate predicate G_v , and a recorded human approval $A_v(x)$. Equation (26) permits issuance only when the current gates pass and a review-eligible maximiser has human approval bound to the same version:

$$\text{issue}_v(x) \implies G_v = \text{pass} \wedge x \in \mathcal{R}(E_v) \wedge A_v(x) = 1. \quad (26)$$

The predicate G_v includes mandatory checks, consistent nonempty state support and robust feasibility; $A_v(x)$ binds human approval to candidate x and version v . Issuance rechecks these conditions and current review eligibility.

Algorithm 1 routes unresolved mandatory or consequential inputs to permitted questioning or a hold, and requires reassessment before renewed approval after evidence or policy changes.

Algorithm 1: *Versioned recommendation procedure for the restricted client selector*

Input: Fixed policy, finite candidate and joint-state sets, evidence, question budget

```

1: Validate role, evidence dependencies and current policy version
2: Filter joint states by accepted valid evidence
3: if evidence is inconsistent or the joint-state set is empty then
4:   Hold for evidence resolution
5: else if a mandatory gate fails then
6:   Request an unresolved permitted input within budget, or hold
7: else
8:   Compute  $\mathcal{F}(E)$ ,  $\mathcal{P}(E)$  and  $W^P(E)$  internally
9:   if  $\mathcal{F}(E) = \emptyset$  then
10:    Hold for human evidence resolution, revised alternatives or scope
11:   else
12:    Assess declared answer partitions and compute  $\mathcal{R}(E)$ 
13:    if  $\mathcal{R}(E) = \emptyset$  then
14:      Request a consequential permitted input within budget, or hold
15:    else
16:      Present  $\mathcal{R}(E)$ , consequences, evidence, gaps and version for review
17:    end if
18:   end if
19: end if
20: On evidence or policy change: invalidate approval and reassess
21: On approval: bind the selected review-eligible maximiser to the current version
22: On issuance: repeat checks and require current bound approval

```

Proposition 3 (gate preservation). Under the stated transition contract and trusted role boundary, every issuance event satisfies Equation (26) at its event time.

The initial state carries no approval. Evidence and policy transitions clear approval and advance the version. Questioning and refusal leave issuance dependent on a subsequent successful assessment. The approval transition binds a selected review-eligible maximiser to the assessed version. The issuance transition re-evaluates the gates and checks membership in the current eligible set and approval/version equality. Each transition preserves the condition that issuance is possible only through these checks, establishing the event-time invariant by induction over the transition sequence.

Proposition 4 (robust feasible-set expansion). Fix X , all utilities and constraints, the active facets and their reference values. For non-empty joint-state sets $\Theta' \subseteq \Theta$, the feasible sets defined by Equation (17) satisfy $\mathcal{F}(\Theta) \subseteq \mathcal{F}(\Theta')$.

Any candidate feasible for every state in Θ satisfies the same constraints over its subset Θ' . Taking a utility minimum over that subset weakly raises each lower bound. Every reference-value inequality satisfied under Θ remains satisfied under Θ' , so each originally feasible candidate remains feasible.

Recommendation permission also depends on mandatory information and versioned approval. An update can reveal a new liability, alter a constraint or invalidate a source; those changes require renewed assessment under the updated model. The inclusion concerns $\mathcal{F}$; Pareto membership, maximiser membership and review eligibility can change after a restriction of support.

Proposition 5 (undominated issuance). Under the issuance contract, an issued candidate is undominated among current robust-feasible candidates with respect to all declared state/facet utility entries.

Equation (26) requires membership in $\mathcal{R}(E)$, which is a subset of $W^P(E)$ and of $\mathcal{P}(E)$. Membership in $\mathcal{P}(E)$ excludes the stated dominance relation by definition, including when every feasible Nash product is zero.

The guarantee depends on supplied utilities and states, whose adequacy for the mandate requires domain assessment.

6 Results

The analytical constructions establish how information, weights and reference floors affect selection. The portfolio–workflow simulation then measures those effects on the declared client and manager outcomes. Further results concern client-criterion sensitivity, clarification, evidence changes and execution by deterministic and model-based workflows.

The liability construction isolates how undisclosed client information changes selection and permission to issue advice. Client value in this example is the product of a return utility and a liquidity utility, the rule of Equation (18). The synthetic client is an entrepreneur seeking advice on EUR 4 million of investment proceeds after a business sale. Three portfolio alternatives allocate different amounts to assets available within the next 12 months. The amount q of a possible payment obligation within that horizon is the context variable. For this illustration, the service’s suitability assessment requires that amount, and each candidate must provide liquid assets $L(x) \geq q$. All other legal, data and risk checks are assumed satisfied. Table 1 fixes the alternatives, with return and liquidity utility scores chosen solely to illustrate the mechanism and reference utilities $d_1 = d_2 = 0$.

Table 1: *Synthetic alternatives and assumed facet utilities.*

Portfolio	Liquid assets EUR million	Other assets EUR million	Return utility U_1	Liquidity utility U_2	Nash product $U_1 U_2$
A	0.4	3.6	9	2	18
B	1.2	2.8	6	6	36
C	2.0	2.0	3	9	27

B maximises the product at an established obligation of EUR 0.2 million; C alone is feasible at EUR 1.8 million. With unresolved support $\{0.2, 1.8\}$ million, C covers both amounts, while the mandatory-information gate holds issuance. An obligation above EUR 2 million requires revised alternatives or a continued hold. The unresolved input is identical whether the undisclosed obligation is EUR 0.2 million or EUR 1.8 million, while the fully informed winners are B and C. This instantiates Proposition 1: robust C covers both amounts, but an accepted answer is needed to distinguish their client-optimal selections.

The joint construction adds the manager and the delivery workflow to this example, so that both declared evaluations take effect. Client values normalize the portfolio products in table 1 by 36. For an operational interpretation of the constructed workflows, ω_1 uses general case review and ω_2 uses template-supported review. We assume reusable templates reduce delivery burden for A and C, while B requires additional exception review under ω_2 . Both workflows are assumed to deliver the same assessed client outcome for a given portfolio. Manager values are stipulated as $\hat{V}_W(x, \omega) = 1 - c(x, \omega)$, where c is a normalized burden index: $(1/4, 1/2, 1/2)$ for A, B and C under ω_1 , and $(0, 3/4, 0)$ under ω_2 . Zero denotes the least burden on this illustrative scale. Empirical calibration of the burden index remains prospective.

The established EUR 0.2 million obligation makes all three portfolios financially feasible. The example stipulates that all six pairs pass the base admissibility conditions; domain review must establish which pairs can actually satisfy a mandate’s client-interest duties. The client and manager floors are the values of reference service (C, ω_1) , namely $(3/4, 1/2)$. The floors retain (B, ω_1) , (C, ω_1) and (C, ω_2) .

Figure 3 shows that (B, ω_1) and (C, ω_2) are the two efficient pairs meeting both floors; choosing between them requires a client–manager trade-off.

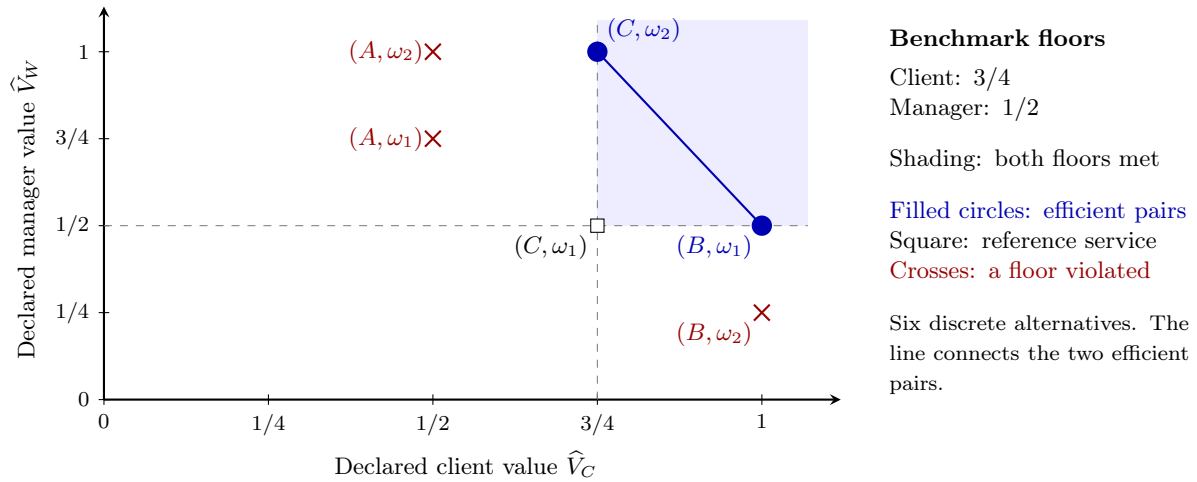


Figure 3: *Client and manager values of the constructed alternatives.*

Note. The six portfolio–workflow pairs use the client values and workflow burdens defined in the text. Shading marks the benchmark region; the connecting segment indicates the trade-off between the two efficient discrete alternatives. Source: exact joint-objective construction.

At $\lambda = 1/2$, Equation (9) selects (C, ω_2) with score $7/8$, compared with $3/4$ for (B, ω_1) . At $\lambda = 3/4$, (B, ω_1) wins with $7/8$, compared with $13/16$ for (C, ω_2) . The two tie at $\lambda = 2/3$, each scoring $5/6$. At the lower weight the selected pair preserves the client benchmark and improves the manager evaluation; at the higher weight it preserves the manager benchmark and improves the client evaluation. The weight makes an explicit choice along the declared trade-off, while the floors restrict permissible sacrifices.

