Q-OPS: WHEN SELECTIVE QUANTUM MACHINE LEARNING FAILS—A COST-AWARE QUANTUM-KERNEL BANDIT AUDIT FOR HUMAN–AI ALLOCATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Advanced computation creates economic value only when its decision contribution exceeds its resource cost. Quantum-assisted AI makes this urgent: benchmarks ask whether a quantum routine changes an output, but organizations must decide before invocation whether it is worth acquiring under partial feedback and hard constraints. We ask whether pre-invocation context can guide a learner to decline or acquire Quantum Approximate Optimization Algorithm (QAOA) evidence. In synthetic human–AI task-mode allocation, a classical procedure fixes allocations and routes inadmissible cases to review, while quantum samples remain evidence rather than decisions. We introduce Q-OPS (Quantum-Optional Policy Selection), an audit that casts decline, depth-one or depth-two QAOA as priced contextual-bandit actions and compares linear, radial-basis-function (RBF), and six-qubit fidelity-kernel representations using ridge upper-confidence bounds. Across 320 held-out instances (221 eligible), the warm-started depth-one distribution improves normalized score over uniform support sampling by 0.331; an oracle invokes on 55.7% and earns 4.245 cumulative expected proxy reward. Yet every learner invokes on 100% and earns -2.398, -0.686, or -0.477; aggregate losses include mandatory depth-two calls costing 0.66 although their largest archived gain is 0.494. The evidence separates reference-relative output quality, pre-invocation predictability, proxy net evidence value, and downstream decision value, which is not measured. Practitioners and policymakers should demand a decline option, resource price, pre-purchase information boundary, matched alternatives, and non-tradeable constraints. Limited to exact simulation on enumerable synthetic supports, Q-OPS motivates warm-state and classical-reference controls, fresh-instance mechanism tests, and participatory studies of costs and stakeholder-defined outcomes.

1 INTRODUCTION

Organizations allocate labor, tasks, and computational attention before acting; expensive analysis may help difficult cases and waste resources on routine ones. A policy that uses information available before a call to choose decline or invocation performs *selective computation*. Its economic logic separates four claims. *Reference-relative output quality* asks whether purchased output improves a declared comparison distribution; *Pre-invocation predictability*, whether context identifies valuable calls; *proxy net evidence-acquisition value*, whether that gain exceeds price; and *Downstream decision value*, whether used evidence improves an actual decision. Metareasoning treats computations as resource-limited actions (Russell & Wefald, 1991), while contextual bandits learn values from chosen-action feedback (Li et al., 2010).

Quantum-assisted learning makes the distinction visible. QAOA and alternating-operator methods can concentrate probability on high-scoring states (Farhi et al., 2014; Hadfield et al., 2019); quantum kernels encode classical inputs through state-overlap similarities (Havlíček et al., 2019; Schuld & Killoran, 2019); and quantum policies appear in reinforcement learning (Jerbi et al., 2021). Yet a structured output or expressive representation does not establish that a learner can identify, before paying, when the evidence improves a decision enough to justify its cost. Calls for matched classical references and explicit scope make that gap consequential (Bowles et al., 2024; Koch et al., 2026).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

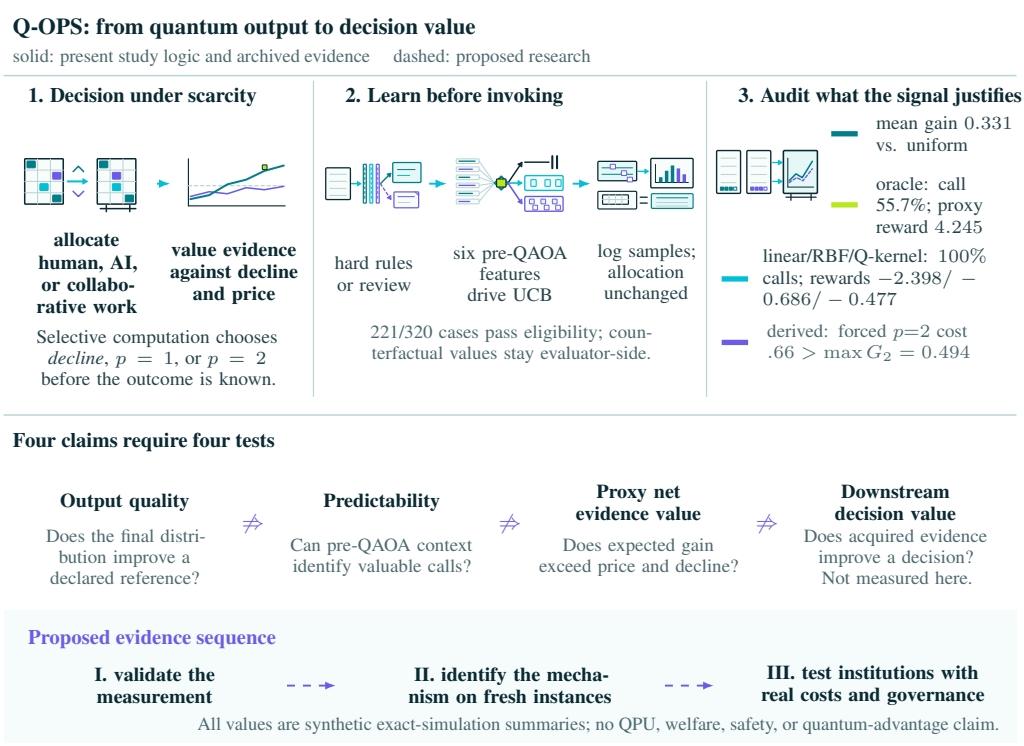


Figure 1: **Semantic argument and evidence boundary.** Solid cyan links the audited sequence; dashed violet marks proposed research. At $\lambda = .33$, depth-one gain is relative to uniform, rewards are cumulative proxy net evidence values over 221 eligible cases, and learned-policy means average 30 permutations. The depth-two inequality is derived. The four constructs do not imply one another; downstream value is unmeasured. Technical scenes are conceptual: samples never alter allocation, and price is normalized rather than observed cost. No QPU or quantum-advantage claim is made.

Kernelized upper-confidence-bound (UCB) contextual bandits connect nonlinear prediction with exploration (Valko et al., 2013; Zenati et al., 2022). DAK-UCB routes prompts among costly generative models (Jafari & Farnia, 2026); quantum-bandit studies examine kernel complexity or quantum-neural-tangent-kernel UCB (Huang et al., 2026b;a). Missing is priced quantum evidence competing with decline behind requirements reward cannot buy away—*non-tradeable constraints*—while quantum computation separately generates evidence and represents context. We introduce neither a new UCB algorithm nor a regret theorem.

We ask: *Can pre-invocation information reveal when simulated QAOA evidence is worth its proxy cost, and what does failure teach about the bridge from output to decision value?* In synthetic human–AI *task-mode allocation*, a classical procedure assigns each task a human-only, AI-only, or collaborative mode; rules send inadmissible cases to review; and an eligible bandit chooses decline or QAOA depth one or two. We call this audit *Quantum-Optional Policy Selection* (Q-OPS). It is not an employee study: quantum samples never become assignments.

Across 320 held-out instances, final warm-started QAOA distributions improve a weak uniform reference and hindsight values admit selective use, yet every learner invokes on every eligible case and the quantum kernel offers no practical gain over matched radial-basis-function (RBF) features. Aggregate losses also include forced sampling of depth two, which is dominated at the primary price. The overview in Figure 1 previews three contributions: (i) a protocol fixing information, decline, price, and feasibility; (ii) an audit retaining favorable output shifts, oracle headroom, policy failure, and this confound; and (iii) a four-part framework followed by measurement, mechanism, and institutional tests.

