Contracts for Compute: Incentivizing Verifiable Effort in Tool-Using LLM Agents via Receipt-Contingent Payments

Liz Lemma Future Detective

January 16, 2026

Abstract

Modern agentic systems (2026) expose rich, verifiable traces of computation—tool-call logs, cryptographic receipts, and execution proofs—while internal reasoning effort remains hidden. We propose a principal–agent framework that treats LLM 'effort' (compute budget, search depth, tool planning intensity) as the hidden action and studies contracts that condition on both task outcomes and verifiable compute receipts. Building on principal–agent reinforcement learning with contracts and subgame-perfect equilibrium (SPE), we formalize receipt-contingent payments in a sequential environment and characterize when near-first-best effort is implementable with small subsidies under limited liability. Our main theoretical contribution is a tractable separability parameter—the total-variation gap between receipt distributions under high vs low effort—that governs the minimal expected subsidy needed for incentive compatibility. When receipts are sufficiently separable, a principal can implement efficient effort with bounded expected payments; when receipts are unavailable, outcome-only contracts suffer a lower bound on required subsidy proportional to the informativeness gap between outcomes and receipts. We further derive robust 'nudging' margins that stabilize best responses under model drift and approximation error, complementing the source paper's analysis of contract fragility and nudging. Finally, we outline an empirical protocol for tool-use benchmarks that treats receipts as contractible outcomes and evaluates learned principals under black-box validation against independently trained agents.

Table of Contents

2. 2. Static receipt-contracting model: binary effort, joint outcome–receipt distributions; limited liability; minimal implementation as an LP; closed-form optimal event-based contracts via likelihood ratios.

3. 3. Receipt separability and subsidy bounds: define Δ_s; derive upper bounds on minimal expected payment and conditions for near-first-best; discuss interpretability and how to estimate Δ_s from telemetry.

4. 4. Outcome-only benchmark and proxy gap: analyze contracts that can condition only on y; derive lower bounds and welfare gaps vs receipt-contracting; clarify when outcomes already suffice (observed-action-like regimes).

5. 5. Sequential extension (MDP): per-state contracting with receipts; SPE conditions; decomposition into per-state LPs with continuation values; total discounted subsidy bounds; when closed form carries over vs when numerical DP is needed.

6. 6. Robustness and nudging under drift: uncertainty sets for $P_s(r \mid e)$ and approximation error in learned Q-values; robust IC margins; explicit additional-subsidy bounds; diagnostic for when robustness is expensive (small Δ_s).

7. 7. Learning and implementation: model-free RL for principal policy and agent response; estimating receipt/outcome distributions; practical contract classes; two-phase training/black-box validation protocol; monitoring for compliance.

8. 8. Experiments design (not full writeup): tool-use benchmarks, compute receipts as contractible signals; compare (i) outcome-only, (ii) receipt-contingent, (iii) constant-proportion subsidy; stress tests under distribution shift; metrics (welfare, spend, compliance, robustness).

9. 9. Extensions and policy discussion: manipulation-resistant receipts, privacy constraints on telemetry, multi-principal competition, multi-agent coordination; implications for procurement and platform governance; limitations and future work.

1 Introduction and motivation: paying agents to think (2026)

In 2026, a large share of economically relevant work mediated by machine learning systems can be described, unromantically, as "paying agents to think." A user, firm, or platform (the principal) delegates a task to an LLM service (the agent), while caring about accuracy, safety, latency, and cost. Yet the principal typically cannot observe the agent's internal compute choices—how long it searched, whether it verified intermediate claims, how many tools it invoked, or whether it stopped early. This gap between what the principal values (high-quality outcomes) and what the agent privately controls (compute intensity and diligence) is a textbook moral hazard problem, but with a modern twist: the agent's "effort" is literally an allocation of expensive compute and tool-use budget, and its traces can sometimes be logged.

The core contracting challenge is familiar. If the principal could directly specify compute, it would simply require the agent to run the expensive checks and pay accordingly. In most deployments, however, the principal sees only a delivered answer (and perhaps some ex post evaluation) and must rely on incentive-compatible payment schemes. Outcome-only incentives are blunt: many tasks are noisy, long-horizon, or hard to score reliably, so the mapping from "worked hard" to "got a better outcome" can be weak. In such environments, attempts to elicit diligence via outcomes alone can require large subsidies, can induce gaming, or can push the agent toward risk-seeking behaviors that raise variance without improving reliability. These difficulties are amplified by limited liability constraints that are natural in practice (we rarely fine an API provider after a mistake; we at most withhold payment or pay bonuses).

What has changed since earlier discussions of moral hazard in computation is the increasing availability of *receipts*: verifiable telemetry about the agent's behavior that is not itself the final outcome. Examples include cryptographically signed tool-call logs, proofs of retrieval queries, attestations of sandboxed execution, counts of verifier invocations, or hardware-backed measurements of compute. Even when privacy or engineering constraints require these receipts to be coarse, they can be substantially more informative about diligence than the final task outcome. The simple economic point is that contracting power depends on *informativeness about hidden action*. Receipts create additional contractible signals that may separate high effort from low effort even when outcomes do not, thereby reducing the payment needed to implement diligence under limited liability.

This paper formalizes that intuition in a principal–agent model designed to speak to current LLM procurement and platform governance. We view "effort" as a hidden choice of compute/search depth/tool intensity, and we

allow the principal to condition payments on observable outcomes as well as verifiable receipts. The model is deliberately spare: we focus on when and how receipts change the feasible set of incentive schemes, and on what can be said in closed form about the cost of implementation. Our emphasis is not on the specific engineering of receipts, but on the economic object they induce: the divergence between receipt distributions under high and low compute. When this divergence is large, small payments can create large incentive wedges; when it is small, contracting becomes expensive and fragile.

A second motivation is methodological. Much of the modern literature on learning systems frames the problem as reinforcement learning (RL): an agent takes actions, receives rewards, and optimizes expected discounted return. But many real deployments are better described as *principal–agent RL*, where a principal designs the reward/payoff process to induce desired behavior from an optimizing agent whose internal choices are not fully observed. In that lens, a contract is a reward-shaping rule, and receipts are auxiliary observations that can enter the reward. Our analysis complements algorithmic reward design by emphasizing the constraints of limited liability, the need for equilibrium reasoning, and the central role of signal informativeness in determining subsidy requirements. The results deliver a language for comparing regimes: outcome-only evaluation, receipt-based monitoring, and hybrid schemes that combine the two.

The equilibrium concept also matters for practice. Many platform settings resemble a sequential relationship: tasks arrive over time, the principal updates terms based on state, and future opportunities depend on current performance. We therefore frame contracting in a Markov (state-dependent) manner and use subgame-perfect equilibrium (SPE) to capture credible continuation behavior. This is distinct from a one-shot Stackelberg view in which the principal commits once to an entire intertemporal payment policy. In many procurement and API contexts, full commitment is unrealistic: terms can be renegotiated, and users can switch providers. SPE is the appropriate discipline when contracts are offered repeatedly and must remain optimal in every continuation. At the same time, the Markov structure preserves tractability and aligns with how real systems are managed (pricing and evaluation policies depend on observable task class, risk tier, or other state variables).

Our main contributions can be summarized as follows. First, we provide a sharp characterization of how receipt informativeness translates into minimal expected payment under limited liability. The key statistic is a *separability* measure: the extent to which the distribution of receipts shifts when the agent exerts high compute rather than low compute. When separability is positive, we show that simple event-based contracts—pay a fixed bonus if and only if a carefully chosen receipt event occurs—can implement high effort, and we give an explicit upper bound on the expected subsidy required. The structure mirrors classic results in moral hazard with nonnegative transfers:

optimal incentives concentrate payments on states that are most diagnostic of effort, which here correspond to receipts with high likelihood ratios.

Second, we compare receipt contracting to outcome-only contracting through what we call a *proxy gap*. When outcomes are weakly responsive to effort (for instance, when success is largely determined by task difficulty or stochastic evaluation), outcome-only incentives become expensive: to compensate the agent for costly compute, the principal must pay large bonuses on rare successes or pay broadly across outcomes, both of which raise expected transfers. Receipts can dramatically reduce this cost when they are more sensitive to effort than outcomes. Our bounds make this comparison transparent by expressing the minimal expected payment in each regime as a cost term divided by a separability term, and thus relating welfare differences to the ratio of receipt versus outcome informativeness.

Third, we extend the logic to sequential settings. When tasks are repeated, continuation values enter the agent's incentive constraint: deviating today may alter future states and future payments. We show how to bound the total discounted subsidy required to sustain high effort along the equilibrium path by decomposing the dynamic problem into per-state incentive problems plus a continuation-value correction. This provides a practical accounting identity for "how expensive it is to pay for thinking" over time, and it highlights where receipts matter most: states with high compute cost and low separability are the bottlenecks.

Fourth, we address robustness. Receipt models can drift: logging pipelines change, tool APIs update, and the mapping from compute to telemetry is imperfectly understood. We therefore study misspecification in total variation and derive a simple "nudging" rule: add slack to the incentive constraint that scales with model uncertainty and inversely with squared separability. The implication is a concrete design principle for practice: if receipts barely separate effort, then not only are incentives expensive, they are also brittle to small measurement errors; conversely, strong receipts buy both efficiency and robustness.

We close the introduction with two limitations that guide interpretation. We model the agent as risk-neutral and focus on limited liability, which fits many contracting environments but abstracts from risk-sharing and reputation concerns. We also assume receipts are verifiable and non-manipulable; in reality, receipts can be noisy, strategically obfuscated, or constrained by privacy policies, and designing trustworthy telemetry is itself a technical problem. Our contribution is not to solve these engineering challenges, but to provide a clear economic calculus for why they are worth solving and what is gained when they are solved well.

The roadmap is as follows. In the next section, we formalize the static receipt-contracting problem with binary effort and derive the minimal-expected-payment contract as a linear program, obtaining closed-form event-based solutions via likelihood-ratio reasoning. Subsequent sections build the sequen-

tial extension, the outcome-only lower bound and proxy gap comparison, and the robustness analysis under receipt drift, returning throughout to the practical question that motivates the theory: when do logs and proofs let us pay less, and get more, for the same amount of "thinking"?