In the terms of Proposition 2, programme (C) selects (B, ω_1) , programme (W) selects (C, ω_2) , and both thresholds equal $2/3$. At $\lambda = 1/2$ the concession pair of Equation (10) is $\Delta_C = 1/4$ and $\Delta_W = 1/2$, twice the minimum that Equation (11) requires.

Figure 4 shows selection moving from (C, ω_2) to (B, ω_1) as the client weight crosses $\lambda = 2/3$, where the two tie.

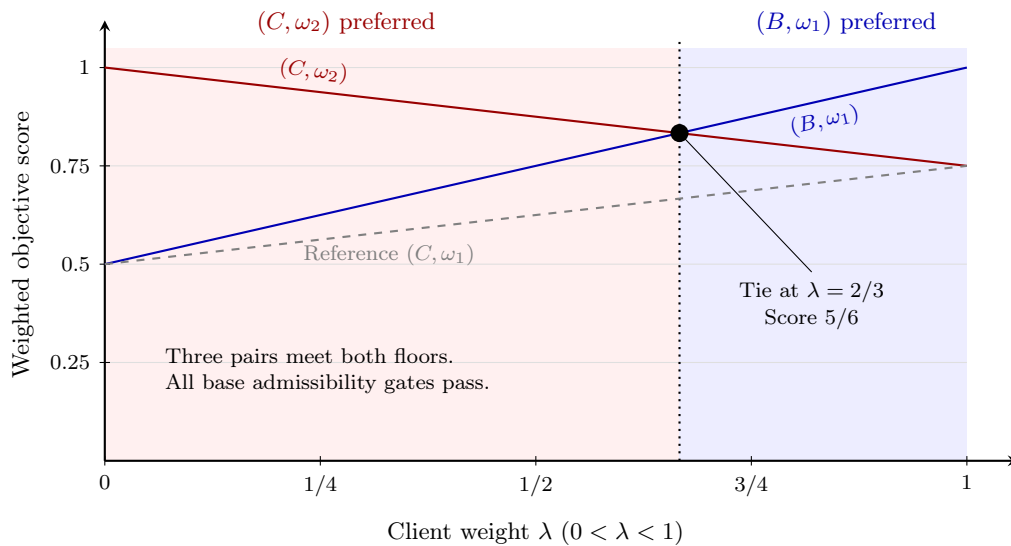


Figure 4: *Objective scores across client weights.*

Note. The three pairs meet both benchmark floors, with all base admissibility gates satisfied. Endpoint values show limits of the specified open weight interval. Source: exact joint-objective construction.

A capacity constraint excluding (C, ω_2) makes (B, ω_1) the winner at $\lambda = 1/2$, even though the excluded alternative has the higher objective score. Missing required information holds every alternative. A further capacity construction admits only (A, ω_2) and (B, ω_2) ; each violates one floor, so the joint problem has no feasible solution and holds. The reference service remains the evaluation benchmark while capacity makes it unavailable. Two further capacity settings isolate each floor’s effect on selection. With (C, ω_2) unavailable and $\lambda = 1/4$, an objective-only comparison selects (A, ω_2) at

7/8; the client floor excludes it, and the constrained rule selects (B, ω_1) at 5/8. With only (B, ω_2) and (C, ω_1) available and $\lambda = 3/4$, the objective-only winner is (B, ω_2) at 13/16; the manager floor excludes it, and (C, ω_1) wins at 11/16. The objective-only calculations isolate the floors' effects; the operative rule enforces both floors throughout. The eight configurations use constructed manager service evaluations; observed workflow expenditures enter the separate resource comparison below.

The two-pair capacity case gives a concrete review-screen test. Suppose the client confirms both service options and an independent cost assessment establishes their viability. With (B, ω_2) and (C, ω_1) hard-admissible, the client comparator before floors is $a_H = (B, \omega_2)$, while the manager floor leaves $a_C = a^* = (C, \omega_1)$. Thus $\Delta_C = 0$, $\Delta_C^{\text{bench}} = 1/4$ and $\Delta_C^{\text{tot}} = 1/4$. With zero omitted client benefit, Equation (14) returns the policy for revision even if policy review has been recorded. Choosing (B, ω_2) as the agreed reference instead gives floors $(1, 1/4)$, selects that pair and reduces both concessions to zero.

A further construction isolates a delivery restriction. Let three services have client–manager values $(10, 0)$, $(9, 1)$ and $(1, 20)$, zero floors and $\lambda = 1/10$. Their scores are 1, 9/5 and 181/10, so the third is selected. If a documented delivery requirement excludes the first service, it enters admissibility and the second becomes $a_H = a_C$. The remaining client concession is 8, which requires its own supported benefit assessment. These exact constructions also check that zero concession still requires policy review, that benefit evidence identifies the current comparator and policy, and that empty feasible sets hold. They implement the proposed review screen alongside the separate simulation.

The adversarial variations retain both party evaluations and both benchmark floors. Required-input omission produces a hold; accepted support refinement changes the robust admissible set; weight and reference translations change the selected pair. Table 2 reports these effects using $B_1 = (B, \omega_1)$, $C_1 = (C, \omega_1)$ and $C_2 = (C, \omega_2)$. The support-refinement condition uses a separate case policy: certified liability bounds satisfy its required-information gate, and the exact amount is optional. Except where varied explicitly, the reference is C_1 , values are those of fig. 3, and the weight is 1/2. All 15 exact outcomes matched their hand-derived witnesses; the six common configurations matched the retained evaluator.

Table 2: *Controlled analytical stresses on the full joint objective.*

Variation	Controlled change	Joint outcome
Required evidence	Established input becomes missing at $\lambda = 3/4$.	B_1 to hold.
Liability support	Accepted support narrows from $\{0.2, 1.8\}$ to $\{0.2\}$ million EUR at $\lambda = 3/4$.	C_2 to B_1 .
Weight translation	Authorized $\lambda = 3/4$ is transmitted as 1/2.	B_1 to C_2 .
Reference translation	Authorized reference C_1 is transmitted as B_1 .	C_2 to B_1 .
Client-value constraint	Add a hard requirement $\hat{V}_C \geq 1$.	C_2 excluded; B_1 selected.
Workflow effect	Reduce $\hat{V}_C(C_2)$ from 3/4 to 1/2, holding other values fixed.	C_2 fails its client floor; B_1 selected.
Affine representation	Transform $\hat{V}_C$ to $2\hat{V}_C + 7$ and $\hat{V}_W$ to $3\hat{V}_W - 2$, including benchmarks.	At 3/4: B_1, C_2 tie. At 9/11: B_1 restored.

The hard client-value requirement excludes C_2 before scalarisation even though it has the higher base weighted score. This checks precedence of an authorized constraint; institution-specific legal review supplies its substantive content. The workflow variation makes client value depend on delivery, relaxing the original example's invariance assumption. Under the affine transformation, changing λ from 3/4 to 9/11 preserves every pairwise score difference up to a common positive factor 24/11. Leaving the weight unchanged instead changes the effective policy. Inputs and preferences in these stress cases are authored.

The simulation illustrates the joint objective on mandates with financial content: both evaluations now depend on financial and service quantities. Constructed mandates, scales and weights make their effects explicit in a worked implementation. For each pair $a = (x, \omega)$ and scenario, define payment-adjusted economic wealth as $Y = A_T + H_{\text{scheduled}} + L_{\text{due}} - B_{\text{unpaid}}$. Here A_T is terminal financial wealth, $H_{\text{scheduled}}$ scheduled household spending, L_{due} the liability valued at its payment-date exchange rate, and B_{unpaid} total unpaid bills, including service charges. Fees and transaction costs remain losses in this accounting measure; funding adequacy is enforced separately. Equation (27) gives the declared client and manager evaluations used in the joint objective:

$$\begin{aligned}\hat{V}_C(a) &= \frac{\mathbb{E}[Y(a)] - A_0 - \rho_C(\mathbb{E}[Y(a)] - \min_{\theta \in \Theta} Y(a; \theta)) + v_C s_\omega}{0.1A_0}, \\ \hat{V}_W(a) &= \frac{\pi(a)}{0.01A_0} + 0.1s_\omega.\end{aligned}\tag{27}$$

Here A_0 is initial wealth, ρ_C the client's declared downside weight, v_C its euro value per unit of service quality s_ω , and $\pi(a)$ manager fee income less personnel and fixed costs. Expectations use three authored scenarios with probabilities

0.25, 0.50 and 0.25; Θ is their support. Both reference floors use these same definitions. The client and manager scales are fixed at 10% and 1% of initial wealth. Hard constraints separately enforce liquidity reserves, loss capacity, issuer concentration, funded obligations and supervisory capacity.

The four complete evaluation mandates admit ten, five, ten and seven pairs, respectively, after applying hard constraints and both reference floors. Each retains at least two undominated client–manager trade-offs. Both floors admit every hard-feasible pair in all eight original and corrected states; their restrictive effects are assessed in the analytical constructions. At client weights 0.95, 0.90, 0.85 and 0.98, the business-sale, retirement, currency-liability and concentration mandates select growth/standard, balanced/advisory, growth/specialist and growth/advisory services. These names identify portfolio and delivery choices; the instruction treatments use the same agent architecture.