2 DECISION PROBLEM AND LEARNING PROTOCOL

2.1 A CONTROLLED ALLOCATION ENVIRONMENT

A *task-mode allocation* specifies how each synthetic task is performed. Let $i \in \mathcal{I}$ index tasks, $r_i \in \mathcal{R}$ their synthetic regions, and $y_i \in \mathcal{M}_i \subseteq \{\mathrm{H}, \mathrm{A}, \mathrm{HA}\}$ their execution modes: human-only, AI-only, or human–AI collaboration. Human-mandatory tasks exclude A. One generated instance, rather than a person or organization, is the unit of analysis.

The evaluator ranks candidate allocations under one declared set of priorities. It combines scenario-adjusted value, direct cost, capacity overflow, cluster complementarity and handoff terms, regional AI-exposure dispersion, review overflow, disallowed modes, and service and capacity chance-constraint excesses. With all terms defined in Appendix B,

$$\begin{aligned} S(y) = \bar{V}(y) - C(y) - 4\bar{O}(y) + B_{\mathrm{cl}}(y) - P_{\mathrm{cl}}(y) \\ - 2\ \mathrm{sd}_{r\in\mathcal{R}}\{q_r(y)\} - 9[R(y) - B_R]_+ - 600H(y) \\ - 150[\rho_s(y) - .20]_+ - 45[\rho_c(y) - .20]_+. \end{aligned} \tag{1}$$

Higher $S(y)$ means that an allocation ranks better under these stipulated weights. The score has no natural monetary unit and is not a social-welfare function. In particular, $\mathrm{sd}_{r\in\mathcal{R}}\{q_r(y)\}$ measures only regional AI-exposure dispersion, not fairness.

Some requirements are *non-tradeable*: rather than accepting a score penalty in exchange for violating them, the system withholds the learning decision and requests review. We call their deterministic conjunction the *eligibility contract*. Feasibility is evaluated separately:

$$\mathcal{F}(y) = 1 \iff y_i \in \mathcal{M}_i\ \forall i, \quad R(y) \le B_R, \quad \rho_s(y) \le .20, \quad \rho_c(y) \le .20. \tag{2}$$

A coordinate-greedy search and fixed cluster destroy–repair routine obtain the classical incumbent y^C (Ropke & Pisinger, 2006). Learning proceeds only if $\mathcal{F}(y^C) = 1$ and a four-task retained core has a strictly feasible state; otherwise the system requests review. Review is an external rule, not a learned action, and no outcome is imputed (Appendix B).

The QAOA statevector and uniform reference share the strictly feasible support $\mathcal{Z}$. Hence Eq. (1)'s disallowed-mode, review-overflow, service-failure, and capacity-failure penalties vanish on every $z \in \mathcal{Z}$: they shape the incumbent and gate, but not within-support gain. The remaining terms define the energy landscape.

2.2 SELECTIVE COMPUTATION AND DECISION CHRONOLOGY

For each eligible instance: (1) generate and repair; (2) apply the gate; (3) compute six pre-QAOA features; (4) choose a_0 (decline), a_1 ($p = 1$), or a_2 ($p = 2$); (5) sample only the chosen depth; and (6) update only that learner. "Pre-invocation" marks time, not low cost: four features enumerate a retained support of at most 81 states. Counterfactual outcomes serve fixed-policy and oracle audits only; the learner sees selected-action feedback. Decline has known reward zero.

Simulated QAOA supplies the optional evidence. A four-task core contains at most $3^4 = 81$ assignments; its strictly feasible states form $\mathcal{Z}$. An alternating-operator statevector uses $E(z) = -S(z)$, one-mode-move adjacency, and a warm state biased toward y^C. For $p \in \{1, 2\}$,

$$|\psi_p(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \prod_{\ell=1}^{p} e^{-i\beta_\ell H_M} e^{-i\gamma_\ell H_C} |\psi_0\rangle. \tag{3}$$

Both depths receive 64 parameter candidates and 128 output samples. Counts are matched, not parameter-space coverage. This dense CPU statevector study has no device topology, noise, or QPU execution (Appendix C).

2.3 FROM OUTPUT QUALITY TO PROXY NET EVIDENCE VALUE

For invocation action a_p, *reference-relative output quality* is the unitless, range-normalized expected gain over the declared uniform reference:

$$G_p(x) = \frac{\mathbb{E}_{z\sim p_p}[S(z)] - \mathbb{E}_{z\sim U(\mathcal{Z})}[S(z)]}{\max_{z\in\mathcal{Z}} S(z) - \min_{z\in\mathcal{Z}} S(z)}. \tag{4}$$

A positive G_p favors the final warm-started distribution over uniform on the same support; it does not isolate QAOA layers from the warm state. Exact G_p supports audit. With price λ, costs are $C(a_0) = 0$, $C(a_1) = 1$, and $C(a_2) = 2$. The 128 selected-action samples give

$$\widehat{G}_{p,t} = \frac{128^{-1} \sum_{j=1}^{128} S(z_{t,j}) - \mathbb{E}_{z \sim U(\mathcal{Z}_t)}[S(z)]}{\max_{z \in \mathcal{Z}_t} S(z) - \min_{z \in \mathcal{Z}_t} S(z)}, \qquad r_t = \widehat{G}_{p,t} - \lambda C(a_p), \tag{5}$$

with $r_t = 0$ under decline. Thus audit values and learning feedback remain distinct.

Total proxy resource is $\sum_t C(a_t)$. Expected *proxy net evidence-acquisition value* is

$$\mu(x, a_p) = G_p(x) - \lambda C(a_p), \qquad \mu(x, a_0) = 0, \tag{6}$$

Positive μ favors invocation; negative μ favors known-zero decline. Cost is normalized, not measured time, money, energy, or hardware expense. Because evidence never changes y^C, this reward is not downstream decision value or organizational benefit.

3 Learning Strategies and Diagnostic Logic

3.1 Matched context representations

The context $x \in \mathbb{R}^6$ contains relative classical gain, greedy–repair disagreement, score skewness, range-normalized dispersion, retained-graph density, and warm-state distance. Pilot statistics standardize it; evaluation statistics do not. Four features require the enumerable retained support, so x is pre-QAOA but neither low-cost nor scalable. The experiment audits selective learning after this preprocessing, not computational savings.

A *kernel* represents context similarity without assuming reward is linear in the original features. Let $|\chi(x)\rangle$ denote the six-qubit state encoding context x. A feature map in the style of Havlíček et al. (2019) yields the CPU-exact fidelity kernel

$$k_Q(x, x') = |\langle \chi(x) | \chi(x') \rangle|^2. \tag{7}$$

The radial-basis-function (RBF) and quantum maps share 24 pilot landmarks and rank 20; the linear map uses the six standardized features. Feedback, tuning, and updates are identical. Only the kernels are rank-matched: linear and Nyström vectors may differ in norm, and no model has an intercept. Thus RBF–quantum is the tighter comparison (Appendix C); classical coverage is not exhaustive.