2 Static receipt contracting with binary effort

We begin with a one-shot version of the relationship at a fixed state s, suppressing the state index to keep notation light. The principal commits to a limited-liability payment rule $b : \mathcal{Y} \times \mathcal{R} \to \mathbb{R}_{\geq 0}$, the agent privately chooses effort $e \in \{0, 1\}$ (interpreted as a higher compute/search/tool-use intensity when $e = 1$), and then a joint signal (Y, R) is realized with distribution $P(y, r \mid e)$. The principal values only the outcome, $v(Y)$, while the agent incurs effort cost $c(e)$ with incremental cost $\Delta c := c(1) - c(0) > 0$. Payoffs are

$$U^P = \mathbb{E}\big[v(Y) - b(Y, R)\big], \qquad U^A = \mathbb{E}\big[b(Y, R) - c(e)\big], \tag{1}$$

where expectations are taken under the induced distribution of (Y, R) given the agent's equilibrium effort.

The static question we study in this section is intentionally narrow: if the principal wants to implement high effort $e = 1$, how expensive does the cheapest incentive scheme have to be under limited liability? This isolates the pure "incentive cost of compute" from other considerations (risk sharing, participation rents, or dynamic reputation) that we treat separately in later sections. In this static benchmark, the principal's value term $v(Y)$ is unaffected by the payment rule once we fix the target action $e = 1$, so the principal's contracting problem reduces to minimizing expected transfers subject to incentive compatibility.

Let

$$B_e := \mathbb{E}\big[b(Y, R) \mid e\big] = \sum_{y \in \mathcal{Y}} \sum_{r \in \mathcal{R}} P(y, r \mid e)\, b(y, r) \tag{2}$$

(with the obvious integral analogue when signals are continuous). The agent chooses $e = 1$ iff

$$B_1 - c(1) \;\geq\; B_0 - c(0) \qquad \Longleftrightarrow \qquad B_1 - B_0 \;\geq\; \Delta c. \tag{3}$$

Limited liability implies we can raise B_1 but cannot lower B_0 through fines; incentives must be created by promising bonuses on some realizations of the observable signal.

Minimal implementation as a linear program. Fix a target action $e = 1$. Among all nonnegative payment rules $b(\cdot, \cdot) \geq 0$ that satisfy (3), we

define the *minimal expected payment* as the value of

$$\begin{aligned} \min_{b:\mathcal{Y}\times\mathcal{R}\to\mathbb{R}_{\geq 0}} \quad & \mathbb{E}\big[b(Y,R) \mid e=1\big] \\ \text{s.t.} \quad & \mathbb{E}\big[b(Y,R) \mid e=1\big] - \mathbb{E}\big[b(Y,R) \mid e=0\big] \;\geq\; \Delta c. \end{aligned} \tag{4}$$

This is a linear program (LP): the objective and constraint are linear in the decision variables $\{b(y,r)\}$, and limited liability is a set of coordinate-wise nonnegativity constraints.

Two immediate observations sharpen the economic content. First, only the distributional shift in the *contractible signal* matters. Formally, if we define the augmented signal $Z := (Y,R)$, then the LP depends on the pair of measures $\{P(\cdot \mid 1), P(\cdot \mid 0)\}$ over $\mathcal{Z} := \mathcal{Y} \times \mathcal{R}$ and not on any other primitives. Second, because b enters only through expectations, there is never a reason to pay on signal realizations that are *less* likely under high effort than under low effort: if $P(z \mid 1) \leq P(z \mid 0)$, then paying at z relaxes the constraint weakly in the wrong direction. In an optimal solution we can set $b(z) = 0$ on all such z without violating (3) and while (weakly) lowering expected payments.

Extreme-point structure: bonuses on diagnostic events. The LP (4) has a single nontrivial inequality constraint besides nonnegativity. This geometry implies that optimal solutions take a stark form: incentives are provided by concentrating payment on the most diagnostic realizations of the signal. To see the logic cleanly, consider any measurable event $E \subseteq \mathcal{Z}$ and a simple *event contract*

$$b(z) = x \cdot \mathbf{1}\{z \in E\} \qquad \text{for some } x \geq 0. \tag{5}$$

Under this contract,

$$B_1 = x\,P(E \mid 1), \qquad B_0 = x\,P(E \mid 0), \tag{6}$$

so the incentive constraint (3) becomes

$$x\big(P(E \mid 1) - P(E \mid 0)\big) \;\geq\; \Delta c. \tag{7}$$

Whenever the probability gap $P(E \mid 1) - P(E \mid 0)$ is positive, the cheapest x satisfying (7) is

$$x(E) = \frac{\Delta c}{P(E \mid 1) - P(E \mid 0)}, \tag{8}$$

and the corresponding expected payment under high effort is

$$B_1(E) = x(E)\,P(E \mid 1) = \Delta c \cdot \frac{P(E \mid 1)}{P(E \mid 1) - P(E \mid 0)}. \tag{9}$$

This simple calculation already delivers the key intuition: to make effort cheap to implement under limited liability, we want contractible events whose probability is much larger under $e = 1$ than under $e = 0$. Paying on such events creates a large incentive wedge per expected dollar. Conversely, if every event has only a small probability gap, then any bonus that generates Δc units of incentive must be large in expectation.

A particularly transparent (and often nearly sharp) bound comes from choosing an event E that maximizes the raw probability gap $P(E \mid 1) - P(E \mid 0)$. For that choice, (8) yields a bonus level proportional to Δc divided by this maximal gap, and (9) implies the expected payment is no larger than that same ratio (since $P(E \mid 1) \leq 1$). We will elevate this maximal-gap statistic in the next section because it provides a single, interpretable measure of how "separating" a telemetry channel is.

Closed-form optimality via likelihood ratios (Neyman–Pearson logic). While maximal-gap events give clean bounds, the LP (4) can be solved more sharply by ranking signal realizations by likelihood ratios. Assume for the moment that $\mathcal{Z}$ is finite. Write $p_1(z) := P(Z = z \mid 1)$ and $p_0(z) := P(Z = z \mid 0)$, and define the *likelihood ratio* $\ell(z) := p_1(z)/p_0(z)$ (with the convention that $\ell(z) = +\infty$ when $p_0(z) = 0$ and $p_1(z) > 0$). Then for any candidate payment vector $\{b(z)\}_{z \in \mathcal{Z}}$, the IC constraint can be written as

$$\sum_{z \in \mathcal{Z}} (p_1(z) - p_0(z))\, b(z) \;\geq\; \Delta c. \tag{10}$$

Because the objective is $\sum_z p_1(z) b(z)$, a natural "bang-per-buck" index is

$$\frac{p_1(z) - p_0(z)}{p_1(z)} = 1 - \frac{1}{\ell(z)}. \tag{11}$$

Paying on a realization with larger $\ell(z)$ yields a larger incentive wedge per unit of expected payment under $e = 1$. The LP therefore places all payment weight on realizations with the highest likelihood ratios, and in generic finite cases an optimal solution pays a single bonus on a single most-diagnostic realization (or splits across ties). More generally, when we restrict attention to event contracts of the form (5), the optimal event $E^\dagger$ solves

$$E^\dagger \in \arg\max_{E \subseteq \mathcal{Z}} \frac{P(E \mid 1) - P(E \mid 0)}{P(E \mid 1)}, \tag{12}$$

equivalently $\arg\min_E \frac{P(E|0)}{P(E|1)}$, and such an $E^\dagger$ is obtained by thresholding $\ell(z)$: include z with $\ell(z)$ above a cutoff, and (if necessary) randomize at the cutoff to hit the constraint at equality. This is precisely the Neyman–Pearson lemma in statistical testing form, with $e = 1$ as the "alternative" and $e = 0$ as the "null": the cheapest way to induce effort is to reward the events that are most informative about effort.

Receipts versus outcomes in the static problem. Finally, we connect the mathematics back to the contracting interpretation. If the principal can contract on the full augmented signal $Z = (Y, R)$, then the preceding analysis applies to Z directly, and receipts can only help relative to outcome-only incentives because they enrich the signal space on which likelihood-ratio screening can be performed. If, instead, the principal is restricted to receipt-only contracts $b(r)$ (or outcome-only contracts $b(y)$), we recover the same LP with $\mathcal{Z}$ replaced by $\mathcal{R}$ (or $\mathcal{Y}$). The substantive difference between regimes is therefore not philosophical but statistical: how much the distribution shifts with effort on the available contractible signal. In the next section we package that shift into a separability index and translate it into explicit subsidy bounds and practical guidance for telemetry design.

Receipt separability as an incentive lever. To summarize the static analysis into a single statistic we can reason about, we now specialize the contractible signal to receipts alone and quantify how strongly receipts move with effort. Fix a state s (suppressed) and consider receipt-only contracts $b : \mathcal{R} \to \mathbb{R}_{\geq 0}$. Let P_1 and P_0 denote the induced distributions of R under $e = 1$ and $e = 0$. We define the *receipt separability* (or compute-signal separability)

$$\Delta \;:=\; \mathrm{TV}(P_1, P_0) \;=\; \max_{E \subseteq \mathcal{R}} \big(P_1(E) - P_0(E)\big) \;=\; \frac{1}{2} \sum_{r \in \mathcal{R}} \big|P_1(r) - P_0(r)\big| \quad (13)$$

(with the obvious integral form when $\mathcal{R}$ is continuous). Economically, Δ measures how much *statistical evidence* the receipt channel can, in principle, provide about whether the agent used high compute. The max-over-events formulation is particularly revealing: it is the largest increase in the probability of passing any binary "audit test" E when the agent exerts $e = 1$ rather than $e = 0$.