Table 3 gives the price of these weights by comparing each selection with the solution of programme (C) over the same feasible set. With client value scaled to $k_C = 0.1A_0$ and manager value to $k_W = 0.01A_0$, Equation (12) becomes $\varepsilon = 10(1 - \lambda)/\lambda$. In the two business-sale states the selected pair is also the client’s best. In the other six states the rule selects the advisory or specialist workflow where the client-best pair uses the standard workflow, and in two of the four mandates a manager euro counts for more than a client euro.

Table 3: *Client concession and manager gain against the client-best pair at the declared weights.*

Mandate	λ	ε	Selected and client-best workflow	Δ_C (EUR)	Manager gain (EUR)
Business sale	0.95	0.53	Standard; identical	0	0
Retirement drawdown	0.90	1.11	Advisory; standard	1,190	3,062
Currency liability	0.85	1.76	Specialist; standard	2,264	10,381
Concentrated holdings	0.98	0.20	Advisory; standard	1,178	6,656

Note. Original and corrected liability states of a mandate give the same differences. Δ_C is the difference in the declared client criterion of Equation (27), converted to euros; the manager gain is the difference in net income π . Clarification charges are zero in both selections. The portfolio is the same within each comparison.

The reference service holds cash at a fee of 0.4%. Every hard-feasible pair meets both of its floors, so $\mathcal{A}_{\text{feas}}(E)$ equals the hard-feasible set in these mandates and the floors restrict nothing here. Consequently $\Delta_C^{\text{bench}} = 0$ here, and table 3 reports the total as well as the weight concession. The six positive concessions identify the additional client-benefit evidence that Equation (14) would require in a substantive mandate review; the simulation implements the declared weighted selection.

Figure 5 shows the client–manager trade-off for the business-sale mandate. Lowering the transmitted client weight from 0.95 to 0.35 redirects the choice from growth/standard to growth/specialist, while both remain admissible.

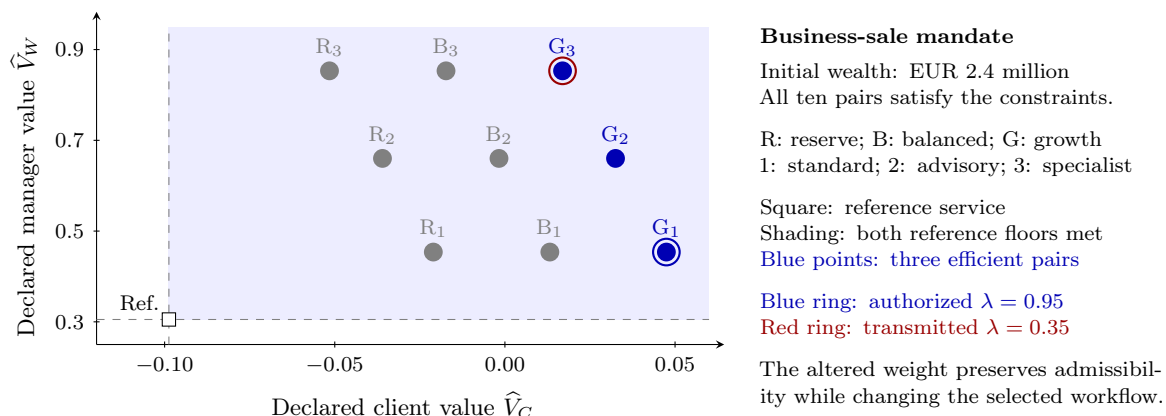


Figure 5: *Economic trade-offs and the effect of a mistranslated joint weight.*

Note. Complete heldout business-sale case, December 2021 decision. Points are the nine portfolio–workflow pairs and the reference service; both axes use the declared outcome scales. Client value includes the forecast risk penalty and service value. The two rings identify the choices under the authorized and altered weights. Each point is a discrete alternative. Source: frozen simulation inputs and exact valuation.

Eight evaluation states combine the original and corrected liability for each mandate. Faithful optimisation of the complete specifications preserves all eight original decisions. Setting the omitted liability to zero produces two recommendations that violate the original liquidity requirement; replacing omitted manager terms with zero hourly cost and 100 available supervisory hours produces two violations of supervision capacity. Transmitting a client weight

of 0.35 changes four optimal choices while each selected pair remains admissible under the original constraints. These are deterministic interventions in specification content. The optimiser follows the received objective in every condition; the original mandate supplies the reference for assessing fidelity. In the terms of Equation (15), the omitted liability acts on the disclosure term, the omitted manager terms and the mistranslated weight act on the translation term, and the execution term is zero under the common deterministic tie rule used in these controls. Figure 6 separates these constraint failures from admissible changes in the authorized trade-off.

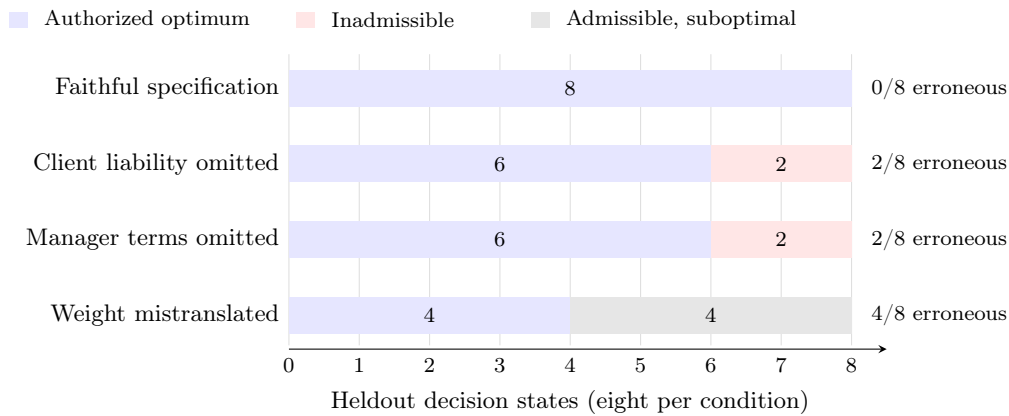


Figure 6: *Consequences of controlled defects in the transmitted specification.*

Note. Four mandate families, each evaluated in its complete and revised state. Exact optimization acts on each transmitted specification; outcomes are scored against the authoritative mandate. Categories partition the eight states in each row. Client omission substitutes zero liability and produces liquidity violations; manager omission substitutes zero hourly cost and 100 supervisory hours and produces capacity violations. Complete and revised states share their base mandate. Source: frozen heldout representation controls.

The instruction comparison keeps the declared decision content fixed across three forms. Each reaches all 32 prescribed terminal decisions: 28 issuances and four justified holds. Every issuance preserves the declared numerical mandate, current simulated approval and the joint optimum; unsupported proposals and violations of the encoded constraints are zero. Regret under the declared objective is zero for all 84 issuances, alongside 12 required holds. The equal-output controls assign identical original values to every tied choice, so tie selection leaves the attribution unchanged in these records. Professional prose and nested-delegation instructions each use 20 questions; the explicit joint-objective treatment uses 26. Its six additional questions incur EUR 450 in nominal client charges and EUR 672 in manager labour cost. Table 4 separates these constructed clarification costs from metered model expenditure.

Table 4: *Execution of the joint objective under three instruction forms.*

Measure	Professional prose	Joint objective	Nested delegation
Correct terminal decisions	32/32	32/32	32/32
Issued recommendations / required holds	28/4	28/4	28/4
Unsafe raw proposals	0	0	0
Issued numerical mandate or constraint violations	0	0	0
Questions	20	26	20
Client question charges (EUR)	1,500	1,950	1,500
Manager question labour (EUR)	2,352	3,024	2,352
Reported model cost (USD)	6.3146	7.3317	6.3162

Note. Four constructed mandates, eight dependent information conditions per mandate, one repetition. Objectives, evidence access, arithmetic and approval controls are shared. Each question incurs a constructed EUR 75 client charge and half an hour of manager labour at the mandate-specific rate. All costs include holds; model expenditure is metered separately.

The comparison contains 96 model episodes; reconstruction reproduces their decisions, evidence transitions and economic outcomes. A public-input reference policy reaches all 64 development and evaluation conditions. The common runtime supplies authoritative source ownership and enforces supported premises, admissibility and current approval. The three instruction forms illustrate equivalent execution of the shared decision rule within those controls.

The subsequent market-path calculation keeps each selected pair fixed and replaces the declared exchange-rate and yield paths with the 2022 observations, retaining the three constructed growth scenarios and their probabilities. The client outcome is Y plus declared service value under the accounting definition above. Each reference service bears the same clarification expenditure as its corresponding episode. Averaging the eight distinct original and corrected mandate

states equally gives a client measure EUR 77,967 below the reference service and manager net income EUR 8,816 above it in every treatment. Manager net income follows from fee income less personnel and fixed costs:

$$\pi(a) = f_{\omega}A_0 - c_W\left(\tau_{\omega} + \frac{n}{2}\right) - 500, \tag{28}$$

with workflow fee rate f_{ω} , hourly personnel cost c_W , workflow hours τ_{ω} , n questions at half an hour each and a fixed cost of EUR 500. No term depends on the market path, so the manager figure would be the same in any year; the client figure alone reflects 2022. Equal-output controls and repeated information conditions enter the decision counts; the economic mean uses each substantive state once. Question-cost totals report clarification burden separately. All obligations are funded. These dependent conditional outcomes differ from the risk-adjusted decision-time client criterion.