3.2 Learning conditional value with an explicit decline action

An upper-confidence-bound (UCB) policy adds an uncertainty bonus to estimated reward. For invocation arm a, let $\varphi_a(x)$ be its linear or Nyström vector. Disjoint ridge-UCB maintains $A_a = I + \sum_{s:a_s=a} \varphi_a(x_s)\varphi_a(x_s)^\top$, $b_a = \sum_{s:a_s=a} r_s \varphi_a(x_s)$, and $\hat{\theta}_a = A_a^{-1} b_a$, with score

$$U_t(a \mid x) = \hat{\theta}_a^\top \varphi_a(x) + \alpha \sqrt{\varphi_a(x)^\top A_a^{-1} \varphi_a(x)}, \qquad a \in \{a_1, a_2\}. \tag{8}$$

After two forced calls per quantum arm, the policy declines iff $\max_a U_t(a \mid x) \leq 0$; otherwise it selects the maximizing quantum arm. Decline therefore has a defined zero score and no parameters. Only the selected arm updates, α is pilot-chosen, and evaluation is online adaptation rather than a frozen test policy.

Outside initialization, Eq. (8) invokes exactly when a quantum score exceeds zero. Because cost enters reward while only quantum arms receive a nonnegative bonus, finite-horizon optimism may cause over-calling; this is not a theorem, because bonuses can shrink. The archive lacks stepwise mean–bonus traces, leaving exploration, value misspecification, and representation insufficiency unresolved.

3.3 Comparators and interpretation

Comparators are decline-all, always-$p = 1$, the pilot-selected fixed action (decline at the primary price), and a per-instance oracle, alongside three learned representations. The oracle uses unavailable

Table 1: Study contract fixed before held-out evaluation.

Component	Specification
Pilot / evaluation	96 / 320 synthetic instances; disjoint seeds
Instance grid	$n \in \{12, 18\}$; stress $d \in \{0, 1, 2, 3\}$; 24 scenarios
Eligibility	Composite-feasible classical incumbent and non-empty strict core
Actions	Decline; QAOA $p = 1$; QAOA $p = 2$
Feedback	Selected action only; decline has known value zero
Matched count	64 parameter candidates and 128 draws per depth; unequal search dimension
Context	Six pre-QAOA diagnostics; four require retained-support enumeration
Feature maps	Linear; RBF Nyström; six-qubit fidelity Nyström
Bandit	Disjoint ridge-UCB; 24 landmarks; rank 20; two warm calls/quantum arm
Primary price	$\lambda = .33$; sensitivity $\{0, .10, .20, .33, .40\}$
Evaluation	30 permutations of one held-out set; compact finite-set summaries

counterfactual values $\arg\max_a \mu(x, a)$; positive oracle value shows headroom, not predictability from x. Appendix G records missing summaries.

Outcomes are expected cumulative proxy reward, call rate, resource use, and *pseudo-regret*: expected loss against the counterfactual best action. Sampled feedback drives updates, but the compact archive retains only the paired quantum-minus-RBF realized-reward difference. A different kernel or trace is not practical gain without better reward and selectivity; neither establishes downstream value.

4 Evidence and Evaluation Design

4.1 Bespoke synthetic generator and pilot revision

All observations are generated by code: there are no operational records, workers, customers, human participants, live APIs, or language-model outputs. The generator crosses task counts $\{12, 18\}$, stress levels $\{0, 1, 2, 3\}$, and seeds. Stress raises sensitivity, review probability, workflow penalties, and correlated disruption probability. Each instance has three synthetic regions, three-task clusters, and 24 disruption scenarios.

The 96-instance pilot uses seeds 1001–1012 and selects $\alpha = 0.25$ for every representation from $\{0.25, 0.50, 1.00, 1.50\}$ across 12 orders. Its first design let sampled quantum states compete with the repaired incumbent; improvement was zero in all 64 eligible cases. A plausible neighborhood-overlap explanation is not causally identified. The study record says that before held-out evaluation, the target changed to Eq. (4), post-quantum context variables were removed, and $\lambda = 0.33$ was fixed near the pilot median depth-one gain, although the compact archive cannot independently time-stamp the sequence. Retaining this failed pilot distinguishes allocation improvement from evidence acquisition.

Evaluation uses 320 disjoint instances (seeds 2001–2040), 30 online-order permutations, and $\lambda \in \{0, .10, .20, .33, .40\}$. Reordering probes sequence sensitivity on the same instances; it does not create 30 independent datasets. No price is selected for favorable evaluation performance.

4.2 Metrics, denominators, and uncertainty

For an eligible sequence of length T, expected cumulative proxy net evidence reward is $\sum_{t=1}^{T} \mu(x_t, a_t)$. Pseudo-regret is

$$\mathcal{R}_T = \sum_{t=1}^{T} \left[\max_{a \in \{a_0, a_1, a_2\}} \mu(x_t, a) - \mu(x_t, a_t) \right]. \tag{9}$$

It uses expected action values rather than a noise-selected realized winner. Ineligible instances are reported in the whole-system eligibility denominator but excluded from bandit reward, invocation, and regret; their unobserved human-review outcomes are not imputed. The archived learning curves advance across all 320 arrivals, while bandit metrics update only at the 221 eligible arrivals. Fixed policies are order-invariant. Learned-policy means and percentile order ranges summarize path

sensitivity across the 30 permutations. Instance-level evidence quantities use paired summaries and 10,000-resample percentile intervals over the 221 eligible cases. These are descriptive finite-set resampling summaries, not population confidence intervals. Seed identifiers recur across the crossed task-count and stress settings, and the compact artifact cannot support a seed-blocked reanalysis.

4.3 Information and reproducibility audit

Pilot and evaluation seeds are disjoint. Feature standardization, landmarks, rank, exploration coefficient, and the pilot-selected fixed action are determined without evaluation outcomes. The context is constructed before QAOA and the online learner observes only selected-action feedback. Because several context variables require exact retained-support enumeration, this chronology does not imply that the selector is computationally cheaper than the call it audits. Counterfactual expected outcomes are restricted to evaluator-side fixed-policy, sensitivity, and oracle summaries. These checks protect the evaluation chronology; they do not establish external validity.

Appendix G records the CPU-only environment, maps claims to the intended anonymous artifact, and distinguishes a smoke test from a full rerun.

5 Results

5.1 RQ1: warm-started outputs improve a weak reference within a shrinking eligibility set

Of 320 held-out instances, 221 (69.1%) pass the eligibility contract; 99 are escalated because the classical incumbent violates composite feasibility. No ineligible instance enters the bandit. Eligibility falls from 97.5–100% at stress zero to 22.5–30% at stress three (Figure 2c). This is deterministic rule behavior, not learned safety.

Within eligible cores, the final warm-started distributions at both QAOA depths shift probability toward higher-scoring states relative to uniform sampling. Mean expected normalized gain is 0.331 for $p = 1$ (descriptive finite-set resampling interval 0.325–0.337) and 0.295 for $p = 2$ (0.287–0.304). Exact optimum probability averages 0.0157 under uniform, 0.2117 under $p = 1$, and 0.1540 under $p = 2$. With equal counts of 64 parameter candidates, $p = 1$ has higher expected gain on 174/221 instances and $p = 2$ on 47. Equal counts do not equalize search density across the two- and four-dimensional parameter spaces. These comparisons neither isolate QAOA layers from the score-biased warm state nor benchmark a competitive classical sampler. They establish only a final-distribution shift on classically enumerable supports, not incremental layer value, intrinsic depth potential, computational complexity, hardware performance, or quantum advantage.