A sharp expected-subsidy bound under limited liability. Receipt separability maps directly into an upper bound on the minimal expected payment required to implement $e = 1$. Consider an event contract $b(r) = x\mathbf{1}\{r \in E\}$, and choose $E^* \in \arg\max_E (P_1(E) - P_0(E))$. Then $P_1(E^*) - P_0(E^*) = \Delta$ by definition, so setting

$$x^* \;=\; \frac{\Delta c}{\Delta} \quad (14)$$

makes the incentive constraint bind:

$$\mathbb{E}[b(R) \mid 1] - \mathbb{E}[b(R) \mid 0] = x^* \big(P_1(E^*) - P_0(E^*)\big) = \Delta c.$$

The expected payment under high effort is

$$\mathbb{E}[b(R) \mid 1] = x^* P_1(E^*) = \Delta c \cdot \frac{P_1(E^*)}{\Delta} \;\leq\; \frac{\Delta c}{\Delta}, \quad (15)$$

since $P_1(E^*) \leq 1$. Thus, whenever $\Delta > 0$, there exists a feasible receipt-only scheme with expected subsidy no larger than $\Delta c/\Delta$. The bound is often close to the true optimum: if the maximizing event E^* also has $P_1(E^*)$ not too small (i.e., high effort triggers the event with reasonable probability), then (15) is within a small constant factor of the minimal expected payment.

Two aspects of (15) deserve emphasis. First, the dependence is *linear* in the incremental compute cost Δc and *inverse* in separability Δ. This makes separability an interpretable "exchange rate" between dollars and incentives: each expected dollar can buy at most Δ units of incentive wedge (in the sense of shifting $\mathbb{E}[b \mid 1] - \mathbb{E}[b \mid 0]$), so achieving a wedge of Δc costs at least on the order of $\Delta c/\Delta$. Second, the argument uses only limited liability and verifiability of receipts; it does not rely on any special structure of $v(\cdot)$, risk aversion, or participation constraints. In that sense, Δ isolates the pure *informativeness* contribution of telemetry to incentive provision.

When is high effort close to first best? In the one-shot benchmark, the principal's implementation decision trades off the value gain from high effort against the required transfer. Let

$$\Delta v \;:=\; \mathbb{E}[v(Y) \mid e = 1] - \mathbb{E}[v(Y) \mid e = 0] \tag{16}$$

denote the principal's incremental gross value from inducing $e = 1$. Because transfers are a deadweight loss for the principal in this risk-neutral, static setup, a sufficient condition for implementing high effort using receipt incentives is

$$\Delta v \;\geq\; \frac{\Delta c}{\Delta}. \tag{17}$$

Condition (17) is intentionally interpretable: high effort is worthwhile when (i) compute is not too costly (Δc small), and/or (ii) receipts are sufficiently separating (Δ large), so that incentives can be created cheaply. Conversely, when telemetry barely changes with compute ($\Delta \approx 0$), limited liability forces implementation costs to explode, and the principal optimally tolerates low effort unless the value gain Δv is enormous.

A closely related budgeting view is often more practical. Suppose the principal can afford expected transfers no larger than $\bar{B}$ under high effort. Then a sufficient condition for implementability is

$$\bar{B} \;\geq\; \frac{\Delta c}{\Delta} \qquad \Longleftrightarrow \qquad \Delta \;\geq\; \frac{\Delta c}{\bar{B}}. \tag{18}$$

Thus Δ can be read as a *minimum telemetry quality* requirement for a given payment budget and compute cost. This perspective is useful in platform settings where budgets (or subsidy policies) are set administratively, while telemetry design is the main engineering degree of freedom.

Interpretation of Δ via testing and classification. The total variation distance has a standard operational meaning that aligns well with our contracting interpretation. Consider the best possible (measurable) test $\phi : \mathcal{R} \to \{0, 1\}$ for distinguishing $e = 1$ from $e = 0$ based on R. Then

$$\Delta = \max_{\phi} \Big(\mathbb{P}(\phi(R) = 1 \mid e = 1) - \mathbb{P}(\phi(R) = 1 \mid e = 0) \Big). \tag{19}$$

That is, Δ is the maximal *true-positive minus false-positive* gap achievable by any audit rule. In words: if we are only allowed to look at receipts and make a binary decision (pay bonus or not), then Δ is the best achievable advantage that high effort has in passing the test. Our subsidy bound is exactly the contracting analogue of this testing limit: to buy Δc units of incentive under limited liability, we must scale up the bonus so that the test advantage Δ is amplified into a payment wedge of size Δc.

This interpretation also clarifies why telemetry coarsening and privacy constraints are costly. Any post-processing of receipts (hashing, binning, adding noise) that makes P_1 and P_0 closer in total variation mechanically reduces Δ, which raises the minimal expected subsidy roughly in proportion to $1/\Delta$. In practice, this forces a concrete trade-off between privacy/overhead and incentive efficiency.

Estimating Δ from telemetry. To use the bound quantitatively, we need a way to estimate Δ from observed logs. Conceptually, the cleanest approach is experimental: run the same task distribution (or state s) under two controlled compute regimes corresponding to $e = 1$ and $e = 0$, record receipts, and form empirical distributions $\widehat{P}_1, \widehat{P}_0$. For discrete receipts, the plug-in estimator

$$\widehat{\Delta} \;=\; \frac{1}{2} \sum_{r \in \mathcal{R}} \big| \widehat{P}_1(r) - \widehat{P}_0(r) \big| \tag{20}$$

is immediate. When $\mathcal{R}$ is large (as is typical for rich logs), direct estimation of the full distributions is sample-inefficient; nevertheless, the max-event formulation suggests a more scalable route: we can estimate Δ by searching over a restricted class of events/tests that are implementable and meaningful for contracting (e.g., thresholds on runtime, number of tool calls, presence of a proof object, verification success, depth of a search tree). Concretely, if $\mathcal{E}$ is a class of candidate events, we can compute the empirical gap

$$\widehat{\Delta}_{\mathcal{E}} \;:=\; \max_{E \in \mathcal{E}} \big(\widehat{P}_1(E) - \widehat{P}_0(E) \big), \tag{21}$$

which is a *lower bound* on Δ (since $\mathcal{E}$ may not contain the optimal event). This lower bound is often exactly what we want for contract design, because it directly identifies an implementable bonus trigger E together with its estimated incentive leverage.

A complementary, high-dimensional method is to train a classifier to predict the compute regime from receipts and translate its performance into a bound on total variation. By (19), any learned test $\widehat{\phi}$ yields a computable lower bound

$$\Delta \;\geq\; \widehat{\Delta}(\widehat{\phi}) \;:=\; \widehat{\mathbb{P}}(\widehat{\phi}(R)=1 \mid e=1) - \widehat{\mathbb{P}}(\widehat{\phi}(R)=1 \mid e=0),$$

where the probabilities are evaluated on held-out data. This approach naturally accommodates complex receipts (vectors of tool traces, lengths, verifier outputs) and aligns with standard ML evaluation pipelines.

We should be explicit about limitations. First, Δ is state-dependent: if task difficulty changes, or if the distribution of prompts shifts, then $P(R \mid e)$ shifts as well. Second, our assumption (H1) that receipts are non-manipulable is substantive: if the agent can directly influence receipts without incurring the intended compute cost, then the estimated Δ will overstate true separability with respect to *effort*, and contracts based on E^* can be gamed. For this reason, in applied settings we view Δ not as a purely statistical quantity but as a joint property of telemetry *and* system integrity (attestation, secure logging, verifier soundness). Under those caveats, receipt separability provides a compact, empirically grounded way to connect telemetry design decisions to concrete subsidy requirements.

Outcome-only contracting as a benchmark. We now ask what can be achieved if the principal cannot condition transfers on receipts and must instead use only the realized task outcome Y. This is a natural benchmark for settings where telemetry is unavailable, privacy policy forbids logging, or the platform's contract language is restricted to end-to-end scores (e.g., pass/fail on a hidden test set). Formally, we restrict attention to limited-liability contracts of the form $b : \mathcal{Y} \to \mathbb{R}_{\geq 0}$. We again fix a state s and suppress it in notation.

As in the receipt case, the key question is whether the distribution of the contractible signal shifts with effort. Let Q_1 and Q_0 denote the induced distributions of Y under $e=1$ and $e=0$. We define *outcome separability*

$$\Delta^y \;:=\; \mathrm{TV}(Q_1, Q_0) \;=\; \max_{A \subseteq \mathcal{Y}} \big(Q_1(A) - Q_0(A)\big), \tag{22}$$

with the usual L^1 representation when $\mathcal{Y}$ is discrete. Economically, Δ^y is the best advantage high compute has in passing any binary test based solely on the observed outcome (e.g., exceeding a score threshold).

A lower bound on required subsidies under outcome-only payments. The outcome-only restriction is not merely a modeling convenience: it imposes a fundamental limit on the maximal incentive wedge attainable

per expected dollar under limited liability. Consider the minimal-expected-payment implementation problem

$$\min_{b\geq 0} \mathbb{E}[b(Y) \mid e=1] \qquad \text{s.t.} \qquad \mathbb{E}[b(Y) \mid 1] - \mathbb{E}[b(Y) \mid 0] \;\geq\; \Delta c.$$

A simple but sharp bound follows directly from the total-variation definition. For any nonnegative function $b(\cdot)$, scaling and the max-over-events representation imply

$$\mathbb{E}[b(Y) \mid 1] - \mathbb{E}[b(Y) \mid 0] \;\leq\; \Delta^y \cdot \sup_y b(y), \tag{23}$$

and, moreover, $\mathbb{E}[b(Y) \mid 1] \geq Q_1(A) \cdot \inf_{y\in A} b(y)$ for any event A. Combining these ideas in the standard LP/extreme-point logic (the cheapest nonnegative transfer concentrates payment on an event), we obtain the benchmark implication: to generate an incentive wedge of Δc, the principal must pay at least on the order of $\Delta c/\Delta^y$ in expectation under $e=1$. One convenient statement is

$$\mathbb{E}[b(Y) \mid 1] \;\geq\; \frac{\Delta c}{\Delta^y}, \tag{24}$$

interpreting $\Delta c/\Delta^y = +\infty$ when $\Delta^y = 0$. Thus if outcomes are (nearly) insensitive to compute, outcome-only incentives are prohibitively expensive (or impossible) under limited liability.