Figure 7 compares forecast and subsequent-path client outcomes using the same euro measure, relative to the same reference service. The five-year government yield rises from approximately -0.48% at the decision date to 2.45% at the end of 2022, exceeding the stipulated adverse-scenario increase of 1.5 percentage points. The reference service holds cash, whose accrual proxy also rises. This identifies a consequential mismatch between the declared market scenarios and subsequent observations. The outcome calculation preserves the constructed growth assumptions, so interpretation concerns that specified combination of historical and authored paths.

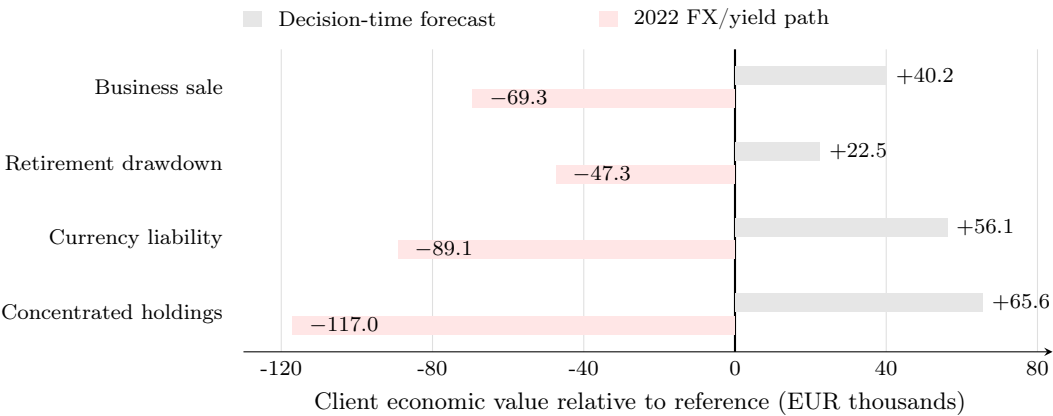


Figure 7: *Forecast and subsequent client consequences on the same economic scale.*

Note. One complete heldout case per family; all three instruction treatments select the same pair. Both bars measure payment-adjusted economic wealth plus constructed service value, relative to the reference, in EUR thousands. The forecast bar averages the three declared valuation scenarios. The 2022 bar uses historical FX and yield observations crossed with the same three constructed growth paths and weights (0.25, 0.50, 0.25). The plotted metric excludes the risk penalty used in decision-time client value. Source: frozen calculations and retained counterfactual cash flows.

The representation checks in table 5 hold the candidate set fixed while changing the aggregation specification. Duplicating the liquidity facet changes the complete-context products from (18, 36, 27) to (36, 216, 243) and selects C. Setting reference values to (0, 4) leaves B and C with products 12 and 15. Positive affine utility transformations with correspondingly transformed references preserve the winner: each facet surplus is multiplied by a positive constant, so every candidate product receives the same positive multiplier. For an exact duplicate facet, splitting an existing exponent between its copies preserves that facet’s product contribution; the exponent policy remains part of the declared representation.

Table 5: *Constructed sensitivity checks for the allocation example.*

Specification	A	B	C	Maximiser
Original product	18	36	27	B
Liquidity represented twice	36	216	243	C
Reference values (0, 4)	Infeasible	12	15	C
Equal-weight utility sum	11	12	12	B, C

Taking the original two-facet specification as the declared representation of the mandate, the duplication in table 5 changes its implemented weighting while retaining every original utility value and financial constraint. The resulting C choice satisfies the duplicated policy; the original policy selects B. The construction separates execution of a specified aggregation rule from preservation of the declared mandate representation. Which representation reflects a particular client’s preferences requires evidence beyond the stipulated utilities.

All 18 constructed aggregation cases retained two to four robust-feasible candidates. Of the 17 equal-weight comparisons, seven had identical winner sets, six had overlapping but unequal sets and four had disjoint sets. Fifteen cases had conflicting rankings among robust-feasible alternatives; these accounted for five identical sets, six overlapping sets and all four disjoint sets. The separately weighted sensitivity also produced disjoint winners. These counts describe related constructions selected to examine policy behaviour.

A strict disagreement occurs with normalized utilities $A=(1, 0.3)$, $B=(0.6, 0.6)$ and $C=(0.3, 0.9)$, zero references and common feasibility. Nash products of 0.30, 0.36 and 0.27 select B. Equal-weight sums of 0.65, 0.60 and 0.60 select A. The shared constraints admit both selections; their relative desirability depends on the client criterion used to assess the chosen scalarisation. Exact witnesses and reference checks reproduced all 18 Nash calculations.

A further construction isolates the uncertainty objective. With fixed zero reference values, candidate X has utility vectors (1, 9) and (9, 1) in two joint states; candidate Y has (2, 2) in both. Both candidates satisfy all constraints. The product of componentwise minima scores X at 1 and Y at 4, selecting Y. The worst joint-state product, $\min_{\theta} \prod_i (U_i(x; \theta) - d_i)$, scores X at 9 and Y at 4, selecting X. The two policies protect different quantities even though the joint states are identical. If X incurs zero monetary loss and Y incurs EUR 100 in each state, the unresolved robust choice has EUR 100 regret. Comparing only the singleton winners would miss this clarification opportunity because both states select X. Equation (24) also compares each resolved choice with the unresolved choice, detecting the difference; the executed case asks at a EUR 10 threshold.

In the constructed comparisons, statewise Pareto filtering removed a dominated candidate whose equal robust score had blocked an eligible alternative, and excluded a dominated choice when all Nash products were zero. Assessing answer cells detected a partial answer that changed the selected portfolio although the unresolved and all singleton choices agreed. Candidate-specific eligibility also allowed an unblocked maximiser to proceed while a tied peer awaited clarification. The full numerical witnesses and comparator traces are reported in the supplementary analytical results.

The partial-answer case shows why the answer partition matters. A declared input separates two of three states from the third. In that two-state cell, changing from A to B increases the client criterion by $4 - 1 = 3$ and reduces monetary loss by EUR 100. The rule asks at a client threshold of 1 or a monetary threshold of EUR 10. A further answer identifying either remaining state changes the preferred candidate back to A, with a client-criterion gain of $9 - 4 = 5$. Selection can therefore reverse as evidence arrives under the componentwise robust criterion.

The client-criterion test also detects consequences at equal monetary loss. Candidate A has utilities (1, 100) and (100, 1) in two states; B has (2, 2) in both, with zero references and zero losses. The unresolved products are 1 and 4, while each singleton selects A with product 100. Resolving the state yields a client-criterion gain of $100 - 4 = 96$, triggering a question at a threshold of at most 96. The monetary comparator leaves B eligible for review.

The diagnostics also left consequential information unresolved in two constructions. With flat Nash products and equal monetary losses, both diagnostics were zero although each answer changed which candidate was undominated. With two optional binary inputs, each individual diagnostic was zero while joint resolution changed the winner from B, with product 4, to A, with product 100. Zero individual consequences therefore leave both dominance changes and joint materiality possible. The supplementary witnesses give the utilities and answer partitions.

When an optional answer would restore an empty robust-feasible set, the procedure held for human evidence resolution. Its diagnostics require a feasible current maximiser, so the autonomous question request was rejected. Externally accepted evidence for either answer restored review of the corresponding candidate.

The finite procedure passed 18 checks comprising 26 constructed traces and 14 issuance events. Evidence withdrawal cleared approval even when an irrelevant source or a second independent corroborating source left mandatory support intact. Withdrawal of the sole required source held the decision until accepted replacement evidence restored support and a new approval authorized issuance. Expiry produced the same reassessment under the explicitly advanced logical clock; earlier issuance events remained in the record. Role checks rejected unauthorized evidence, policy and approval operations at the trusted application programming interface (API) boundary. These results establish conformance to the supplied evidence and feasibility labels.

Applying two Swiss service policies to the same allocation states changed permission to proceed. Knowledge and experience and all other relevant service inputs are stipulated as established. The portfolio-based policy classifies the material liability as required financial information; the transaction-specific policy limits required inputs to knowledge and experience. Both retain robust candidate C with zero unmet-liquidity loss, while the portfolio-based policy asks and then holds after refusal and the transaction-specific policy permits review. The difference follows from the declared

applicability mapping and conservative institutional hold, with the financial state set held fixed.

Table 6 contrasts the mandatory amount with an optional communication-format preference that leaves utilities and constraints unchanged.

Table 6: *Decision traces under four synthetic information conditions.*

Condition	Available context	Model outcome
Complete	Required information established; $q = 0.2$ million.	B proceeds to human review.
Missing pivotal input	Required amount unresolved between 0.2 and 1.8 million.	Ask and hold; reassess after an adequate answer.
Missing optional context	$q = 0.2$ million established; communication-format preference absent.	B proceeds to review; optional gap remains recorded.
Required disclosure withheld	Client declines to supply the necessary obligation information.	Hold and refer to the advisor; mandatory checks remain active.

A currency mismatch illustrates how a changed valuation can invalidate an earlier liquidity assessment. Assume liquid euro assets of EUR 1.2 million and an unhedged CHF 1.4 million payment obligation, with interest, taxes and execution costs omitted. Each date values the same assumed exposure separately. If e_t is the observed number of Swiss francs per euro, its euro cost is $q_t = 1,400,000/e_t$, and the liquidity shortfall is $\max(0, q_t - 1,200,000)$. Table 7 reports the calculation using observed ECB reference rates (European Central Bank, [n.d.-b](#)).

Table 7: *Constructed CHF liability valued at observed ECB rates.*

Date	CHF per EUR	Liability in EUR	Shortfall in EUR
14 January 2015	1.2010	1,165,695.25	0.00
15 January 2015	1.0280	1,361,867.70	161,867.70
16 January 2015	1.0128	1,382,306.48	182,306.48

Using the 14 January valuation on either subsequent date would retain an apparent surplus despite the shortfall at the current rate.