The $p = 1$ samples increase mean score by 16.32 points and reduce direct cost by 3.41 relative to uniform support means. They also raise collaboration frequency by 5.65 percentage points (95% interval 5.25–6.05), reduce regional AI-exposure dispersion by 0.75 points (0.46–1.05), and raise capacity-violation frequency by 0.48 points (0.33–0.65). The composite score therefore hides small operational trade-offs; none is a measured organizational outcome.

5.2 RQ2: learned policies never decline, while aggregate losses include forced depth two

At $\lambda = 0.33$, the hindsight oracle invokes $p = 1$ on 123/221 eligible instances (55.7%), declines the remainder, and obtains 4.245 cumulative expected proxy net evidence reward (Table 2). This establishes a selective expected-value regime under privileged counterfactual values, not predictability from the observed context. Linear, RBF, and quantum-kernel UCB instead invoke a quantum action on every eligible instance and obtain -2.398, -0.686, and -0.477. Decline-all obtains zero, and decline is also the pilot-selected best fixed action. On a per-eligible-case scale, the oracle and quantum-kernel values are 0.0192 and -0.0022; the cumulative totals should not be read as large operational effects or downstream decision improvements.

The aggregate reward sign does not isolate failure to decline. At the primary price, depth two costs $2\lambda = 0.66$, while its maximum archived gain is 0.494. It is therefore dominated by decline on every recorded eligible case, yet each learner is forced to call both quantum arms twice. Because all 221

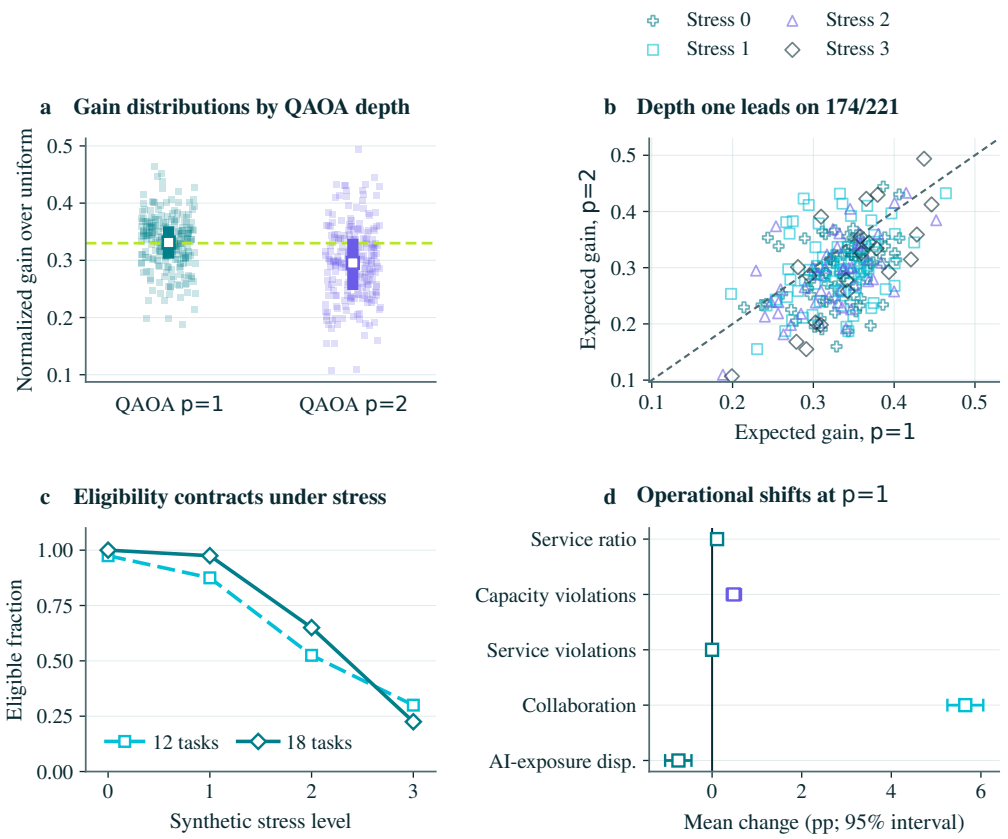


Figure 2: **Descriptive held-out evidence diagnostics.** (a) Expected reference-relative gain of the final warm-started QAOA distributions under equal parameter-candidate counts; the dashed line is the primary depth-one proxy price. (b) Paired depth comparison, colored by synthetic stress. (c) The deterministic eligibility contract becomes restrictive under stress; each cell contains 40 instances. (d) Mean operational-metric differences between 128 depth-one samples and exact uniform means on the same strictly feasible supports. Bars are descriptive 95% finite-set instance-bootstrap intervals, not population confidence intervals; seed identifiers recur across grid cells. The denominator is 221 eligible cases. Uniform is a weak reference, and AI-exposure dispersion is not a fairness claim.

eligible cases receive a quantum call, resource accounting implies mean depth-two call counts of 7.47 (linear), 2.57 (RBF), and 2.07 (quantum kernel). The quantum-kernel policy therefore almost stops choosing depth two after initialization but still substitutes depth one for decline on the remaining cases. The compact archive cannot identify the exact reward contribution of forced versus later calls. The 100% invocation rate establishes failure to recover the oracle's decline pattern; the negative total cannot be attributed solely to autonomous post-initialization behavior.

The quantum kernel makes slightly different sequence-dependent choices from RBF. Across matched orders, its cumulative pseudo-regret is lower by 0.208 on average, with percentile order range $[-0.482, 0]$; realized reward is higher by 0.205, range $[0, 0.489]$. The endpoints at zero, shared 100% invocation rate, and negative proxy net evidence rewards provide no practically meaningful quantum-kernel advantage in this study.

5.3 RQ3: THE AUDIT NARROWS, BUT DOES NOT IDENTIFY, THE FAILURE MECHANISM

The cost scan separates three regimes. At $\lambda \leq .20$, nearly every eligible $p = 1$ action has positive net value, so always-$p = 1$ nearly matches the oracle and learning to decline is unnecessary. At $\lambda = .33$,

Table 2: Held-out policy results at $\lambda = .33$ (221 eligible of 320). "Reward" is proxy net evidence-acquisition value, not downstream decision value. Means are over 30 order permutations; fixed policies are order-invariant. Reward and regret are cumulative; dividing by 221 gives an eligible-case mean. Every learned policy includes two forced calls to each quantum arm.

Policy	Cum. proxy reward	Pseudo-regret	Call rate	Resource
Decline all	0.000	4.245	0.0%	0.0
Always $p = 1$	0.257	3.988	100%	221.0
Linear UCB	−2.398	6.643	100%	228.5
RBF-kernel UCB	−0.686	4.930	100%	223.6
Quantum-kernel UCB	−0.477	4.722	100%	223.1
Oracle (hindsight)	4.245	0.000	55.7%	123.0