A concrete illustration: pass/fail outcomes. The bound (24) is especially transparent when $Y \in \{0,1\}$ is a binary success indicator. Let $p_1 = \mathbb{P}(Y=1 \mid e=1)$ and $p_0 = \mathbb{P}(Y=1 \mid e=0)$, so $\Delta^y = |p_1 - p_0|$. The natural outcome-only contract pays x upon success: $b(Y) = x\mathbf{1}\{Y=1\}$. The incentive constraint is

$$x(p_1 - p_0) \;\geq\; \Delta c, \qquad \Rightarrow \qquad x \;\geq\; \frac{\Delta c}{p_1 - p_0},$$

and the expected payment under high effort is $\mathbb{E}[b(Y) \mid 1] = p_1 x \geq p_1 \Delta c/(p_1 - p_0) \geq \Delta c/(p_1 - p_0) = \Delta c/\Delta^y$. In this canonical scoring environment, (24) is not merely a loose asymptotic statement: it describes the exact scaling of the cheapest implementation. The economic punchline is that if extra compute mainly affects *how* the answer is produced (e.g., longer reasoning traces) rather than *whether* it is correct, then $p_1 \approx p_0$ and outcome-only payment must blow up.

The proxy gap: why receipts can dominate outcomes. Outcome-only contracts are a special case of receipt contracts in which the contractible signal is a low-dimensional proxy for the agent's hidden action. The natural comparison is therefore between Δ^y and receipt separability Δ from (13). The ratio

$$G \;:=\; \frac{\Delta}{\Delta^y} \tag{25}$$

summarizes what we might call the *proxy gap*: how much more informative receipts are than outcomes about whether the agent actually expended high compute. When $G \gg 1$, receipts can reduce the minimal expected subsidy by a large factor relative to outcome-only contracting. Indeed, combining the receipt upper bound $\mathbb{E}[b(R) \mid 1] \lesssim \Delta c/\Delta$ with the outcome-only lower bound (24) yields the qualitative welfare claim:

$$\text{(minimal outcome-only expected payment)} \gtrsim \frac{\Delta c}{\Delta^y} \qquad \text{vs.} \qquad \text{(feasible receipt expected payment)} \lesssim$$

Thus, holding fixed the compute cost Δc, the achievable implementation cost differs by a factor comparable to G. Practically, this is the formal sense in which telemetry can be "worth paying for": it expands the set of implementable effort policies under any fixed transfer budget, and it does so precisely when outcomes are a weak proxy for effort.

Welfare gaps under a payment budget. The proxy gap becomes especially policy-relevant in environments with hard budget caps (procurement rules, platform subsidy schedules, or internal spend limits). Suppose the principal can afford at most $\bar{B}$ in expected transfers under high effort. Outcome-only contracting can implement $e = 1$ only if $\bar{B} \geq \Delta c/\Delta^y$ (at least as a necessary condition by (24)), whereas receipt contracting can implement $e = 1$ whenever $\bar{B} \geq \Delta c/\Delta$ (sufficient by our earlier construction). Therefore, whenever

$$\frac{\Delta c}{\Delta} \leq \bar{B} < \frac{\Delta c}{\Delta^y}, \tag{26}$$

the principal can induce high compute with receipts but *cannot* do so with outcomes alone, no matter how cleverly payments are shaped over $\mathcal{Y}$. In such regions, the welfare loss from outcome-only contracting is not a marginal inefficiency but a discrete feasibility gap: the principal is forced into low effort even if high effort would be value-enhancing absent contract constraints.

When do outcomes suffice? (Observed-action-like regimes). The outcome-only benchmark is not meant to suggest that receipts are always necessary. There are important regimes in which outcomes already act like a nearly observed action. In our notation, this corresponds to Δ^y being large: the distribution of Y shifts sharply with compute. In the extreme, if Y deterministically reveals effort (e.g., $Y = y_1$ under $e = 1$ and $Y = y_0$ under $e = 0$ with disjoint support), then $\Delta^y = 1$ and the lower bound (24) becomes $\mathbb{E}[b(Y) \mid 1] \geq \Delta c$, which is essentially the first-best "pay cost" benchmark. More generally, outcome-only incentives are effective when (i) evaluation is fine-grained enough that marginal quality improvements are detectable, and (ii) noise in Y is small relative to the effect of compute. In such cases, the engineering priority may be to improve evaluation (increase Δ^y) rather than to log more receipts (increase Δ).

This clarifies a useful design principle: when we cannot or do not want to rely on receipts, we can sometimes recover incentive power by making outcomes more informative—for instance, replacing a single pass/fail test with a battery of independent checks, using stronger verifiers, or eliciting richer graded outputs. Each of these interventions can increase Δ^y by making the outcome distribution more sensitive to compute, thereby reducing the outcome-only subsidy required by (24).

Limitations of the outcome-only view. Finally, we emphasize what the outcome-only benchmark leaves out. First, Y is often delay-prone and potentially manipulable (distribution shift, dataset leakage), whereas receipts are closer to the agent's internal action and can be secured via attestation. Second, even when outcomes are informative on average, they may be too sparse to support fine-grained per-task incentives, whereas receipts can provide dense signals (timeouts, tool-usage patterns, proof objects) that vary at the right temporal resolution. For these reasons, Δ^y is best interpreted as a best-case contracting limit under an intentionally austere signal space. The next step is to return to the sequential setting and show how receipt leverage composes across states once continuation values enter the incentive constraint.

Sequential extension: Markov contracting with receipts. We now lift the analysis from a single state to a controlled stochastic process. The key economic question is whether the "receipt leverage" captured by Δ_s can be used repeatedly, state by state, without creating intertemporal distortions or requiring the principal to solve a fully history-dependent mechanism design problem. Our modeling choice of *Markov contracts* is designed precisely to make this composition transparent: at each visit to state s, the principal posts a contract $b_s(\cdot)$ that depends only on the current contractible signal, and the continuation game is summarized by the next state S'. In this sense, the sequential setting does not change what a receipt *is*; it changes what effort *does*, because effort may now affect not only the current receipt distribution but also the distribution of future states.

Bellman representation and the dynamic IC constraint. Fix a Markov strategy profile (contracts and effort choices). The agent's continuation value at state s admits the standard Bellman form

$$V^A(s) \;=\; \max_{e\in\{0,1\}} \Big\{ -c_s(e) + \mathbb{E}[b_s(Y,R) \mid e,s] + \gamma\, \mathbb{E}\big[V^A(S') \mid e,s\big] \Big\}, \quad (27)$$

where S' is the next state drawn from the equilibrium transition rule given (s,e) (i.e., integrating out (Y,R) and any exogenous shocks). Implementing high effort at s requires that $e=1$ solve (27). Writing the one-step deviation

constraint explicitly yields

$$\mathbb{E}[b_s(Y,R) \mid 1,s] - \mathbb{E}[b_s(Y,R) \mid 0,s] \;\geq\; \Delta c_s + \gamma\Big(\mathbb{E}\big[V^A(S') \mid 0,s\big] - \mathbb{E}\big[V^A(S') \mid 1,s\big]\Big). \tag{28}$$

The additional term is the dynamic analogue of an incentive spillover: if low effort makes it more likely the agent lands in states with higher continuation payoff, then current transfers must also offset this future advantage.

For compactness, define the *required incentive wedge* at s (given continuation values) as

$$\kappa_s \;:=\; \Delta c_s + \gamma\,\Delta V_s^A, \qquad \Delta V_s^A := \mathbb{E}\big[V^A(S') \mid 0,s\big] - \mathbb{E}\big[V^A(S') \mid 1,s\big]. \tag{29}$$

Then (28) becomes $\mathbb{E}[b_s \mid 1] - \mathbb{E}[b_s \mid 0] \geq \kappa_s$. Two polar cases are worth keeping in mind. If high effort tends to move the system to *better* states for the agent (in the sense of larger V^A), then $\Delta V_s^A < 0$ and dynamics *help* incentives by reducing the needed current wedge. If instead low effort preserves "easy" future rents, then $\Delta V_s^A > 0$ and the principal must subsidize more today.

Per-state decomposition as a minimal-implementation LP with continuation values. The principal's sequential problem is, in principle, a coupled fixed point: contracts determine effort, effort shapes transitions, transitions determine $V^A(\cdot)$ and $V^P(\cdot)$, and those values feed back into optimal contracts. The main structural simplification is that, in a finite-horizon game (or in a discounted infinite-horizon game under standard boundedness assumptions), subgame perfection lets us treat continuation values as *given* when solving the contract choice at the current state. Concretely, suppose we are at a subgame in which the continuation strategy profile from period $t+1$ onward is fixed, hence $V^A(\cdot)$ is pinned down. Then, to implement $e = 1$ at s as cheaply as possible subject to limited liability, the principal solves the per-state program

$$\min_{b_s \geq 0} \; \mathbb{E}[b_s(Y,R) \mid 1,s] \qquad \text{s.t.} \qquad \mathbb{E}[b_s(Y,R) \mid 1,s] - \mathbb{E}[b_s(Y,R) \mid 0,s] \;\geq\; \kappa_s. \tag{30}$$

Relative to the static LP, the only change is that Δc_s is replaced by κ_s. Economically, we can interpret $\gamma\,\Delta V_s^A$ as an *effective incremental cost of effort* created by the future: it is the amount of discounted continuation utility that high effort must compensate the agent for foregoing.

Closed-form carryover: the same event-pay contract, with an adjusted wedge. Because (30) has the same linear structure as the static problem, the extreme-point logic carries over essentially verbatim. In particular, if the principal is willing to use receipts alone (or to ignore Y when it is

not helpful), then for any state with $\Delta_s > 0$ there exists a receipt-contingent contract that implements $e = 1$ and satisfies

$$\mathbb{E}[b_s(R) \mid 1, s] \;\le\; \frac{\kappa_s}{\Delta_s} \;=\; \frac{\Delta c_s + \gamma\, \Delta V_s^A}{\Delta_s}. \tag{31}$$

Moreover, as in the one-shot case, an optimal minimal-expected-payment contract can be chosen to pay a constant on a single event $E_s^* \subseteq \mathcal{R}$ maximizing the probability gap $P_s(E \mid 1) - P_s(E \mid 0)$. The object doing all the work is still separability Δ_s; sequentiality enters only through the wedge κ_s.