The workflow comparisons test whether supplied instructions produce the specified questions, holds and recommendations from partial evidence, while recording delivery resources. All workflows reached the 36 reference decisions in the structured pilot (table 8), with consistent selections across repetitions. Each made 18 question attempts. Unsafe proposals, issued violations, unnecessary holds and schema failures were zero.

Table 8: *Development-pilot outcomes and resources across 36 records per workflow.*

Measure	Deterministic	Single agent	Multi-agent
Correct terminal decisions	36/36	36/36	36/36
Correct permitted recommendations	24	24	24
Required holds	12	12	12
Unsafe raw proposals / proposals	0/24	0/24	0/24
Issued violations / issued outputs	0/24	0/24	0/24
Reported model cost (USD)	0.0000	0.9758	3.1034

The multi-agent workflow incurred 3.18 times the single agent’s reported model cost. The deterministic workflow used the shared calculator and incurred zero model charges. Implementation, local computation and professional review costs fall outside this accounting boundary.

The retirement construction sets assets at EUR 300,000 and annual pension income at EUR 24,000. Essential expenditure is EUR 5,000 per month and a separate EUR 60,000 payment falls within a three-year reserve horizon. Equation (29) gives the required liquid reserve R :

$$R = \max(b - p, 0)h + o = \max(60,000 - 24,000, 0) \cdot 3 + 60,000 = 168,000 \text{ EUR.} \quad (29)$$

Here b, p are annual expenditure and pension income, h is the horizon in years and o the separate obligation. Only R180, the EUR 180,000 reserve alternative, covers R ; the other reserves are EUR 60,000 and EUR 120,000. A competing questionnaire records EUR 3,000 monthly expenditure, implying a EUR 96,000 requirement and a different feasible set. The signed reconciliation establishes which figure governs. Expenditure and the separate obligation are required fields in the case policy.

The second construction separates the investment's price change from its complete cash flows. Equation (30) calculates available assets A and the net investment result P :

$$\begin{aligned} A &= s + d - f = 82,000 + 4 \cdot 5,500 - (2,000 + 4,000) = 98,000, \\ P &= A - i = 98,000 - 100,000 = -2,000. \end{aligned} \quad (30)$$

Here s denotes sale proceeds, d retained distributions, f separately debited fees and taxes, and i purchase cost, all in EUR. A near-term EUR 90,000 liability makes the EUR 80,000 reserve insufficient, while EUR 100,000 exceeds available assets A . R90, the EUR 90,000 alternative, satisfies both bounds. The EUR 2,000 loss records the investment result P ; feasibility uses available assets. All distributions remain available and the separate charges exhaust the stipulated fees and taxes.

Table 9 gives the four issuance and three hold conditions in each narrative family. The withdrawal grade requires valid initial proposal and approval before evidence loss; a blanket initial hold fails that sequence.

Table 9: *Information variants and conformance outcomes in both narrative families.*

Information variant	Required evidence transition	Terminal outcome
Complete evidence	Current records support calculation and approval.	Issue
Two jointly missing inputs	Obtain both required records; either adequate question order is accepted.	Issue
Refusal	A required input remains unresolved after the client declines.	Hold
Resolvable conflict	Signed reconciliation identifies the superseded assertion.	Issue
Unresolved conflict	The reconciliation route returns unavailable.	Hold
Withdrawal after approval	Evaluator invalidates the source dependencies and approval; workflow responds to the changed state.	Hold
Missing presentation preference	Financial inputs remain complete and unchanged.	Issue

Under fixed responses, all three interfaces reached the required outcomes in all 14 cases; reconstruction reproduced those outcomes.

The initial narrative comparison yielded 42, 30 and 22 correct decisions for deterministic, single-agent and multi-agent workflows, respectively (table 10). Model-workflow rates were 71.4% and 52.4%; correct recommendations numbered 14 and seven out of 24 issuance cases. Each workflow asked 30 relevant questions and zero redundant questions.

Table 10: *Initial narrative outcomes across three repetitions, with 42 records per workflow. Costs cover 37 matched records per workflow.*

Measure	Deterministic	Single agent	Multi-agent
Correct terminal decisions	42/42	30/42	22/42
Correct permitted recommendations	24	14	7
Correct required holds	18	16	15
Failed workflows	0	12	20
Inadmissible raw proposals / proposals	0/54	1/37	0/20
Reported model cost (USD)	0.0000	7.7222	13.3133

The calculator accepted exactly the required extracted fields and read remaining quantities from the public fixed inputs. Additional fields caused all rejected calculator calls, while repeated successful calculations also consumed the budget. Eight single-agent workflows and 16 multi-agent workflows exhausted the 24-call calculator budget. Response-format errors caused three further single-agent failures and four multi-agent failures. The remaining single-agent failure involved a missing matching calculation. These results measure calculation-interface adherence alongside document interpretation and evidence control.

One single-agent issuance attempt selected the feasible, currently approved R90 alternative but attached an unsupported source link to the fixed sale proceeds. Its premise set also lacked a matching successful calculation. The adapter blocked the attempt; scoring retained it as an inadmissible raw proposal. The enforced issuance invariant recorded zero violations for all workflows. The deterministic, single-agent and multi-agent workflows completed six, four and three of the six withdrawal trajectories, respectively. The other five model attempts failed before initial approval.

Six response-format failures lacked complete final usage totals, two in the single-agent workflow and four in the multi-agent workflow, affecting five case repetitions. Reported costs cover 37 matched records per workflow, while decision analysis includes all 42. The multi-agent cost was 1.72 times the single-agent cost within that subset (table 10), including metered failures and required holds.

The revised narrative interface completed all 126 assignments, with 42 correct decisions and all six withdrawal trajectories per workflow (table 11). This revision changed extraction and calculator interaction under the retained case policies; the Pareto and answer-partition rules in Algorithm 1 were assessed separately in deterministic constructions. Failures, validation errors and enforced gate violations were zero. All model proposals were admissible at their evidence versions. Each workflow asked 30 relevant questions and zero redundant questions. Every calculator event returned a single feasible candidate, so these results concern extraction and evidence control.

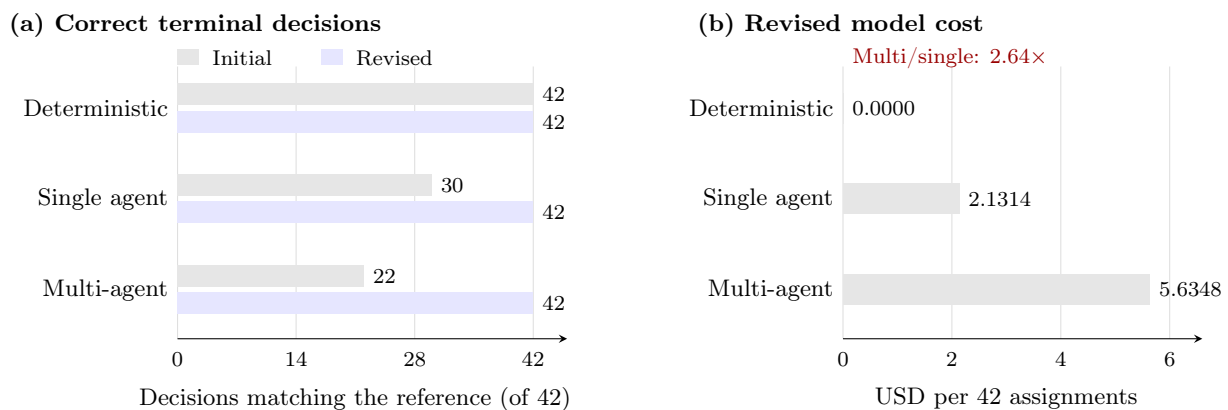
Table 11: *Revised-interface outcomes across three repetitions, with complete accounting for all 42 records per workflow.*

Measure	Deterministic	Single agent	Multi-agent
Technical completion	42/42	42/42	42/42
Correct terminal decisions	42/42	42/42	42/42
Correct permitted recommendations	24	24	24
Correct required holds	18	18	18
Failed workflows	0	0	0
Inadmissible raw proposals / proposals	0/54	0/54	0/54
Reported model cost (USD)	0.0000	2.1314	5.6348

The revised comparison has complete usage accounting for all 42 records per workflow. The multi-agent workflow incurred 2.64 times the single agent’s reported model cost, including the costs of required holds (table 11). Both used the shared arithmetic-receipt interface. The live run required zero correction attempts.

Independent reconstruction checked all 252 narrative records and their event chronology. The initial run contained 24 deterministic, 14 single-agent and seven multi-agent issued outputs; the revised run contained 24 per workflow. Across these 117 outputs, the checker found zero violations of the supplied financial, evidence, provenance and approval requirements. These results assess the retained emissions under the authored requirements; domain validity remains with the case-policy review.

Figure 8 contrasts the initial conformance gaps with the revised 42/42 reference outcomes for every workflow. Revised model expenditures across all 42 assignments are USD 0, USD 2.13 and USD 5.63 for deterministic, single-agent and multi-agent execution. On these two observed dimensions, the deterministic workflow strictly dominates both model arrangements and the single agent strictly dominates the multi-agent arrangement: conformance is equal and model expenditure is lower. This descriptive comparison includes correct recommendations and required holds and retains the experimental grammar, supplied policy and resource-accounting boundary.