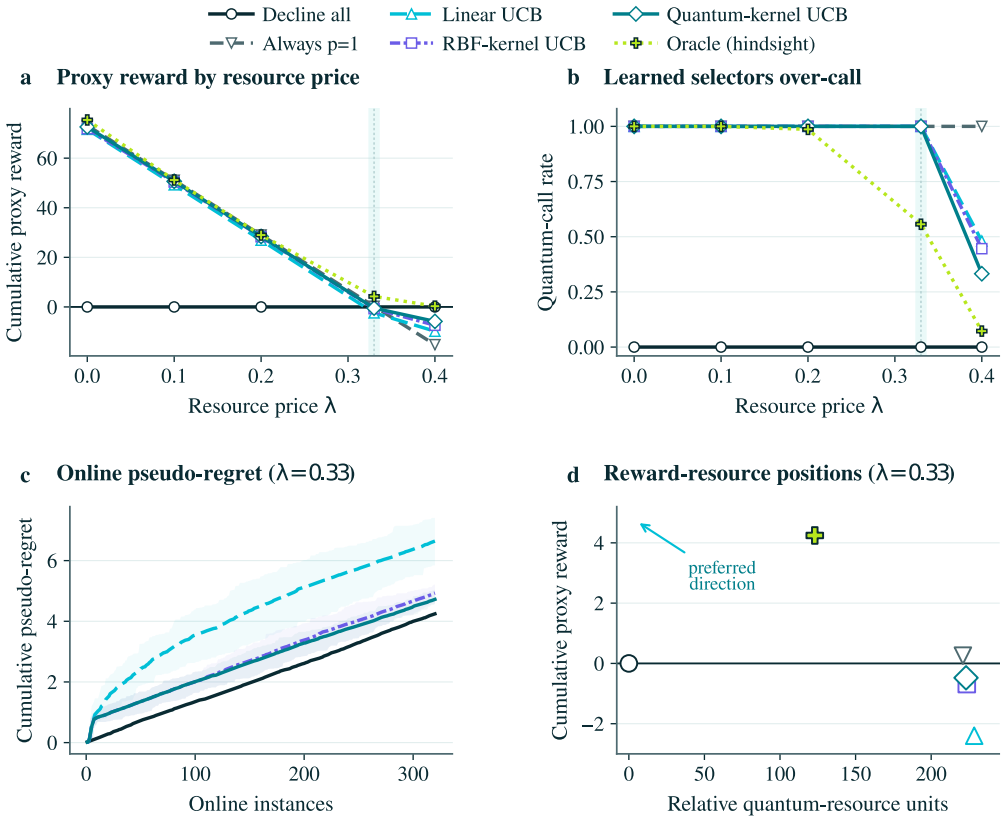


Figure 3: **Descriptive selective-policy audit.** (a) Expected proxy net evidence reward and (b) invocation rate over the declared price grid; the narrow band marks $\lambda = 0.33$. Coincident traces are genuine. (c) Cumulative pseudo-regret at the primary price; the x-axis advances over all 320 arrivals, but only 221 eligible arrivals update the bandit. Ribbons give the 2.5–97.5 percentile range across 30 permutations of the same held-out set, not independent replications. (d) Expected reward and normalized proxy-resource use at the primary price. Oracle values use counterfactual expected outcomes and are unavailable online; no panel measures downstream allocation value.

selectivity matters but the learners do not recover it. At $\lambda = .40$, quantum-kernel UCB lowers its invocation rate to 33.2%, compared with 44.6% for RBF and 48.0% for linear UCB, but still earns -5.822 while decline earns zero. Lower call rate alone is therefore insufficient; selected calls must have gain above price.

The results rule out absence of a final-distribution shift relative to uniform and absence of a privileged invoke/decline regime. They also show that replacing the matched RBF map with the tested fidelity map does not bridge the gap. Equation (8) and the all-call traces remain consistent with weak conditional-value prediction, a positive uncertainty bonus over the available horizon, or both; they do not prove any one mechanism. Forced depth-two exploration further prevents a clean attribution of the negative reward sign. The current records contain neither mean–bonus score traces nor a fresh out-of-sample action-value predictability experiment, so choosing among these explanations would exceed the evidence.

6 Discussion and Future Research

The central finding is a broken value chain, not a representation victory. Warm-started depth-one output clears a weak uniform reference, and counterfactual values reveal selective headroom; the policies nevertheless fail to decline. Q-OPS separates output quality, pre-invocation predictability, proxy net evidence value, and downstream decision value. This synthetic experiment addresses the first three but does not measure the fourth.

DAK-UCB contributes a routing objective, mixture extension, regret analysis, and positive applications (Jafari & Farnia, 2026); Q-OPS audits a zero-valued decline option and supplies neither a new UCB method nor comparable theory. It cannot identify the kernel-complexity mechanism in Huang et al. (2026b). The RBF and quantum kernels are rank-matched, but linear and kernel features are not norm-matched and lack intercepts; the enumerable cores also fall outside application-scale quantum-advantage tests (Koch et al., 2026). The contribution is evaluation logic for warranted inference.

The mechanism remains unresolved. Forced depth-two exploration is unprofitable at the primary price; later all-call behavior could reflect context, scaling, representation, optimism, initialization, or horizon. A nonnegative quantum-arm bonus against zero-valued decline makes UCB optimism plausible, not proved. Posterior sampling (Agrawal & Goyal, 2013), a baseline-conservative bandit (Kazerouni et al., 2017), exploitation-only selection, mean–bonus traces, and positive/null controls should be tested on fresh instances, not tuned on this held-out set.

For practitioners and policymakers, the implication is procedural, not pro- or anti-quantum. An adoption case should state decline, pre-purchase information and computation, price, non-tradeable constraints, downstream use, and matched alternatives. Here four features require support enumeration; price is not measured cost, the evaluator is not welfare, and the gate is not learned safety. Q-OPS offers procurement questions, not organizational-benefit estimates.

These boundaries define three dependent levels (Appendix A): Level I adds warm-state and classical controls, cheap-context ablations, and a downstream consumer; Level II isolates mechanisms on fresh seed blocks; Level III tests measured costs and stakeholder-defined outcomes under participatory design and independent review. Appendix F gives one computational bridge. Until then, this synthetic exact-simulation evidence supports no general impossibility, QPU utility, scaling, safety, fairness, welfare, worker-benefit, or deployment claim.

7 Conclusion

Q-OPS shows why quantum evidence cannot be valued from output structure alone. Warm-started QAOA improves a uniform reference and oracle values reveal selective headroom, but the UCB policies never decline, losses include forced sampling of a depth-two arm dominated at the primary price, and the quantum kernel does not practically beat matched RBF. The four-part test—output quality, pre-invocation predictability, proxy net evidence value, and downstream decision value—makes this negative result constructive. Adoption should begin with decline, full resource accounting, matched alternatives, and non-tradeable constraints; research should progress from measurement to mechanism to institutions.

STATEMENTS EXCLUDED FROM THE MAIN-TEXT PAGE LIMIT

AI USE STATEMENT (REQUIRED)

Generative AI tools assisted literature discovery, research ideation and methodological critique, drafting and restructuring, language editing, code and figure debugging, and audit planning. They did not generate experimental observations, labels, synthetic instances, decisions made by the evaluated policies, or mathematical proofs, and no generative AI system is part of the method. The authors checked AI-assisted literature suggestions against primary sources, checked quantitative statements against the archived manuscript evidence and machine-readable plotted data, reviewed the revised source and compiled PDF, and take responsibility for the final text, claims, and artifacts.

ETHICS STATEMENT (RECOMMENDED)

This study uses only synthetic tasks and exact CPU simulation; it contains no participants, employee records, personal data, or deployed decision system. Human–AI task-mode allocation is an evaluation environment, not evidence about workers or organizations. The scalar evaluator is not social welfare, the exposure-dispersion term is not a complete fairness measure, and rule-based feasibility is not learned safety. Applying such a system to people could create displacement, surveillance, discrimination, and contestability risks. Any field study would require participatory objective design, privacy safeguards, independent governance review, and validation with affected communities.