This is the sense in which the sequential problem "decomposes": *holding fixed* the continuation values induced by future play, the current-state incentive provision problem reduces to the same one-step LP on the current signal space. The principal does not need to design a long menu of intertemporal payments to exploit receipts; she can, without loss of optimality for incentive provision, use a locally targeted payment rule that triggers on the most effort-informative receipt event at that state.

Constructing an SPE by backward induction (finite horizon). In a finite-horizon model, this decomposition yields an explicit equilibrium construction. At the terminal date T, continuation values are zero, so $\kappa_s = \Delta c_s$ and the optimal contract at T is exactly the static receipt contract. Given the period-T contracts, we can compute the induced $V_T^A(s)$ and hence $\Delta V_{T-1}^A(s)$ for each state at $T-1$. Plugging these into (30) gives the period-$T-1$ minimal implementation contracts, and so on backward to date 0. The resulting strategy profile is subgame perfect by construction: at each subgame, the principal chooses a contract that is optimal given the continuation, and the agent best responds by choosing $e = 1$ because the dynamic IC constraint is satisfied.

This backward-induction perspective also clarifies when the earlier "first-best everywhere" statement is meaningful. If the planner's first-best calls for $e = 1$ in every state along the equilibrium path, then the principal can implement exactly that policy provided $\Delta_s > 0$ on those states, with the per-visit expected transfer bounded by (31). Summing along the realized path yields the advertised total-subsidy bound:

$$\mathbb{E}\left[\sum_{t=0}^{T} \gamma^t\, b_{S_t}(Y_t, R_t)\right] \;\le\; \sum_{t=0}^{T} \gamma^t\, \mathbb{E}\left[\frac{\Delta c_{S_t} + \gamma\, \Delta V_{S_t}^A}{\Delta_{S_t}}\right]. \tag{32}$$

A convenient coarse corollary obtains by upper-bounding ΔV_s^A by the range of continuation values, but the sharper message is that dynamics matter only insofar as they create (or destroy) a continuation-value advantage to shirking.

When does the closed form suffice, and when do we need numerical dynamic programming? The construction above is "closed form" in the narrow sense that, conditional on knowing κ_s and Δ_s, we can write down an optimal payment rule at each state (an event-triggered contract) and an explicit bound on its expected cost. What is *not* closed form in general is the mapping $s \mapsto V^A(s)$ (and hence $s \mapsto \kappa_s$), because V^A is endogenous to the entire future contracting policy. There are, however, important regimes where κ_s becomes simple. If transitions are (approximately) independent of effort, or if effort affects only current receipts but not future states, then $\Delta V_s^A \approx 0$ and we revert to the static wedge $\kappa_s \approx \Delta c_s$ period by period. More generally, if the effect of effort on future states is monotone and known (high effort stochastically improves S' for both players), then $\Delta V_s^A \leq 0$ and the static bound $\Delta c_s/\Delta_s$ is conservative.

Outside these regimes—large state spaces, rich transition dependence on (Y, R), or strategic interactions in which contracts affect the agent's future rents in subtle ways—one should view (31) as a *local* characterization embedded in a global fixed point. Computing the equilibrium can then be posed as a dynamic program with incentive constraints: iterate on candidate continuation values, solve the per-state LP (30) to obtain implied policies and payoffs, and update values until convergence. This is exactly where the "engineering" complexity enters: not in designing complicated transfers, but in estimating the objects ($P_s(\cdot \mid e)$, transitions, and value functions) that determine κ_s and the relevant separating events.

Why the sequential view sets up robustness concerns. The sequential extension therefore delivers a clean conceptual takeaway: receipts buy us per-state incentive leverage, and Markov structure lets that leverage compose. At the same time, sequentiality highlights a fragility that the one-shot model hides: both Δ_s and the continuation wedge κ_s are model-dependent objects, and errors can accumulate across time through misestimated transitions or misspecified receipt likelihoods. This motivates the next step, where we study robustness and "nudging" margins that protect the IC constraints when the principal computes contracts using an estimated model rather than the true $P_s(r \mid e)$.

Robustness and nudging under drift. Our bounds so far are *comparative-statics clean* but *estimation fragile*: the principal computes a separating event (and hence a contract) using an estimated receipt model, while the implemented system may drift. In practice, drift can come from innocuous sources (logging changes, tool-version updates, caching policies) or strategic ones (an agent learns to route computation to mimic "high-effort" telemetry). The economic issue is that limited-liability incentive provision relies on a *probability gap*; if that gap is misestimated, the implemented contract may

accidentally subsidize low effort or fail to motivate high effort.

We therefore treat the principal as choosing contracts from an estimated model $\widehat{P}_s(r \mid e)$ and an estimated continuation-value object (e.g., a learned Q-function), while the true environment may differ within an uncertainty set. The goal is not full minimax mechanism design, but a transparent *nudging rule*: add a simple IC margin that restores incentive compatibility under bounded misspecification, and quantify the associated extra expected subsidy.

Uncertainty sets for receipt likelihoods. Fix a state s. We assume the true receipt distributions lie in a total-variation ball around the estimated ones:

$$\mathrm{TV}\Big(P_s(\cdot \mid e), \widehat{P}_s(\cdot \mid e)\Big) \;\le\; \eta_s, \qquad e \in \{0,1\}. \tag{33}$$

This is deliberately coarse: it captures any misspecification that perturbs event probabilities by at most η_s, without committing to a parametric form. The key inequality we repeatedly use is that for any event $E \subseteq \mathcal{R}$,

$$\Big|P_s(E \mid e) - \widehat{P}_s(E \mid e)\Big| \;\le\; \eta_s, \qquad e \in \{0,1\}, \tag{34}$$

and therefore the incentive-relevant probability gap is perturbed by at most $2\eta_s$:

$$\Big|\big(P_s(E \mid 1) - P_s(E \mid 0)\big) - \big(\widehat{P}_s(E \mid 1) - \widehat{P}_s(E \mid 0)\big)\Big| \;\le\; 2\eta_s. \tag{35}$$

Robust IC via a slackened wedge. Recall that the dynamic IC constraint at s takes the form

$$\mathbb{E}[b_s(Y,R) \mid 1, s] - \mathbb{E}[b_s(Y,R) \mid 0, s] \;\ge\; \kappa_s, \tag{36}$$

where $\kappa_s = \Delta c_s + \gamma\, \Delta V_s^A$ is the required incentive wedge given continuation values. Suppose we restrict attention (without loss for minimal subsidy) to receipt-only event contracts of the form $b_s(r) = x\,\mathbf{1}\{r \in E\}$ for some $x \ge 0$ and event $E \subseteq \mathcal{R}$. Under the estimated model, the IC constraint is

$$x\Big(\widehat{P}_s(E \mid 1) - \widehat{P}_s(E \mid 0)\Big) \;\ge\; \kappa_s. \tag{37}$$

Under drift, the true gap may be smaller. A sufficient robustification is to replace κ_s by $\kappa_s + \xi_s$ in (37), where ξ_s offsets the worst-case reduction in the probability gap induced by (35). Concretely, if we select an event E with estimated gap $\widehat{\Delta}_s(E) := \widehat{P}_s(E \mid 1) - \widehat{P}_s(E \mid 0)$, then the true gap satisfies

$$P_s(E \mid 1) - P_s(E \mid 0) \;\ge\; \widehat{\Delta}_s(E) - 2\eta_s.$$

Thus, choosing

$$x \;=\; \frac{\kappa_s}{\widehat{\Delta}_s(E) - 2\eta_s} \qquad \text{(provided } \widehat{\Delta}_s(E) > 2\eta_s\text{)} \tag{38}$$

guarantees (36) for all true models satisfying (33). Equivalently, one can implement this by solving the estimated-model problem but adding a slack margin

$$\xi_s(E) \;=\; \kappa_s \cdot \frac{2\eta_s}{\widehat{\Delta}_s(E) - 2\eta_s}, \tag{39}$$

which is exactly the extra wedge the principal must "nudge" into the contract to cover worst-case drift.

When we choose E optimally under the estimated model (e.g., a maximizing event $\widehat{E}_s^* \in \arg\max_E \widehat{\Delta}_s(E)$), write $\widehat{\Delta}_s := \widehat{\Delta}_s(\widehat{E}_s^*)$. Then the robust expected payment under high effort is bounded by

$$\mathbb{E}[b_s(R) \mid 1, s] \;=\; x\, P_s(E \mid 1) \;\le\; \frac{\kappa_s}{\widehat{\Delta}_s - 2\eta_s}, \tag{40}$$

using $P_s(E \mid 1) \le 1$. Comparing (40) to the non-robust proxy $\kappa_s/\widehat{\Delta}_s$ yields the incremental "price of robustness"

$$\frac{\kappa_s}{\widehat{\Delta}_s - 2\eta_s} - \frac{\kappa_s}{\widehat{\Delta}_s} \;=\; \kappa_s \cdot \frac{2\eta_s}{\widehat{\Delta}_s(\widehat{\Delta}_s - 2\eta_s)} \;=\; O\!\left(\eta_s \,\frac{\kappa_s}{\widehat{\Delta}_s^2}\right) \quad \text{when } \eta_s \ll \widehat{\Delta}_s. \tag{41}$$

This recovers the central curvature: robustness costs scale like η_s/Δ_s^2, so weakly separating receipts are disproportionately fragile.