14 authored cases, three repetitions; revised interface evaluated on the same cases.
Every revised calculation had one feasible candidate. Cost covers provider model charges.

Figure 8: *Narrative conformance and revised model expenditure.*

Note. Panel (a) shows observed conformance in the initial and revised narrative comparisons; panel (b) shows revised model expenditure with complete accounting. Sources: retained narrative comparisons summarised in tables 10 and 11.

7 Discussion

The simulations draw attention to choices made before an agent begins to optimise. In the constructed cases, an omitted liability led to a liquidity shortfall, omitted manager terms breached supervision capacity, and a mistranslated

weight changed the selected service. Duplicating a client facet also changed the objective's behaviour. These examples suggest a practical reason to review disclosure and mandate translation alongside execution. Proposition 1 gives the corresponding information limit: selecting a client-optimal outcome across states with disjoint optima requires complete decision inputs that distinguish those states.

Nested delegation offers a way to organise that review around the manager's dual role. The manager agrees the mandate with the client and translates it into instructions for the system. In the proposed framework, mandate review addresses the objectives, scales, service terms and evidence criteria; execution review checks their application and current authorization. Making those responsibilities explicit could help reviewers trace a questionable recommendation to the choices or information on which it depends.

In the joint objective, benchmark floors preserve both declared evaluations relative to a reference service, and positive weights select a Pareto-efficient pair among the alternatives satisfying the constraints and floors. Law and mandate determine those alternatives before scalarisation. This ordering gives a place in the specification to MiFID II Article 24's client-interest duties and Article 54(1)'s retained responsibility for automated suitability assessment (European Securities and Markets Authority, *n.d.*, paras. 1 and 10; European Commission, 2017, art. 54(1)). The choice of utility scales, benchmarks and commercial criteria still calls for an assessment against those duties. Even the interpretation of λ depends on the scales: rescaling manager value while holding the weight fixed changes the effective trade-off.

The manager-floor example makes the stakes of choosing a reference service concrete. Preserving manager value at $1/2$ selects (C, ω_1) with client value $3/4$, while excluding (B, ω_2) with client value 1 and manager value $1/4$. The weight concession is zero; the benchmark and total concessions are $1/4$. Under the proposed review rule in Equation (14), that concession calls for policy review and supporting client-benefit evidence. A necessary delivery restriction instead enters admissibility and changes the comparator. Client-confirmed preferences, agreed service terms and domain assessment provide grounds for accepting the policy or revising the reference service.

The economic simulation illustrates a similar tension using constructed fees, personnel costs, supervision hours and service-quality values. In six of eight states, the declared weight selected a higher service tier than the client-best pair, including with a client weight close to one. The floors were nonbinding in those cases, so the reported client amounts equal the total concessions. In the 2022 simulation, expected benchmark preservation also coexisted with a lower subsequent client outcome and higher manager net income. That euro comparison applies the same household-payment adjustment and service valuation to both services; it measures a different quantity from the risk-adjusted decision criterion. The example suggests that review should extend to the forecasts and valuations supporting the mandate.

Service classification matters as well. In the Swiss illustration, the common hold policy adds an institutional restriction to the mapped FinSA information duty; the EU necessary-information gate follows Article 54(8) (Swiss Confederation, 2018, arts. 11–14; European Commission, 2017, art. 54(8)). A candidate can remain financially feasible while a change in the required-information classification changes permission to proceed. Recording the basis of each restriction could make such decisions easier to explain. The same applies to the distinction between evidence sufficiency and permission to collect it: the specification records both the required information and its permitted acquisition routes (European Securities and Markets Authority, 2016; European Parliament and Council of the European Union, 2016, arts. 5–6; Swiss Confederation, 2020, arts. 6–7).

Clarification raises further questions about how the mandate handles uncertainty. Statewise Pareto filtering removes the demonstrated dominated alternatives, including the zero-product case. Candidate-specific assessment allows an eligible maximiser to proceed while another awaits clarification. The answer-partition test detects the supplied partial-answer consequence, and the client-criterion diagnostic detects a gain from changing candidates even at equal euro losses. These properties hold for the declared states, facets, answer partitions and thresholds. The partial-answer example is particularly relevant to mandate design: the componentwise robust comparison can reverse as evidence arrives, even when every fully resolved state initially selects the same candidate. Whether that behaviour suits the client depends on the chosen treatment of uncertainty.

There are several reasons to seek further information. Required gaps govern permission to proceed; optional answers may improve selection, reveal dominance or restore feasibility. The restricted selector evaluates declared conditional gains under a fixed policy. An answer that would restore an empty feasible set still needs human resolution, while a discretionary question about the full joint objective would need assessment against both parties' evaluations and the workflow. The simulation assigns costs and source owners to questions. Choosing a sequence of questions would require a further account of their costs and possible answers, with probabilities for an expected-value treatment. Calibration on representative mandates could examine consequential omissions, question burden, unnecessary holds and review effort before evaluating a chosen policy on held-out cases.

After a recommendation, evidence changes may call for a fresh assessment. The withdrawal and replacement checks distinguish lost support from surviving independent evidence; the currency example shows how a changed valuation can leave an unchanged obligation underfunded. Such checks could contribute to the testing, monitoring and allocation of responsibility addressed by FINMA (Swiss Financial Market Supervisory Authority FINMA, 2024, secs. 2.1 and 2.4). They also give engineers concrete events that should trigger reassessment or renewed approval.

The narrative runs illustrate how implementation choices can affect execution. Rejected inputs and repeated calculations exhausted budgets in the initial interface, and the adapter blocked a proposal carrying unsupported attribution and missing calculation support. Typed extraction, clearer instructions and shared arithmetic receipts subsequently reached all reference outcomes on the same development cases. This experience suggests that the interface deserves attention alongside the decision rule. Because those changes were introduced together and the correction mechanism remained unused, their individual contributions remain open.

In the instruction comparison, professional prose, the explicit joint objective and nested-delegation instructions reached the same prescribed decisions; professional prose also matched nested delegation on question count. All three received the declared trade-off, both floors, source identities and common enforcement. Their agreement illustrates several ways of communicating the same specification. The representation interventions changed its content and produced departures from the original mandate. Recovering missing inputs used a supplied field schema and authoritative source routes. Recognising an omitted field, choosing an appropriate scale or justifying a reference service would call on professional judgment at an earlier stage.

Workflow design raises a related question about the value of additional roles. In the revised narrative comparison, deterministic execution and the single agent achieved the same conformance as the multi-agent arrangement at lower measured model cost. The two facet-labelled calls and coordinator each received the complete policy and common numerical support, and every revised calculator event had one feasible candidate. The observed agreement therefore concerns extraction and evidence-aware sequencing; the analytical constructions examine selection among competing feasible alternatives. Giving specialists complementary information, with budget-matched repeated single-agent checking as a comparison, could help assess when decomposition adds value.

The service workflow ω and the agent architecture would both need consideration in an applied system. The simulated service tiers differ in fees, quality and supervisory hours, while the instruction treatments hold architecture fixed. The narrative comparison varies architecture and records model expenditure. Relating these choices would require measurements of professional effort, infrastructure cost and delivered service quality. Additional roles may be worthwhile when their information or expertise improves the service enough to justify those costs. The proposed outcome functions offer a place to express and examine that judgment.

8 Conclusion

This paper formulates the task of a delegated AI system as joint maximisation under binding constraints within a nested client–manager–AI relationship. The formulation represents client and manager interests through separate outcome functions, with explicit constraints, reference floors and trade-off weights. It connects responsibility for the mandate to its translation into an executable decision specification. Concession accounting makes the client consequences of the floors and weights available for review, while the distinction between disclosure, translation and execution helps locate possible departures from the intended mandate.

The analytical constructions and simulation illustrate how this framework can guide calculations, clarification questions and approval controls. Within the simulated settings, changes in represented information, constraints and policy choices affected the available or selected decisions. These observations suggest that making such choices explicit could support more structured mandate design and review. The underlying proposal is independent of a particular model or agent architecture and may provide a useful basis for comparable implementations. Professional specification studies and field evaluations could assess whether, and under which conditions, this approach improves decision quality, oversight and client outcomes.

9 Limitations and Future Research

The simulation instantiates the joint objective with constructed mandates, preferences, service terms and growth scenarios. Public exchange-rate and yield observations anchor selected market inputs; they supply current retrieval vintages of historical observations. Client evaluation scales, reference service, quality values and managerial commercial

interests require domain validation. The source-linked legal analysis supplies the context for institution-specific mandate and processing review. Strategic disclosure, hidden effort and incentive-compatible implementation require behavioural models and observations of how the parties respond to the specification.

Information degradation depends on the disclosure and specification maps and available supplementary evidence. The indistinguishability result assumes identical complete decision inputs and fixed policy primitives. Finite state enumeration, omitted-state discovery and evidence sufficiency bound the procedure's applicability. Statewise Pareto filtering and the client-criterion trigger depend on declared cardinal utilities. Flat Nash products can conceal an answer-dependent dominance change even at zero clarification thresholds; undominated issuance concerns the current support. Optional questions that restore an empty feasible set in the restricted client selector require human evidence resolution. The clarification policy examines one-input partitions and possible gains. Joint inputs can conceal a selection gain even when every individual consequence is zero; question probabilities, costs and compound partitions remain open. Applying the legal mapping to an actual mandate requires current provider authorization or exemption, client category, territorial facts and processing arrangements, alongside production identity and storage controls.