REPRODUCIBILITY STATEMENT (RECOMMENDED)

Sections 2–4 specify the chronology, action set, information boundary, reward, comparators, seeds, and aggregation. Appendix B documents the evaluator and retained support, Appendix C the statevector and feature mappings, Appendix D the sensitivity values, and Appendix G the claim and artifact map. The anonymous companion artifact is intended to contain the generator, configurations, row-level records, online traces, tests, and full reproduction command at `https://anonymous.4open.science/r/qops-selective-qml-reproducibility-3426`.

REFERENCES

Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In Sanjoy Dasgupta and David McAllester (eds.), *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 127–135. PMLR, 2013. URL https://proceedings.mlr.press/v28/agrawal13.html.

Daniel S. Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. The complexity of decentralized control of Markov decision processes. *Mathematics of Operations Research*, 27(4): 819–840, 2002. doi: 10.1287/moor.27.4.819.297. URL https://pubsonline.informs.org/doi/10.1287/moor.27.4.819.297.

Joseph Bowles, Shahnawaz Ahmed, and Maria Schuld. Better than classical? the subtle art of benchmarking quantum machine learning models, 2024. URL https://arxiv.org/abs/2403.07059. arXiv:2403.07059; preprint.

Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm, 2014. URL https://arxiv.org/abs/1411.4028. arXiv:1411.4028.

Drew Fudenberg and Eric Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, 1986. doi: 10.2307/1911307. URL https://www.jstor.org/stable/1911307.

John Gardiner, Orlando Romero, Brendan Tivnan, Nicolò Dal Fabbro, and George J. Pappas. Learning to coordinate via quantum entanglement in multi-agent reinforcement learning, 2026. URL https://arxiv.org/abs/2602.08965. arXiv:2602.08965; preprint.

Stuart Hadfield, Zhihui Wang, Bryan O'Gorman, Eleanor G. Rieffel, Davide Venturelli, and Rupak Biswas. From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. *Algorithms*, 12(2):34, 2019. doi: 10.3390/a12020034. URL https://www.mdpi.com/1999-4893/12/2/34.

Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019. doi: 10.1038/s41586-019-0980-2. URL https://www.nature.com/articles/s41586-019-0980-2.

Yuqi Huang, Vincent Y. F. Tan, and Sharu Theresa Jose. Quantum-enhanced neural contextual bandit algorithms, 2026a. URL https://arxiv.org/abs/2601.02870. arXiv:2601.02870; preprint.

Yuqi Huang, Vincent Y. F. Tan, and Sharu Theresa Jose. Balancing expressivity and learnability in quantum kernel bandit optimization. In Emilija Perković and Daniel Malinsky (eds.), *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 2274–2313. PMLR, 2026b. URL https://proceedings.mlr.press/v337/huang26c.html.

Donya Jafari and Farzan Farnia. DAK-UCB: Diversity-aware prompt routing for LLMs and generative models. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=nnN2TKlS5C.

Sofiene Jerbi, Casper Gyurik, Simon C. Marshall, Hans J. Briegel, and Vedran Dunjko. Parametrized quantum policies for reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 34, pp. 28362–28375, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/eec96a7f788e88184c0e713456026f3f-Abstract.html.

Abbas Kazerouni, Mohammad Ghavamzadeh, Yasin Abbasi-Yadkori, and Benjamin Van Roy. Conservative contextual linear bandits. In *Advances in Neural Information Processing Systems*, volume 30, pp. 3910–3919, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/bdc4626aa1d1df8e14d80d345b2a442d-Abstract.html.

Thorsten Koch, David E. Bernal Neira, Ying Chen, Giorgio Cortiana, Daniel J. Egger, Raoul Heese, Narendra N. Hegade, Alejandro Gomez Cadavid, Rhea Huang, Toshinari Itoko, Thomas Kleinert, Pedro Maciel Xavier, Naeimeh Mohseni, Jhon A. Montanez-Barrera, Koji Nakano, Giacomo Nannicini, Corey O'Meara, Justin Pauckert, Manuel Proissl, Anurag Ramesh, Maximilian Schicker, Noriaki Shimada, Mitsuharu Takeori, Víctor Valls, David Van Bulck, Stefan Woerner, and Christa Zoufal. The quantum optimization benchmarking library. *Nature Computational Science*, 6 (6):653–671, 2026. doi: 10.1038/s43588-026-00991-1. URL https://www.nature.com/articles/s43588-026-00991-1.

David M. Kreps, Paul Milgrom, John Roberts, and Robert Wilson. Rational cooperation in the finitely repeated Prisoners' Dilemma. *Journal of Economic Theory*, 27(2):245–252, 1982. doi: 10.1016/0022-0531(82)90029-1. URL https://www.sciencedirect.com/science/article/pii/0022053182900291.

Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pp. 661–670, New York, NY, USA, 2010. Association for Computing Machinery. doi: 10.1145/1772690.1772758. URL https://dl.acm.org/doi/10.1145/1772690.1772758.

Stefan Ropke and David Pisinger. An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation Science*, 40(4):455–472, 2006. doi: 10.1287/trsc.1050.0135. URL https://pubsonline.informs.org/doi/10.1287/trsc.1050.0135.

Stuart J. Russell and Eric H. Wefald. Principles of metareasoning. *Artificial Intelligence*, 49(1–3):361–395, 1991. doi: 10.1016/0004-3702(91)90015-C. URL https://www.sciencedirect.com/science/article/pii/000437029190015C.

Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical Review Letters*, 122(4):040504, 2019. doi: 10.1103/PhysRevLett.122.040504. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.122.040504.

Michal Valko, Nathaniel Korda, Rémi Munos, Ilias Flaounas, and Nello Cristianini. Finite-time analysis of kernelised contextual bandits. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pp. 654–663. AUAI Press, 2013. URL https://dl.acm.org/doi/10.5555/3023638.3023705.

Houssam Zenati, Alberto Bietti, Eustache Diemert, Julien Mairal, Matthieu Martin, and Pierre Gaillard. Efficient kernelized UCB for contextual bandits. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 5689–5720. PMLR, 2022. URL https://proceedings.mlr.press/v151/zenati22a.html.

A Appendix guide and future-research program

Appendix B expands the synthetic evaluator and generator. Appendix C gives the statevector and kernel constructions. Appendix D records price-sensitivity and pilot quantities used in the main claims. Appendix E develops the mechanism-identification studies, Appendix F preserves a longer-horizon coordination hypothesis as a computational bridge, and Appendix G states the artifact boundary and claim ledger.

Table 3 converts the paper's consequential limits into a three-level program. The levels are dependencies rather than rhetorical stages: a downstream field study cannot validate an evidence measure whose source or decision relevance remains unidentified, and a mechanism claim should not be transported to organizations before it generalizes on fresh synthetic instances.

Table 3: Three-level future-research program. Every row is proposed work. Levels are sequenced dependencies, not evidence already collected.