Robustness with approximation error in learned continuation values. Receipt drift is not the only misspecification. In sequential problems, the wedge κ_s depends on continuation values; in modern implementations these are often produced by function approximation (value networks, fitted Q-iteration, etc.). Let $\widehat{V}^A$ denote the principal's estimate of the agent's continuation value under the continuation equilibrium, and suppose the estimation error is bounded uniformly:

$$\left\| \widehat{V}^A - V^A \right\|_\infty \;\le\; \varepsilon. \tag{42}$$

Then the induced error in the continuation wedge satisfies

$$\left| \gamma\Big(\mathbb{E}[V^A(S') \mid 0, s] - \mathbb{E}[V^A(S') \mid 1, s]\Big) - \gamma\Big(\mathbb{E}[\widehat{V}^A(S') \mid 0, s] - \mathbb{E}[\widehat{V}^A(S') \mid 1, s]\Big) \right| \;\le\; 2\gamma\varepsilon, \tag{43}$$

since each expectation shifts by at most ε. Therefore, a contract computed using $\widehat{\kappa}_s := \Delta c_s + \gamma(\mathbb{E}[\widehat{V}^A(S') \mid 0, s] - \mathbb{E}[\widehat{V}^A(S') \mid 1, s])$ should be nudged by an additional margin $2\gamma\varepsilon$ to remain valid for the true κ_s.

Combining receipt drift (33) and value error (42), a sufficient robust implementation rule is: pick an event E using the estimated receipt model, and set

$$x \;=\; \frac{\widehat{\kappa}_s + 2\gamma\varepsilon}{\widehat{\Delta}_s(E) - 2\eta_s}, \qquad \text{requiring } \widehat{\Delta}_s(E) > 2\eta_s. \tag{44}$$

This yields a clean interpretation: uncertainty in receipts shrinks the usable separability (denominator), while uncertainty in dynamic incentives expands the required wedge (numerator).

A diagnostic: when robustness is expensive (or impossible). The expressions above provide an operational diagnostic. The robust contract is well-behaved when η_s is small relative to $\widehat{\Delta}_s$, in which case the extra subsidy is second-order in η_s as in (41). By contrast, when $\widehat{\Delta}_s$ is small, robustness becomes rapidly expensive; indeed if

$$\widehat{\Delta}_s \;\le\; 2\eta_s, \tag{45}$$

then the robust denominator in (44) is non-positive, meaning that *no* non-negative event-pay contract computed from $\widehat{P}$ can guarantee a positive incentive gap under all models in the TV ball. Economically, (45) says that the principal's uncertainty is large enough to erase the very statistical distinction between high- and low-effort receipts. In such regimes, one must either (i) improve telemetry so that true Δ_s increases, (ii) reduce uncertainty η_s via better calibration/validation, (iii) fall back on outcome-based or hybrid contracts (accepting higher subsidy), or (iv) introduce additional enforcement instruments (audits, hard compute caps, or ex post verification) that effectively enlarge the receipt space.

Why this "nudging" viewpoint is useful. The reason we emphasize margins rather than full robust optimization is practical: the principal can keep using the simple extreme-point structure (pay on one informative event) and only adjust the level x by a transparent safety factor. The resulting comparative statics remain the same but sharpen into an engineering rule of thumb: measure (or lower-bound) separability, upper-bound drift, and scale subsidies by $(\widehat{\Delta}_s - 2\eta_s)^{-1}$ rather than $\widehat{\Delta}_s^{-1}$. This also clarifies what the principal should monitor online: not just average task success, but the *stability* of receipt likelihood ratios across time, since changes that reduce Δ_s have an amplified effect on the required payments.

Looking ahead: learning the objects that robustness needs. These robustness bounds deliberately speak in terms of primitives the principal must estimate: $\widehat{P}_s(r \mid e)$ (to get $\widehat{\Delta}_s$ and candidate events), η_s (to quantify drift), and the continuation values (to get $\widehat{\kappa}_s$ and ε-type error bars). This sets up the next step: how to *learn* principal policies and contracts, and how to validate them against a black-box agent while maintaining a monitoring pipeline that detects when $\widehat{\Delta}_s$ is eroding or when the robust condition $\widehat{\Delta}_s > 2\eta_s$ is close to failing.

3 Learning and implementation

Thus far we have treated the objects that enter contracting—receipt likelihoods, separability, continuation wedges—as primitives. In an LLM setting they are not primitives: they must be estimated from interaction data, and they evolve as models, toolchains, and workloads change. The design problem is therefore *jointly* (i) statistical (learning $P_s(y, r \mid e)$ and the induced Δ_s) and (ii) strategic (anticipating that the agent responds to the contract). In this section we outline an implementation pipeline that keeps the economics transparent: we retain simple contract classes with closed-form incentive leverage, and we learn only the quantities needed to choose among them and to set their levels.

A model-free view of the principal's problem. Operationally, the principal controls a *policy over contracts*. Let $\pi_\theta(\cdot \mid s)$ denote a parametric mapping from a state representation s to a contract parameter (e.g., an event E and a payment level x, or a weight vector on receipt features). The principal observes realized (y, r) and pays $b_\theta(y, r)$; her per-period realized payoff is $v_s(y) - b_\theta(y, r)$. From the principal's perspective, this is an MDP with unknown transition law and an *endogenous* response: the agent's hidden effort e is a best reply to the contract. A pragmatic model-free approach is to treat the agent as part of the environment and to optimize θ directly for long-run payoff, using standard RL (policy gradients, actor–critic, fitted Q-iteration) on trajectories of (s_t, y_t, r_t).

Pure model-free optimization, however, is not enough for our objectives because it may converge to contracts that earn short-run value by *not* inducing high effort (e.g., exploiting easy tasks), or to contracts that pay for receipts that correlate with outcomes only transiently. We therefore advocate a *constrained* learning formulation that uses the IC structure explicitly: the learning system should search over contracts while maintaining an estimated IC wedge with an explicit margin. This keeps the resulting policy interpretable and makes failure modes diagnosable.

Constrained learning with estimated IC. Fix a state s and a contract family $\{b_{\theta,s}\}_{\theta\in\Theta}$ with limited liability. The one-step IC constraint can be written abstractly as

$$\mathcal{I}_s(\theta) \;:=\; \mathbb{E}[b_{\theta,s}(Y,R) \mid 1, s] - \mathbb{E}[b_{\theta,s}(Y,R) \mid 0, s] - \kappa_s \;\geq\; 0,$$

where κ_s is the required wedge (including continuation effects when relevant). In sequential problems we can impose these constraints in expectation under the principal's visitation distribution. A simple Lagrangian approach is then

$$\max_{\theta\in\Theta} \; \mathbb{E}\left[\sum_t \gamma^t \big(v_{S_t}(Y_t) - b_{\theta,S_t}(Y_t, R_t)\big)\right] \quad \text{s.t.} \quad \mathcal{I}_s(\theta) \geq \xi_s \;\; \forall s, \tag{46}$$

where $\xi_s \geq 0$ is a robustness margin chosen from drift/error considerations. In practice, one can implement (46) with primal–dual updates: the contract parameters θ are updated to increase value net of payments, while dual variables penalize IC violations. The economic benefit of this constrained view is that when learning fails, we can ask *why*: was separability low, was κ_s underestimated, or did drift erode the relevant receipt statistics?

Estimating receipt and outcome responses. To evaluate $\mathcal{I}_s(\theta)$ we need counterfactual expectations under $e = 1$ and $e = 0$, yet e is hidden in deployment. We therefore separate estimation into two regimes.

Controlled calibration. In a sandbox, the principal can often induce exogenous variation in effort by controlling compute budgets, tool permissions, or sampling policies—effectively creating "forced" $e = 0$ and $e = 1$ runs. This yields labeled data $\{(s, y, r, e)\}$ from which we can estimate $\widehat{P}_s(r \mid e)$ and $\widehat{P}_s(y \mid e)$ (or directly estimate gaps $\widehat{P}_s(E \mid 1) - \widehat{P}_s(E \mid 0)$ for candidate events E). For finite receipt spaces, plug-in estimates with empirical Bernstein confidence intervals are natural; for large or continuous receipts, we recommend working with a feature map $\phi(r) \in \mathbb{R}^d$ and estimating a low-dimensional score (e.g., a logistic model for e on $\phi(r)$) whose induced level sets define events.

On-policy updating under strategic response. Post-deployment, the agent's best response makes naive inference hazardous: the distribution of r changes when we change θ. We therefore treat online estimation as a monitoring problem (detecting that previously estimated gaps have shifted) rather than as an attempt to identify $P_s(r \mid e)$ from unlabeled data. When identification is required, we can introduce deliberate randomization: occasionally run an *audit mode* contract that strongly incentivizes one effort level, or impose an operationally enforced compute cap/floor for a small fraction of tasks. These interventions create fresh labeled segments that re-anchor the receipt model.

Practical contract classes: from theory to deployable rules. The extreme-point structure suggests paying on informative events, but a raw event $E \subseteq \mathcal{R}$ is rarely a convenient engineering object. We therefore recommend contract classes that preserve the same logic (pay only when receipts look "high effort") while being implementable and auditable:

1. *Threshold contracts on a receipt score.* Choose a scalar score $\sigma(r)$ (e.g., a log-likelihood ratio proxy, a verifier acceptance score, or a weighted sum of receipt features) and pay $x\mathbf{1}\{\sigma(r) \geq \tau\}$. This recovers the event contract with $E = \{r : \sigma(r) \geq \tau\}$ while allowing τ to be tuned for separability.

2. *Piecewise-linear bonuses.* Pay $b(r) = \min\{x_0 + x_1\sigma(r), \bar{b}\}$ with a cap $\bar{b}$ to control tail risk. This is useful when the principal wants smoother incentives or when receipts are noisy.

3. *Hybrid outcome–receipt contracts.* Pay $b(y, r) = b^y(y) + b^r(r)$ with a small outcome component to discourage pathological "receipt gaming" that passes telemetry but degrades quality.

All three can be parameterized compactly, enabling efficient RL over $\theta = (x, \tau)$ or (x_0, x_1, τ) rather than over arbitrary functions.

A two-phase protocol: calibrate, then validate as a black box. Putting the pieces together, a robust implementation pipeline naturally has two phases.