The structured pilot spans six families sharing three motivating source clusters; the narratives span two source-linked families with controlled grammar. Repetitions measure within-case consistency. All workflows receive authored policies and numerical support. The revised narrative interface was developed from failures on the same cases used for its evaluation. All 114 revised calculator events returned one feasible candidate, comprising 57 R90 and 57 R180 results. The deterministic parser is tailored to the document grammar. Initial model costs cover 37 matched records per workflow because six failures lack final usage totals; revised costs cover all 42. The deterministic policy checks establish properties of the Pareto and partition procedure, while the historical model comparisons retain their original policy scope.

The joint-objective comparison has four base mandates, eight dependent conditions per mandate and one repetition per instruction treatment. Evaluation mandates were reserved from prompt development and share their four templates and weights with the development mandates; an interrupted preliminary execution exposed a numerical grading defect, corrected by standardising monetary inputs to cents before the reported comparison. The final treatment instructions were retained. Source records present explicit numerical fields, source owners and bounded question routes, and the common approval layer rejects unsupported or inadmissible proposals. The observed parity concerns that shared environment. Broader language variation, independently authored mandates and budget-matched controls would assess transfer; related conditions should remain grouped in allocation and uncertainty estimates.

A prospective reviewer comparison would assess the framework against a competent governance checklist covering the same controls. Reviewers would diagnose matched defects in mandate representation, requirements translation, premise support and current approval. Decision accuracy, responsibility assignment, unnecessary holds and review effort would remain separate outcomes. Practitioner responses and independent domain assessment of case labels would establish whether organizing the controls through nested delegation improves professional specification and review. Field evaluation would examine client outcomes after oversight and implementation costs.

A prospective screening extension would specify disclosure menus and test truth-telling and participation conditions using client utilities, disclosure costs, outside options and outcomes of possible deviations (Mirrlees, 1976; Rothschild & Stiglitz, 1976). Controlled behavioural studies would examine *generative moral hazard*, the hypothesis that omitted checks can remain concealed behind a plausible recommendation, and its dependence on effort and generation incentives (Holmstrom, 1979; Kalai et al., 2025).

References

- Ante, L. (2026). Specification cost: Relocating the binding constraint in delegation to AI agents with externally supplied objectives. *Technology in Society*, 88, Article 103503. <https://doi.org/10.1016/j.techsoc.2026.103503>
- Bernheim, B. D., & Whinston, M. D. (1986). Common agency. *Econometrica*, 54(4), 923–942. <https://doi.org/10.2307/1912844>
- Bloem, R., Könighofer, B., Könighofer, R., & Wang, C. (2015). Shield synthesis: Runtime enforcement for reactive systems. <https://doi.org/10.48550/arXiv.1501.02573>
- Bokrantz, R., & Fredriksson, A. (2017). Necessary and sufficient conditions for Pareto efficiency in robust multiobjective optimization. *European Journal of Operational Research*, 262(2), 682–692. <https://doi.org/10.1016/j.ejor.2017.04.012>

- Bundesanstalt für Finanzdienstleistungsaufsicht. (2021). Big data and artificial intelligence: Principles for the use of algorithms in decision-making processes. Retrieved October 1, 2026, from https://www.bafin.de/SharedDocs/Downloads/EN/Aufsichtsrecht/dl_Prinzipienpapier_BDAI_en.html
- Cemri, M., Pan, M. Z., Yang, S., Agrawal, L. A., Chopra, B., Tiwari, R., Keutzer, K., Parameswaran, A., Klein, D., Ramchandran, K., Zaharia, M., Gonzalez, J. E., & Stoica, I. (2025). Why do multi-agent LLM systems fail? <https://doi.org/10.48550/arXiv.2503.13657>
- Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50(6), 1431–1451. <https://doi.org/10.2307/1913390>
- D’Acunto, F., Prabhala, N., & Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, 32(5), 1983–2020. <https://doi.org/10.1093/rfs/hhz014>
- European Central Bank. (n.d.-a). *Euro area yield curves* [AAA-rated euro-area government zero-coupon spot yields; one-, five- and ten-year maturities]. Retrieved October 1, 2026, from https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_area_yield_curves/html/index.en.html
- European Central Bank. (n.d.-b). *Swiss franc/euro ECB reference exchange rate* [Series EXR.D.CHF.EUR.SP00.A; daily observations, 2014–2025]. Retrieved September 29, 2026, from <https://data-api.ecb.europa.eu/service/data/EXR/D.CHF.EUR.SP00.A?startPeriod=2014-01-01&endPeriod=2025-12-31&format=csvdata>
- European Commission. (2017). Commission Delegated Regulation (EU) 2017/565 [Consolidated text of 2 August 2022; Articles 54–56 and 91]. Retrieved September 30, 2026, from https://eur-lex.europa.eu/eli/reg_del/2017/565/2022-08-02/eng
- European Parliament and Council of the European Union. (2014). Directive 2014/65/EU on markets in financial instruments [Consolidation dated 6 June 2026; Article 25]. Retrieved September 30, 2026, from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02014L0065-20260606>
- European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Retrieved May 3, 2026, from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679>
- European Securities and Markets Authority. (n.d.). *Article 24 general principles and information to clients* [MiFID II Interactive Single Rulebook; paragraphs 1 and 10]. Retrieved October 1, 2026, from <https://www.esma.europa.eu/publications-and-data/interactive-single-rulebook/mifid-ii/article-24-general-principles-and>
- European Securities and Markets Authority. (2016, October 10). *Suitability and appropriateness: Client unwilling to fully disclose financial information* [ESMA Q&A 1762; ESMA35-43-349, chapter 2, question 7]. Retrieved September 29, 2026, from <https://www.esma.europa.eu/publications-data/questions-answers/1762>
- European Securities and Markets Authority. (2021, January 13). Reminder to firms of the MiFID II rules on reverse solicitation in the context of the recent end of the UK transition period [Public statement ESMA35-43-2509; MiFID II Articles 39 and 42]. Retrieved September 30, 2026, from https://www.esma.europa.eu/sites/default/files/library/esma35-43-2509_statement_on_reverse_solicitation.pdf
- European Securities and Markets Authority. (2022). Guidelines on certain aspects of the MiFID II suitability requirements. Retrieved October 1, 2026, from https://www.esma.europa.eu/sites/default/files/2023-04/ESMA35-43-3172_Guidelines_on_certain_aspects_of_the_MiFID_II_suitability_requirements.pdf
- Federal Data Protection and Information Commissioner. (n.d.). *Duty to provide information*. Retrieved September 29, 2026, from <https://www.edoeb.admin.ch/en/duty-to-provide-information>
- Federal Data Protection and Information Commissioner. (2025, September 24). *AI and data protection*. Retrieved September 29, 2026, from <https://www.edoeb.admin.ch/en/ai-and-data-protection>
- Federal Republic of Germany. (1961). Banking Act, section 32. Retrieved September 30, 2026, from https://www.gesetze-im-internet.de/kredwg/___32.html
- Federal Republic of Germany. (1994). Securities Trading Act, sections 63 and 64 [Consolidated act records last amendment by Article 11 of the Act of 26 May 2026, BGBl. 2026 I No. 156]. Retrieved September 30, 2026, from https://www.gesetze-im-internet.de/wphg/___64.html
- Federal Republic of Germany. (2021). Investment Firm Act, section 15. Retrieved September 30, 2026, from https://www.gesetze-im-internet.de/wpig/___15.html
- Gadioli, A. V. A., Azevedo, P., Frizera-Neto, A., De Souza, A. F., Oliveira-Santos, T., & Badue, C. (2026). PeterAI: An AI multi-agent framework for personalized investment consultancy. In *Neural information processing* (pp. 125–139, Vol. 2754). Springer Nature Singapore. https://doi.org/10.1007/978-981-95-4091-4_9
- Gong, H., Samawi, M., & Medda, F. (2026). Authority–inference separation in agentic finance: First-line control, blockchain enforcement, and replayable assurance [Preprint, version 1]. Retrieved October 1, 2026, from <https://arxiv.org/html/2608.30519v1>
- Grossman, S. J., & Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4), 691–719. <https://doi.org/10.1086/261404>