Level and present limit	Testable question and evidence	Expected intellectual and practical value	Governance gate and sequence
I. Measurement validity. Final QAOA distributions are compared with uniform, confounded with a biased warm state, unused downstream, and predicted partly from exhaustively computed context.	Does gain survive a depth-zero warm-state control and competitive classical generator? Do cheap incumbent features retain predictive value, and do acquired samples change a pre-specified downstream choice? Use fresh synthetic instances, paired outcomes, seed-blocked uncertainty, and classical-compute accounting.	Separates layer contribution, reference sensitivity, feature acquisition cost, and downstream relevance. It prevents "structured output" or pre-QAOA timing from becoming an adoption or efficiency claim.	Preserve the classical allocation and eligibility contract; disclose negative controls and preprocessing cost. Complete before representation or field claims. Status: *proposed*.
II. Learning mechanism. Current summaries cannot separate information, feature scale, representation, exploration, initialization, horizon, or unequal QAOA search density.	Which intervention changes selective behavior? Cross positive/null controls, privileged predictability, UCB mean–bonus traces, exploitation-only, posterior-sampling, and baseline-conservative policies, norm/intercept-matched features, longer horizons, and dimension-scaled QAOA tuning on fresh seed blocks.	Identifies whether failure belongs to context, representation, the online rule, or optimization budget. It guides algorithm design without attributing every failure or success to quantum structure.	Pre-specify hypotheses, tuning, and diagnostics before new evaluation. Begin only after Level I supplies a defensible target. Status: *proposed*.
III. Institutional validity. Synthetic scores, proxy prices, and deterministic review rules do not measure welfare, organizational benefit, or legitimacy.	In a decision domain chosen with affected stakeholders, does selective evidence improve task outcomes after measured time, monetary, energy, review, error, and distributional costs? Use staged simulation, preregistered pilot, and only then a consented field design; units and sampling frame must follow the institution and affected community.	Connects value of computation to accountable adoption: who benefits, who bears cost and error, when review is available, and whether the decision remains contestable. It can inform procurement and policy only within the validated domain.	Require participatory objective design, privacy safeguards, labor-impact assessment, independent ethics/governance review, and stopping rules. Begin after Levels I–II. Status: *proposed*.

B Evaluator, support, and generator summary

For scenario ω, $\bar{V}(y)$ is mean priority-weighted delivered value, $C(y)$ direct mode cost, and $\bar{O}(y)$ mean regional capacity overflow. B_{cl} rewards clusters meeting their collaboration threshold; P_{cl} combines all-AI and handoff penalties. q_r is region-r AI-only exposure, R/B_R review load/capacity, H disallowed modes, and ρ_s, ρ_c service- and capacity-failure rates. Equation (1) is evaluated identically for every candidate source. On the strictly feasible quantum support $\mathcal{Z}$, $H(y) = 0$, $[R(y) - B_R]_+ = 0$, $[\rho_s(y) - .20]_+ = 0$, and $[\rho_c(y) - .20]_+ = 0$. Thus these four penalties enforce eligibility and support construction but vanish in every within-$\mathcal{Z}$ gain calculation.

The synthetic generator uses three regions and three-task clusters. Human, AI, and collaborative modes receive costs and nominal success rates; collaboration is generated above the better single mode, while sensitive and human-mandatory tasks restrict modes. Common and regional shocks perturb demand, human capacity, AI availability, and service. Stress raises sensitivity, review probability, workflow penalties, and disruption. The service target is 0.70, both scenario tolerances are 0.20, and capacity scale is 0.92; `skill` is unused metadata.

The classical incumbent is obtained by coordinate-greedy improvement and 18 destroy–repair iterations with two restarts. The retained core contains four task positions. Eligibility requires composite feasibility, a nonempty

strictly feasible retained support, and a positive score range for Eq. (4). All 221 archived eligible summaries have finite gains; the compact artifact does not retain the support ranges or the runner's zero-range convention. It also does not expose the full algebra for the cluster terms, every generator distribution, or the deterministic tie-breaking rule used to select the four-task core. Those details require the companion runner; the present section is therefore an implementation summary rather than an independently executable specification. The evaluator may enumerate counterfactual actions for auditing, but this information is not part of the online context or feedback.

C STATEVECTOR AND QUANTUM FEATURE-MAP DETAILS

The retained graph joins states differing in one task mode and adds nearest-state links for isolated vertices. If z_w is nearest to the classical incumbent, initial amplitude is proportional to $\exp[-0.35 d_H(z, z_w)]$; cost energies are centered and standardized. At each depth, 64 seeded Latin-hypercube candidates cover $[0, 2\pi]^p \times [0, \pi]^p$ and minimize expected energy, breaking ties by near-optimal mass and optimum probability; the selected circuit supplies 128 draws. Equal candidate counts provide a reproducible budget comparison, not equal search density: they cover two parameters at $p = 1$ and four at $p = 2$. Matching candidate and draw counts also does not match wall-clock cost or establish intrinsic depth potential.

For the quantum context map, pilot-standardized values are clipped to $[-3, 3]$ and mapped to $[0, \pi]$. Two repetitions apply all-qubit Hadamards, single-qubit feature phases, and cyclic pair phases $\frac{1}{2}(\pi - x_j)(\pi - x_{j+1})Z_j Z_{j+1}$, with $x_7 = x_1$; exact 64-amplitude states $|\chi(x)\rangle$ yield the fidelity Gram matrix. The compact paper source does not expose the runner's exact single-qubit phase-gate convention, so independent circuit reconstruction requires the companion artifact. The RBF bandwidth is the pilot median Euclidean distance between standardized contexts (2.6251). Both kernels use the same 24 farthest pilot landmarks and rank 20. The linear vector remains six-dimensional; the implementations use the same ridge coefficient and α but neither equalize feature norms across linear and kernel models nor add an intercept.

The two quantum actions have separate ridge-UCB parameters and update only when chosen. The known decline arm is not estimated. The two forced warm calls per quantum arm are included in invocation, reward, resource, and regret totals. Thus an invocation rate below 100% can still coexist with negative proxy net evidence reward when selected actions are poorly targeted.

D ADDITIONAL QUANTITATIVE AUDIT

At $\lambda = 0$, always-$p = 1$ obtains expected proxy net evidence reward 73.187 and regret 2.222, versus 72.591 and 2.818 for quantum-kernel UCB. At $\lambda = .10$ and .20, always-$p = 1$ is within 0.104 and 0.015 of the oracle. At $\lambda = .40$, the oracle calls 7.2% and earns 0.307; linear, RBF, and quantum-kernel UCB call 48.0%, 44.6%, and 33.2% and earn -9.797, -7.400, and -5.822. These values instantiate the three regimes discussed in Section 5.

At the primary price, depth two costs 0.66 per call, while the largest archived depth-two gain is 0.493790. It consequently has negative expected proxy net evidence value on every eligible case. The mean recorded resources imply 7.47, 2.57, and 2.07 depth-two calls for linear, RBF, and quantum-kernel UCB, respectively, because 221 depth-one calls would consume exactly 221 resource units. Each policy includes two forced depth-two calls. This arithmetic identifies a protocol-level contribution to negative reward, not the exact post-initialization reward. In particular, subtracting an average forced-call penalty would not recover the realized values of the specific warm-start cases.

Pilot-only exploration-coefficient search selected 0.25 for all three representations. At the primary price, mean pilot reward was -2.261 for linear UCB and -0.907 for both kernel policies, so the pilot-selected best fixed action was decline. This is distinct from the evaluation-only observation that always-$p = 1$ has small positive cumulative reward 0.257.

The evidence audit uses two denominators. Eligibility is 221/320 for the whole system. Reference-relative output quality, action choice, reward, resource, and regret use only those 221 eligible cases. The 30 order permutations reuse these cases and therefore describe path sensitivity rather than 30 population replications.

E LEVEL II: PROPOSED MECHANISM DIAGNOSTICS

The following studies are designs, not results. A positive control would generate contexts in which one observed coordinate deterministically identifies whether $G_1(x) > \lambda$; a null control would set both invocation values below zero. Both should retain the same selected-action update code and known-zero decline arm. Failing either control would point to implementation, initialization, or horizon rather than quantum representation.