Phase I (calibration/training). We (i) define the receipt schema and feature map $\phi(r)$; (ii) collect paired datasets under forced low/high effort across representative states; (iii) estimate separability (or a lower confidence bound on Δ_s for candidate events); and (iv) train the principal's contract policy π_θ either by solving per-state LPs inside a planner or via constrained RL with IC margins. The output is not just a contract, but a *model card for incentives*: estimates of which receipts separate effort, how large the gaps are, and where uncertainty is concentrated.

Phase II (black-box validation and rollout). We then freeze the contract policy and evaluate it on a holdout workload where the agent is treated as a black box. The validation goal is not merely higher average y, but *compliance*: do receipt statistics and outcome statistics line up with the calibrated model? We recommend pre-registering acceptance criteria such as (a) realized payment rates, (b) stability of key receipt frequencies, and (c) maintained correlation between the receipt score and downstream quality. Only after passing these checks do we scale traffic.

Monitoring and compliance: what to measure online. Because the contract relies on a probability gap, monitoring should focus on objects that proxy that gap. Concretely, for a deployed threshold contract $x\mathbf{1}\{\sigma(r) \geq \tau\}$, the principal can continuously track the empirical payment indicator $I_t = \mathbf{1}\{\sigma(r_t) \geq \tau\}$ and its conditional versions by state cluster. Large deviations in $\mathbb{E}[I_t \mid s]$ relative to the calibrated baseline indicate drift in receipts *or* a shift in the agent's best response.

Two additional diagnostics help distinguish benign drift from strategic manipulation. First, track the *receipt–outcome link*: if $\sigma(r)$ was designed to be informative about effort, then historically it should predict improved y; a weakening of this relationship is a signature of "hollow receipts." Second, maintain a small stream of *audited tasks* where effort is partially controlled or independently verified (e.g., verified tool traces, reproducible proofs, or duplicated runs). Audits allow the principal to periodically re-estimate $\widehat{P}_s(r \mid e)$ and update η_s-type drift bounds.

Limitations and where additional instruments matter. We should be clear about what this approach does *not* solve. If the state space is extremely large, any per-state estimation of Δ_s is sample-hungry; practical systems will need state aggregation and conservative lower confidence bounds. If receipts are only weakly informative or easily coarsened by privacy constraints, learning cannot manufacture separability, and the feasible contracts revert toward outcome-only subsidies. Finally, if (contrary to our maintained hypothesis) the agent can manipulate receipts directly, then learning "better" receipt models may only accelerate Goodharting; in such settings, audits, cryptographic attestation, and hard compute controls are not optional add-ons but part of the contractible signal design.

The central takeaway is that implementation hinges on a small set of learnable, monitorable statistics—separability of receipts, stability under drift, and continuation wedges—and that we can embed these objects into a model-free control loop without abandoning the economics: the principal learns a contract policy, but remains accountable to explicit IC constraints and explicit robustness margins.

Experimental objectives. Our experimental goal is not to "beat a benchmark" per se, but to test the central comparative statics of the model in an LLM tool-use environment: (i) when receipts are informative about hidden compute/effort, receipt-contingent contracting should reduce the expected subsidy required to elicit high effort relative to outcome-only contracting; (ii) naive pay-for-compute rules (a constant-proportion subsidy) should be less cost-effective than contracts that concentrate payment on the most separating receipt events; and (iii) the advantage of receipts should persist (or fail gracefully) under distribution shift and receipt-model drift. We therefore design experiments around controlled variation in effort, explicit construction of contractible receipt signals, and stress tests that deliberately perturb the mapping from effort to receipts and outcomes.

Benchmarks and a state representation. We instantiate $\mathcal{S}$ using a small set of task clusters that differ in how valuable additional compute and tool depth are. Concretely, we recommend a suite of tool-use benchmarks spanning: (a) multi-hop web search and synthesis; (b) code generation with unit tests; (c) mathematical reasoning with external calculators/provers; and (d) data-wrangling tasks with a sandboxed interpreter. Each task instance induces a state s consisting of coarse, contractible descriptors: task family, input length bucket, time budget bucket, and (when relevant) tool availability. This "state" is not intended to be a full Markov description; it is an engineering device that allows us to estimate separability and contracts at a manageable granularity while still testing whether the economics survives heterogeneous workloads.

Operationalizing effort. To create a ground-truth analogue of $e \in \{0, 1\}$, we use operational controls that are available in many LLM deployments: compute caps/floors (token budgets, wall-clock limits), enforced tool-use policies (maximum number of tool calls, enforced minimum depth), and search parameters (beam width, number of self-consistency samples, number of retrieved documents). High effort $e = 1$ is implemented as a configuration with strictly higher allowed compute/tool intensity; low effort $e = 0$ is the restricted configuration. Importantly, we keep the base model weights fixed within each experimental run so that the induced distributions $P_s(y, r \mid e)$ are attributable to compute/tool differences rather than model identity.

Receipts as contractible signals. We define $\mathcal{R}$ as a collection of verifiable telemetry objects emitted by the runtime: tool-call traces (timestamps, tool names, inputs/outputs hashes), retrieval metadata (number of retrieved documents, diversity statistics), execution artifacts (unit test logs, interpreter transcripts, proof checker acceptance), and coarse compute meters (tokens generated, latency). To reflect practice, we evaluate two receipt schemas. A *rich* schema logs fine-grained tool traces and verifier outcomes; a *coarsened* schema logs only aggregate counts and pass/fail verifications (motivated by privacy and storage constraints). For each schema we predefine a feature map $\phi(r)$ and a scalar score $\sigma(r)$ used by threshold contracts; in a controlled calibration phase (where e is known), we estimate receipt separability via an empirical approximation to

$$\Delta_s \;=\; \sup_{\tau} \Big(\mathbb{P}(\sigma(R) \geq \tau \mid 1, s) - \mathbb{P}(\sigma(R) \geq \tau \mid 0, s)\Big),$$

and we record the approximately optimal event $E_s^* = \{\sigma(r) \geq \tau_s^*\}$ achieving the largest gap within the contract class.

Contract regimes compared. We compare three families that map naturally onto procurement-like payment rules.

1. *Outcome-only contracts.* Payments depend only on benchmark-scored outcomes y (e.g., pass/fail, normalized score). We include (i) a threshold bonus $b(y) = x\mathbf{1}\{y \geq \bar{y}\}$ and (ii) a linear score bonus $b(y) = x \cdot y$ with $b \geq 0$. These contracts are attractive operationally but should struggle when Δ_s^y is small.

2. *Receipt-contingent contracts.* Payments depend on receipt events (or hybrid $b(y, r) = b^y(y) + b^r(r)$ with a small b^y term fixed across conditions). The primary specification is the event bonus $b(r) = x\mathbf{1}\{r \in E_s^*\}$, where E_s^* is chosen from the calibrated class (e.g., a threshold on $\sigma(r)$). We also include a capped piecewise-linear variant to test sensitivity to tail receipts.

3. *Constant-proportion subsidy (pay-for-metrics).* Payments are proportional to a compute proxy, such as tokens or tool-call count: $b(r) = \alpha \cdot m(r)$ with $m(r) \geq 0$. This mirrors common "pay per token/tool" pricing and provides a baseline for how much is lost by ignoring separability and concentrating payments.

Within each regime we tune payment levels to target a common implementation objective (e.g., induce $e = 1$ in the calibration environment at minimal expected payment, subject to limited liability).

Evaluation protocol: calibration, then strategic rollout. We run a two-stage evaluation aligned with the model's identification needs. In *calibration*, we execute forced-e runs to estimate $\widehat{P}_s(r \mid e)$ and $\widehat{P}_s(y \mid e)$ and to select contract parameters (event thresholds, slopes) using only training tasks. We compute an estimated incentive gap for each candidate contract,

$$\widehat{\mathcal{I}}_s \;=\; \widehat{\mathbb{E}}[b(Y,R) \mid 1, s] - \widehat{\mathbb{E}}[b(Y,R) \mid 0, s] - \Delta c_s,$$

and add a robustness margin ξ_s when choosing parameters (implemented as requiring $\widehat{\mathcal{I}}_s \geq \xi_s$). In *rollout*, we evaluate on held-out tasks where the agent is free to choose behavior (to the extent the platform permits) and we treat the system as a black box, measuring realized outcomes, receipts, and spend. When true strategic response is hard to elicit in a single-model setup, we approximate it with a "mixture agent" that selects among available compute configurations based on the realized expected payment (a reduced-form best response calibrated from historical data).

Distribution-shift and drift stress tests. To test robustness, we introduce shifts that alter either $P_s(y \mid e)$, $P_s(r \mid e)$, or both. (i) *Task shift:* new domains (e.g., different websites, new coding libraries), longer contexts, adversarially composed tasks, and changed tool availability. This probes whether a contract tuned on one workload generalizes. (ii) *Receipt drift:* logging changes (coarsening, delayed tool events), verifier version updates, and added latency noise. This directly maps to an increased η_s and should differentially harm receipt contracts. (iii) *Effort-to-receipt decoupling:* configurations where high effort yields diminishing receipt differences (e.g., additional internal sampling without extra tool calls). This creates low Δ_s even if compute increases, and should expose when receipt contracting loses leverage. For each shift we re-run rollout without retuning to measure performance degradation, and we separately evaluate a "nudged" variant that increases margins ξ_s (hence payments) to maintain IC under drift.

Metrics: welfare, spend, compliance, and robustness. We report four primary metrics, each computed per state cluster and aggregated by

the empirical state visitation distribution. *Welfare (principal value net payments):*

$$\widehat{U}^P \;=\; \frac{1}{n}\sum_{i=1}^{n}\big(v_{s_i}(y_i) - b_{s_i}(y_i, r_i)\big).$$

Spend and cost-effectiveness: average payment, payment conditional on success, and the empirical analogue of "subsidy per induced effort" using forced-e counterfactuals from calibration. *Compliance/IC diagnostics:* the realized frequency of the paid event (e.g., $\mathbf{1}\{r \in E^*\}$), stability of receipt statistics relative to calibration baselines, and (when audits are available) the measured gap $\widehat{\mathbb{P}}(E^* \mid e = 1) - \widehat{\mathbb{P}}(E^* \mid e = 0)$ over time. *Robustness under shift:* worst-case (or tail) degradation in $\widehat{U}^P$ and increased spend across the stress tests, together with a decomposition into (a) receipt drift (changes in receipt frequencies) versus (b) hollowing (weakened relationship between receipt score and outcome).