- Hart, O., & Moore, J. (1990). Property rights and the nature of the firm. *Journal of Political Economy*, 98(6), 1119–1158. <https://doi.org/10.1086/261729>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Holmstrom, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 10(1), 74–91. <https://doi.org/10.2307/3003320>
- Howard, R. A. (1966). Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1), 22–26. <https://doi.org/10.1109/TSSC.1966.300074>
- Inderst, R., & Ottaviani, M. (2009). Misselling through agents. *American Economic Review*, 99(3), 883–908. <https://doi.org/10.1257/aer.99.3.883>
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305–360. [https://doi.org/10.1016/0304-405X\(76\)90026-X](https://doi.org/10.1016/0304-405X(76)90026-X)
- Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). Why language models hallucinate. <https://doi.org/10.48550/arXiv.2509.04664>
- Kuhn, L., Gal, Y., & Farquhar, S. (2023). CLAM: Selective clarification for ambiguous questions with generative language models. *Workshop on Challenges in Deployable Generative AI at the International Conference on Machine Learning*. Retrieved September 29, 2026, from <https://arxiv.org/abs/2212.07769>
- Kurz, W. (2026). Multi-jurisdictional legal identity assurance for capability gating: A design-science proposal for tiered, reusable identity assurance of natural, juridical, and machine entities. *Swissi AI Journal*. Retrieved September 29, 2026, from <https://journal.swissi-ai.institute/en/doi/5kdnql4rsq27>
- Milsom, R. (2025). Benchmarking LLM agents for wealth-management workflows. Retrieved September 29, 2026, from <https://arxiv.org/abs/2512.02230>
- Mirrlees, J. A. (1976). The optimal structure of incentives and authority within an organization. *The Bell Journal of Economics*, 7(1), 105–131. <https://doi.org/10.2307/3003192>
- Nash, J. F. (1950). The bargaining problem. *Econometrica*, 18(2), 155–162. <https://doi.org/10.2307/1907266>
- NeuFin. (n.d.). *NeuFin Decision Assurance Envelope* [Specification v1.0.0]. Retrieved September 29, 2026, from <https://www.neufin.ai/developers/decision-assurance-envelope>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Racioppi, G. (2026). Mandato: Protocol-level enforcement of digitally signed mandates on AI agent actions with cryptographically chained audit trails [Preprint, version 1; architectural comparison based on the abstract]. Retrieved October 1, 2026, from <https://arxiv.org/abs/2608.14074>
- Republic of Austria. (2018a). Securities Supervision Act 2018, section 56 [Section 56 version effective 9 June 2022; RIS date selection 30 September 2026; document NOR40244333]. Retrieved September 30, 2026, from <https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20009943&Paragraf=56&FassungVom=2026-09-30>
- Republic of Austria. (2018b). Securities Supervision Act 2018, sections 21, 23 and 24 [RIS date selection 30 September 2026; section 21 effective 3 January 2018, section 23 effective 25 July 2025 and section 24 effective 1 February 2023]. Retrieved September 30, 2026, from <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20009943&FassungVom=2026-09-30>
- Rothschild, M., & Stiglitz, J. E. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4), 629–649. <https://doi.org/10.2307/1885326>
- Sreenivas, N. K., & Larson, K. (2026, August 18). Contracting for LLM delegation: Moral hazard in technology and effort choice [Preprint, version 1]. <https://doi.org/10.48550/arXiv.2608.18232>
- Srivastava, V. (2026, September). *AI Advice Evidence Benchmark: A framework for scoring investor context in AI-generated wealth decisions*. NeuFin. Retrieved September 29, 2026, from <https://www.neufin.ai/research/ai-advice-evidence-benchmark>
- Steuer, R. E., & Na, P. (2003). Multiple criteria decision making combined with finance: A categorized bibliographic study. *European Journal of Operational Research*, 150(3), 496–515. [https://doi.org/10.1016/S0377-2217\(02\)00774-9](https://doi.org/10.1016/S0377-2217(02)00774-9)
- Stoughton, N. M. (1993). Moral hazard and the portfolio management problem. *The Journal of Finance*, 48(5), 2009–2028. <https://doi.org/10.1111/j.1540-6261.1993.tb05140.x>
- Swiss Banking Ombudsman. (2019). *Poor performance of an investment fund* [Case 2019/06]. Retrieved September 29, 2026, from <https://bankingombudsman.ch/en/poor-performance-of-an-investment-fund/>
- Swiss Banking Ombudsman. (2020). *Damage due to faulty investment advice* [Case 2020/24]. Retrieved September 29, 2026, from <https://bankingombudsman.ch/en/damage-due-to-faulty-investment-advice/>

- Swiss Banking Ombudsman. (2023). *Unauthorized purchases of fund units as part of an investment advisory mandate* [Case 2023/14]. Retrieved September 29, 2026, from <https://bankingombudsman.ch/en/unauthorized-purchases-of-fund-units-as-part-of-an-investment-advisory-mandate/>
- Swiss Confederation. (1911). Federal Act on the Amendment of the Swiss Civil Code (Part Five: The Code of Obligations) [SR 220; Articles 398 and 400]. Retrieved October 1, 2026, from https://www.fedlex.admin.ch/eli/cc/27/317_321_377/en
- Swiss Confederation. (2018). Federal Act on Financial Services [German authoritative text, status 1 March 2024; Articles 10–14, 15, 20 and 25–26]. Retrieved September 30, 2026, from <https://www.fedlex.admin.ch/filestore/fedlex.data.admin.ch/eli/cc/2019/758/20240301/de/pdf-a/fedlex-data-admin-ch-eli-cc-2019-758-20240301-de-pdf-a.pdf>
- Swiss Confederation. (2020). Federal Act on Data Protection [German authoritative text, status 1 September 2023; Articles 6, 21 and 30–31]. Retrieved September 29, 2026, from <https://www.fedlex.admin.ch/filestore/fedlex.data.admin.ch/eli/cc/2022/491/20230901/de/pdf-a/fedlex-data-admin-ch-eli-cc-2022-491-20230901-de-pdf-a.pdf>
- Swiss Federal Council. (2019). Financial Services Ordinance [German authoritative text, status 1 January 2022; Article 2 territorial scope]. Retrieved September 30, 2026, from <https://www.fedlex.admin.ch/filestore/fedlex.data.admin.ch/eli/cc/2019/759/20220101/de/pdf-a/fedlex-data-admin-ch-eli-cc-2019-759-20220101-de-pdf-a.pdf>
- Swiss Financial Market Supervisory Authority. (2024, October 31). Circular 2025/2 on rules of conduct under FinSA and FinSO [Effective 1 January 2025; German authoritative circular; marginal numbers 2, 4 and 13–14]. Retrieved September 30, 2026, from https://www.finma.ch/de/~media/finma/dokumente/dokumentencenter/myfinma/rundschreiben/finma-rs-2025-02-20241122.pdf?hash=B0B038372817C2443D4066505B340890&sc_lang=de
- Swiss Financial Market Supervisory Authority FINMA. (2024). FINMA Guidance 08/2024: Governance and risk management when using artificial intelligence. Retrieved September 30, 2026, from <https://www.finma.ch/en/~media/finma/dokumente/dokumentencenter/myfinma/4dokumentation/finma-aufsichtsmittelungen/20241218-finma-aufsichtsmittelung-08-2024.pdf>
- Tirole, J. (1986). Hierarchies and bureaucracies: On the role of collusion in organizations. *The Journal of Law, Economics, and Organization*, 2(2), 181–214. <https://doi.org/10.1093/oxfordjournals.jleo.a036907>
- Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, 25(1), 77–89. <https://doi.org/10.1057/ejis.2014.36>
- WealthSchema. (2026, September). *FiduciaryBench* [Version 2026.9]. Retrieved September 29, 2026, from <https://www.wealthschema.com/benchmark>
- Xie, Y., Zhu, C., Zhang, X., Zhu, T., Ye, D., Qi, M., Chen, H., & Zhou, W. (2026). From spark to fire: Modeling and mitigating error cascades in LLM-based multi-agent collaboration. <https://doi.org/10.48550/arXiv.2603.04474>
- Zhang, Y., & Zhang, T. (2025). How generative AI reshapes principal-agent relationships: Mitigating adverse selection and moral hazard. *IET Conference Proceedings, 2025*(22), 49–67. <https://doi.org/10.1049/icp.2025.2956>
- Zhuang, S., & Hadfield-Menell, D. (2020). Consequences of misaligned AI. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (pp. 15763–15773, Vol. 33). Curran Associates, Inc. Retrieved October 1, 2026, from https://proceedings.neurips.cc/paper_files/paper/2020/file/b607ba543ad05417b8507ee86c54fcb7-Paper.pdf

Declarations

CITE THIS ARTICLE (BIBTEX)

```
@article{saijguex5c23zfm5,
  author = {Kurz, W. and Magg, R. and Kollberg, F. and Stricker, W. and Marx, S. and Reinhardt, F. and Dedić, V.},
  title = {Multi-Agent AI as a Nested Principal-Agent Problem in Private Wealth Management: Mandate Representation
    and Evidence Control in Switzerland, Germany and Austria},
  journal = {Swissi AI Journal},
  year = {2026},
  volume = {2026},
  note = {Article SAIJ-guex5c23zfm5},
  issn = {3043-1921},
  doi = {10.5281/zenodo.23095194},
  url = {https://journal.swissi-ai.institute/doi/guex5c23zfm5}
}
```

PAPER INFORMATION

AFFILIATIONS	¹ Swissi Institute for AI, office@swissi-ai.institute ; ² Hochschule für Wirtschaft und Umwelt Nürtingen-Geislingen ³ Faculty of Information Technology and Engineering (FITI), Belgrade
CORRESPONDING AUTHOR	Walter Kurz, me@walterkurz.com
ARTICLE TYPE	Research Article
SWISSI ADDRESS	Sw2:academic01:obj:p1:guex5c23zfm5ipnsrcfym5ceo7vk2wtki74vqeszxxrlmfgy32q:08038f13

PAPER DECLARATIONS

AUTHOR CONTRIBUTIONS	All authors contributed equally to this work.
FUNDING	This research received no external funding.
CONFLICTS OF INTEREST	The authors declare no conflicts of interest.
DATA AND CODE AVAILABILITY	The authors retain the source code, constructed mandates, market inputs, recorded responses, experimental configurations and verification instructions. Access to these research records is restricted. Market inputs include source URLs and content hashes.
USE OF AI TOOLS	GPT-6.1 Astra and Claude Opus 5.5 were used for the simulation.

JOURNAL INFORMATION

PUBLISHER	Swissi AI Journal
EDITORIAL DATES	Submitted 06/2026; Revised 09/2026; Published 10/2026
LICENCE	CC BY 4.0
OPEN ACCESS	https://journal.swissi-ai.institute/authors#open-access
CREDIT TAXONOMY	https://journal.swissi-ai.institute/authors#credit-taxonomy
JOURNAL CONTACT	https://journal.swissi-ai.institute journal@swissi-ai.institute