A privileged information-sufficiency audit would fit action-value predictors using counterfactual expected outcomes for all three arms, with development and evaluation instances separated prospectively. Because it uses feedback unavailable online, its scores would diagnose whether the six features carry value-relevant information but could not be reported as bandit performance. A paired context ablation should compare the two incumbent-derived diagnostics with the four support-enumeration diagnostics, reporting both predictive value and classical computation. Representation tests should normalize feature vectors to a common scale, add or remove intercepts symmetrically, and tune ridge and exploration parameters separately on development data. This would distinguish information content and representation fit from online exploration.

Finally, the UCB score should be logged as estimated mean plus uncertainty bonus for every arm and step. Evaluation would compare the original exploratory policy with a fixed exploitation-only rule after pilot training, posterior sampling, and a baseline-conservative rule. A persistent positive mean would implicate value misspecification; a positive bonus concentrated on the quantum arms would implicate uncertainty or horizon; failures of the positive/null controls would expose action or reward plumbing. A separate QAOA study should scale the parameter-search budget with dimension or add local refinement before interpreting $p = 2$ versus $p = 1$. Any choice motivated by the current held-out outcomes requires a fresh seed block before supporting a performance claim.

F Level III bridge: Prospective dynamic testbed

One future direction asks whether longer-horizon private information changes the value of selective computation. Repeated interaction, payoff-relevant private signals, behavior-updated trust, and a finite call budget could jointly create a coordination bottleneck not present in the one-shot benchmark (Fudenberg & Maskin, 1986; Kreps et al., 1982; Bernstein et al., 2002). For a continuation policy π, let $V^{\pi}_{t:T}$ denote gross stakeholder-defined value from time t onward, including future service and harms but excluding the current invocation fee. A history-conditioned value of computation (Russell & Wefald, 1991) is then

$$\mathrm{VOC}^{\pi}_{t}(a) = \mathbb{E}[V^{\pi}_{t:T} \mid h_t, a_t = a] - \mathbb{E}[V^{\pi}_{t:T} \mid h_t, a_t = a_0] - \lambda C(a), \tag{10}$$

where h_t is the observed history and the same continuation policy governs later choices in both expectations. A recent preprint studies entanglement-assisted coordination in a sequential decentralized setting (Gardiner et al., 2026); it does not imply an advantage for this allocation or invocation problem.

Figure 4 is deliberately prospective. Its mechanism would be supported only if a selectively invoked quantum resource improves held-out net value in the joint high-continuation, informative-signal regime while maintaining calibrated abstention elsewhere and surviving matched message-passing, belief-state, and recurrent classical baselines. Failure under stronger communication, unstable trust, or full resource pricing would remain informative.

G Artifact structure and claim ledger

The intended anonymous companion artifact specifies three distinct commands: unit tests, figure generation, and full reproduction. The full command is designed to regenerate 96 pilot and 320 evaluation records, select hyperparameters from pilot data, evaluate every policy and price over 30 orders, rebuild figures, and compare row-level records within 10^{-12}. A SHA-256 manifest is intended to detect silent changes. The compact paper repository additionally checks headline values against plotted CSVs but contains neither the full row-level experiment runner nor the selected-action traces. Neither that check nor a compilation smoke test is equivalent to the full experiment rerun. The bound paper artifact omits full summaries for always-$p = 2$ and uniform-random comparators, so the main paper does not count them as evaluated.

The archived study reports CPU-only Linux execution with 9 Intel Xeon Platinum 8370C vCPUs, 21 GiB RAM, Python 3.12.13, and NumPy 2.3.5; producing the 320 evaluation rows took 571.21 seconds. This is an environment and runtime record, not hardware-performance evidence.

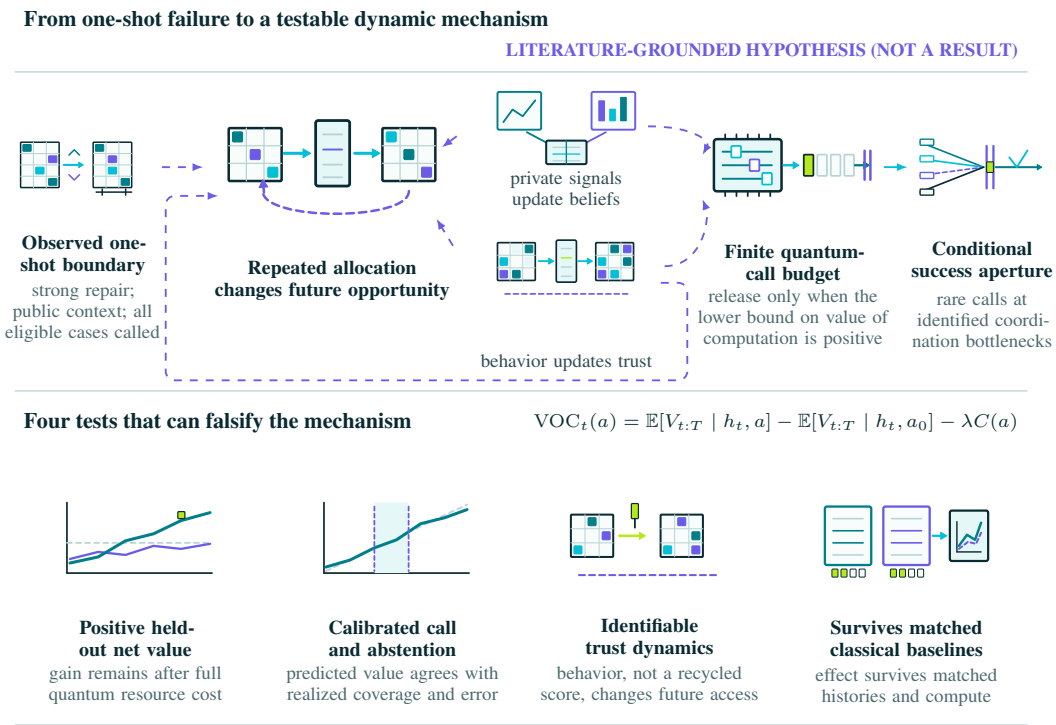


Figure 4: **Literature-grounded mechanism from the observed one-shot failure to a falsifiable dynamic success regime.** The central system makes the missing causal bridge explicit: private observations update beliefs, realized behavior updates a partner-trust matrix, those states affect future allocation, and a finite quantum accelerator token is released only when the lower bound on value of computation is positive. The right-hand aperture opens only when continuation value, informative signals, identifiable trust, and resource-aware selectivity jointly hold. The lower bench specifies four observable tests. All violet paths are hypotheses synthesized from cited literatures, not results of the present experiment.

Table 4: Claim ledger. "Supported" is restricted to the synthetic exact-statevector study.

Status	Claim
Supported	Final warm-started depth-one and depth-two QAOA distributions have higher mean feasible-support score than uniform on eligible held-out cases; the design does not isolate the layers from the warm state.
Supported	The expected-value oracle is selective at $\lambda = .33$, whereas the tested UCB policies invoke on every eligible case and have negative proxy net evidence reward under the implemented forced-exploration protocol.
Supported	The tested quantum kernel has no practical advantage over the matched RBF kernel in this experiment.
Unresolved	Whether all-call behavior is primarily due to context information, feature scaling, representation fit, exploration uncertainty, initialization, or horizon.
Not supported	Quantum speedup, QPU utility, noise robustness, scaling, or measured monetary, latency, or energy savings.
Not supported	Downstream allocation improvement, operational safety, worker benefit, welfare, fairness, or institutional value.