Ablations and limitations. To connect outcomes tightly to the theory, we include ablations that manipulate separability directly: coarsen receipts to reduce Δ_s, remove tool-trace features from $\sigma(r)$, and restrict the event class to suboptimal thresholds. We also vary Δc_s by changing the compute gap between $e = 0$ and $e = 1$, testing whether required payments scale linearly as predicted. Finally, we emphasize a limitation: without a genuinely strategic agent that can adapt its internal policies to the contract, some Goodharting channels will be underexplored. For that reason, the most informative experimental signal is not merely higher benchmark scores, but whether receipt-contingent contracts deliver the predicted *payment–separability tradeoff* across environments and whether this tradeoff breaks precisely when separability erodes.

4 Extensions and policy discussion.

The model is intentionally spare: a single principal, a single agent, binary hidden effort, and verifiable receipts that shift in distribution with compute. That sparseness is a feature for comparative statics, but it also tells us exactly where practical deployments will strain the assumptions. In this section we discuss four extensions that matter for real procurement and platform governance—manipulation-resistant receipts, privacy constraints on telemetry, competition across principals, and multi-agent coordination—and we close with limitations and a concrete future-work agenda.

Manipulation-resistant receipts and the boundary of (H1). The key institutional assumption behind receipt contracting is (H1): the agent cannot choose r directly except through the costly effort margin. In practice,

this is not a metaphysical claim; it is an engineering and governance target. If the agent can cheaply inflate a receipt (e.g., generate spurious tool calls, pad token counts, or selectively log favorable traces), then Δ_s becomes a poor proxy for true effort and the event E_s^* can be Goodharted. The theory suggests a useful diagnostic: receipt contracting is only as good as the *effort elasticity* of the log distribution. When manipulation channels open, the effective separability collapses, $\Delta_s \downarrow 0$, and the minimal-implementation bound $\Delta c_s/\Delta_s$ diverges.

Two design patterns help restore an (approximate) (H1). First, *attestation and provenance*: receipts should be emitted by components the agent cannot rewrite, such as a platform-controlled tool router, a trusted execution environment, or a cryptographic transcript with hash chaining and server-side timestamps. Second, *verifier-coupled receipts*: instead of paying for raw activity (calls, tokens), we pay for events that are hard to generate without genuine work, such as unit tests passing, proof checker acceptance, or consistency checks across multiple independent tools. Economically, these constructions aim to ensure that, for any cheap "receipt inflation" strategy, the induced distribution $P_s(r \mid e)$ either changes negligibly or triggers detectable anomalies that reduce expected payment. A complementary governance tool is randomized auditing: if the platform occasionally re-executes tool calls or requests reproduction seeds, then receipt manipulation becomes a mixed strategy problem with expected penalties (or forfeited bonuses) even under limited liability.

Privacy constraints and the price of coarsening. Receipt richness is rarely free. Telemetry can contain sensitive user content, proprietary retrieval results, or security-relevant traces. Platforms therefore coarsen, aggregate, or privatize logs, and this interacts mechanically with incentives: any privacy mechanism that reduces information about e reduces Δ_s. The comparative static is immediate: holding Δc_s fixed, any privacy-induced reduction in separability raises the subsidy required to induce high effort.

This point can be made operational. If a privacy filter applies a randomized mechanism Π to raw receipts R to produce released receipts $\widetilde{R} = \Pi(R)$, then the relevant separability becomes

$$\widetilde{\Delta}_s \;=\; \mathrm{TV}\big(P_s(\widetilde{r} \mid 1),\, P_s(\widetilde{r} \mid 0)\big),$$

and data processing implies $\widetilde{\Delta}_s \le \Delta_s$. For example, if Π satisfies an ε-differential-privacy constraint at the receipt level, then the likelihood ratio between $P_s(\widetilde{r} \mid 1)$ and $P_s(\widetilde{r} \mid 0)$ is bounded, which in turn upper-bounds total variation and can force $\widetilde{\Delta}_s$ to be small unless ε is permissive. This creates an explicit policy frontier: stronger privacy (smaller ε or heavier coarsening) reduces incentive leverage, and the system must either tolerate lower effort or pay more to sustain it.

A practical compromise is to separate *content* from *effort evidence*. We can often retain manipulation-resistant, privacy-preserving receipts by logging only (i) tool identities and counts, (ii) cryptographic commitments to tool I/O (hashes) without raw content, and (iii) pass/fail verifier outcomes. Where even that is too revealing, we can use secure aggregation across tasks or delayed release, but then robustness margins ξ_s become more important: privacy mechanisms can introduce additional drift and noise, effectively increasing η_s and requiring larger nudges to preserve incentive compatibility.

Multi-principal competition and "incentive arbitrage." Many LLM deployments occur in markets with multiple principals: enterprise customers, platforms, and intermediaries simultaneously purchase model outputs and can each impose a contract. A simple extension indexes principals by $j \in \{1, \dots, J\}$ with contracts $b_s^{(j)}(y, r)$, so the agent internalizes total expected payment $\sum_j \mathbb{E}[b_s^{(j)}(Y, R) \mid e, s]$ net of effort cost. Competition can cut both ways. On one hand, it can discipline rents: if receipts allow precise incentive targeting, principals may achieve high effort with lower expected transfers, and competitive pressure pushes contracts toward minimal-implementation forms. On the other hand, competition can create *externalities in telemetry*: if one principal pays on a receipt event that is easy to inflate, the agent may reshape behavior in ways that degrade other principals' outcomes, and the relevant $P_s(y, r \mid e)$ becomes endogenous to the portfolio of contracts.

This suggests a platform-governance role for standardization. If the platform defines canonical receipt schemas, audit rules, and manipulation penalties, then principals compete on thresholds and bonus levels rather than on incompatible measurement regimes. Moreover, without coordination, competition may induce a "race to the simplest metric" (e.g., tokens) because it is easiest to specify and verify, even when it is a poor separator of true effort. The model clarifies why that is inefficient: proportional pay-for-metrics typically fails to concentrate payments on E_s^* and therefore wastes budget whenever the metric is only weakly correlated with effort.

Multi-agent systems and coordination across components. Modern tool-using systems are often modular: a planner, retriever, coder, and verifier may be separate agents or services, each with its own effort choice and cost. Let effort be a vector $e = (e_i)_{i=1}^N$ with costs $\sum_i c_{s,i}(e_i)$ and receipts $r = (r_i)_{i=1}^N$ that may be individually attributable. The principal then faces a joint implementation problem: induce high effort in the components that matter, while avoiding overpaying those whose receipts do not separate.

Two economic frictions emerge. First is *free-riding*: if outcome y is a joint product, outcome-only bonuses induce each component to rely on others. Receipts mitigate this by enabling component-specific incentives when $\Delta_{s,i} := \mathrm{TV}(P_s(r_i \mid e_i = 1), P_s(r_i \mid e_i = 0))$ is large. Second is *attribu-*

tion: when receipts are coupled (e.g., planner choices affect retriever logs), a naive per-agent event contract may create perverse substitution across modules. The natural design response is to define receipts that are both attributable and verifiable (e.g., per-module test suites, provenance tags for retrieved sources, or sandboxed execution logs) and to use hybrid contracts that pay on module-specific events while retaining a small outcome term to align system-level objectives.

Technically, the minimal-payment logic generalizes: the principal solves an LP over nonnegative payments on the joint receipt space, and extreme-point solutions again concentrate payment on separating events—but now the event can be a *pattern* across modules. This highlights a governance benefit of modular receipts: if we can maintain approximate conditional independence of r_i given e_i, then incentive design decomposes and avoids combinatorial explosion in $\mathcal{R}$.

Implications for procurement and platform governance. Receipt contracting reframes AI procurement from "pay for outcomes" toward "pay for verified process evidence," with clear guardrails. First, procurement should begin with an *informativeness audit*: estimate Δ_s (and its drift) for candidate receipts and reject metrics that do not separate effort. Second, contracts should be written to minimize Goodhart pressure: concentrate payments on hard-to-fake events, cap exposure to tail receipts, and include periodic recalibration of E_s^* as tools, models, and workloads change. Third, platforms should treat telemetry standards as market infrastructure. A well-defined receipt API, verifiability guarantees, and privacy-preserving attestations expand the feasible set of contracts and reduce the need for blunt, expensive outcome subsidies.

From a policy perspective, the model also suggests a narrow and testable case for transparency mandates: not "full logging," but *auditable evidence of effort-linked events* that improves separability without revealing sensitive content. Conversely, overly restrictive telemetry rules can backfire by forcing contracting onto outcomes alone, which is precisely the regime where the required subsidies can become arbitrarily large when Δ_s^y is small.

Limitations and future work. We close by being explicit about what the model does not yet capture. Risk neutrality and limited liability are useful baselines, but many vendors are risk averse and may demand insurance-like premia for volatile receipt bonuses. Binary effort is a simplification; real systems allocate compute continuously and choose among many tool policies. Our treatment of drift via η_s is deliberately reduced-form; in practice, drift can be strategic (agents adapt to the contract) and adversarial (agents search for receipt loopholes). Finally, we have assumed commitment to contracts and verifiability of receipts; where either fails, repeated-game reputation and

audit enforcement become central.

These limitations point to concrete extensions: (i) continuous effort and multi-dimensional actions (e, a) with endogenous receipt design; (ii) online learning of contracts with exploration costs and adversarial robustness; (iii) equilibrium analysis under multi-principal competition with shared telemetry standards; and (iv) mechanism design for multi-agent systems with attribution constraints and collusion. Our view is that the model is most valuable precisely because it sharpens these questions: it turns vague debates about "paying for compute" into measurable objects—Δ_s, Δ_s^y, and η_s—that can be estimated, stress-tested, and governed.