TRUTHFUL ONLINE PROCUREMENT FROM MULTI-TIER LLM PROVIDERS UNDER BUDGET CONSTRAINTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Procuring services from external large language model (LLM) providers allows AI platforms to exploit quality–cost trade-offs across service tiers. Doing so under a hard budget requires both eliciting providers' private tier-specific costs and learning their unknown request-dependent qualities. We formulate this joint problem as repeated procurement with vector-valued private costs, contextual quality feedback, and per-round payment budgets. For known qualities, we propose TRIP, a deterministic mechanism that combines bid-independent posted menus with leave-one-out payment caps to ensure truthfulness, individual rationality, and budget feasibility. We establish an approximation certificate under explicit conditions on realized budget utilization and winner payment rates. For unknown qualities, we develop TRIP-MC, which learns from contextual semi-bandit feedback and supplies TRIP with optimistic quality estimates that respect tier ordering. The online mechanism preserves history-conditioned, current-round truthfulness and per-round budget feasibility. Under a linear contextual model, we prove sublinear contextual γ-regret conditional on a per-round γ-approximate oracle, and using TRIP as this oracle requires its certificate conditions to hold in the compared rounds. Experiments in procurement markets constructed from MMLU rollout traces show gains in procurement value and cumulative reward over truthful baselines.

1 INTRODUCTION

Modern AI applications increasingly assemble capabilities from external model and agent services rather than rely on a single in-house model. Model routers and agent orchestrators may query several third-party providers for decomposition, fallback, verification, or parallel execution Feng et al. (2026). They also reserve advanced service tiers (e.g., larger context windows, additional test-time compute, stronger reasoning capabilities, or richer tool use) for requests where the expected quality gain justifies the additional expense Geiping et al. (2026). This pattern turns model selection into a budget-constrained procurement problem: the platform must acquire heterogeneous AI capabilities while respecting a hard budget. This heterogeneity is also an opportunity. Always invoking the strongest tier can spend scarce resources on requests that cheaper alternatives could handle, whereas always choosing the cheapest tier forgoes quality on requests that benefit from additional capability. Effective procurement can instead exploit differences in quality and cost across provider-tier pairs, reserving expensive services for requests whose marginal quality gains justify their additional cost and thereby extracting more value from a limited budget Poon et al. (2026).

Realizing this opportunity is difficult because a quoted price is neither a provider's true cost nor a guarantee of quality. Each provider has private, tier-specific operating costs and may strategically misreport them. Because a provider can offer several tiers, its private type is a cost vector rather than a scalar bid, creating a multi-parameter mechanism-design problem Weinberg & Zhou (2022). At the same time, the platform does not know how well each provider–tier pair will perform on a particular request and must learn its contextual quality from realized outcomes Gu et al. (2026). Exploration therefore cannot be treated as a purely statistical exercise: it changes allocations and payments, and can compromise incentive guarantees or exhaust the available budget. The platform must learn which tiers are valuable while eliciting truthful costs and enforcing a spending limit.

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.

Existing approaches capture only subsets of this problem. Budget-feasible procurement handles strategic costs and hard payment constraints, but typically assumes known values, static allocations, and scalar seller types Singer (2010). Standard bandit methods learn unknown rewards but treat costs as public problem data Badanidiyuru et al. (2018), while cost-aware LLM routing similarly optimizes against fixed API prices Poon et al. (2026); Shirkavand et al. (2026). Taken together, these tools do not provide an online mechanism that learns contextual quality, elicits truthful tier-specific cost vectors, and keeps realized payments within the budget in every round.

We first study offline multi-tier procurement with known tier qualities and design TRIP, a deterministic, truthful, and budget-feasible mechanism with a conditional approximation certificate. We then consider unknown contextual tier qualities and develop TRIP-MC, which learns from semi-bandit feedback and invokes TRIP using optimistic estimates. The resulting policy preserves conditional current-round incentive and budget guarantees; its sublinear-regret result holds under an explicit per-round approximate-oracle assumption. Our main contributions are as follows:

(1) We formulate a repeated procurement problem for multi-tier LLM services that jointly addresses private vector-valued costs, unknown contextual qualities, and stringent per-round payment budgets. This formulation bridges multi-parameter truthful procurement with contextual online learning.

(2) We design TRIP, a deterministic offline mechanism that employs bid-independent posted menus and leave-one-out payment caps to ensure truthfulness, individual rationality, and budget feasibility. We establish bounds on the optimal procurement value and provide an approximation certificate under specific conditions regarding realized budget utilization and winner payment rates.

(3) We develop TRIP-MC, which integrates TRIP with optimistic quality estimates that adhere to tier ordering and are derived from contextual semi-bandit feedback. This mechanism ensures history-conditioned, current-round truthfulness and per-round budget feasibility. Under a linear contextual model and a per-round γ-approximate oracle assumption, we prove sublinear contextual γ-regret.

(4) We evaluate both mechanisms using trace-backed procurement markets constructed from MMLU rollouts, comparing procurement value, cumulative reward, and realized budget utilization against truthful baselines and a non-truthful evaluation ceiling.

Related Work *Budget feasibility.* Budget-feasible reverse auctions maximize a known valuation while keeping payments to strategic sellers within a fixed budget Singer (2010); Liu et al. (2024). Most mechanisms and approximation results assume a single-parameter seller that reports one cost Bei et al. (2012); Chen et al. (2011); Gravin et al. (2020). Multi-unit extensions permit several identical units with a common per-unit cost Chan & Chen (2014); Wu et al. (2019), whereas our provider offers mutually exclusive, heterogeneous tiers with a private cost vector. These mechanisms are typically static and assume known values; our setting must learn contextual provider–tier quality.

Online learning with strategic agents. Truthful bandit mechanisms have been studied in settings such as sponsored search with unknown click-through rates and scalar private values Babaioff et al. (2014). Bandits with knapsacks impose resource budgets but generally treat consumption as non-strategic Badanidiyuru et al. (2018); Immorlica et al. (2022). Closer procurement-oriented work learns worker quality while eliciting a scalar cost, or learns multi-unit supply with a per-unit cost and capacity Biswas et al. (2015); Bhat et al. (2019); Liu et al. (2025). These models do not capture a provider that strategically reports costs over heterogeneous tiers, nor the interaction between contextual tier learning and a hard payment budget.

Cost-aware LLM routing. LLM routers and cascades improve query-dependent quality–cost trade-offs by selecting among models or APIs Chen et al. (2023); Poon et al. (2026); Xia et al. (2024); Ding et al. (2024); Hu et al. (2024); Pan et al. (2026). They establish the value of adaptive model selection, but treat endpoints and API prices as exogenous. They do not model a provider's privately costed tier menu, design payments to induce truthful reports, or distinguish a hard mechanism-payment budget from an inference-cost objective. A full discussion appears in Appendix A.

2 Problem Formulation

As illustrated in Figure 1, we study dynamic procurement from strategic LLM providers under per-round budget constraints over T rounds. In round $t \in \{1, \dots, T\}$, the platform receives a buyer

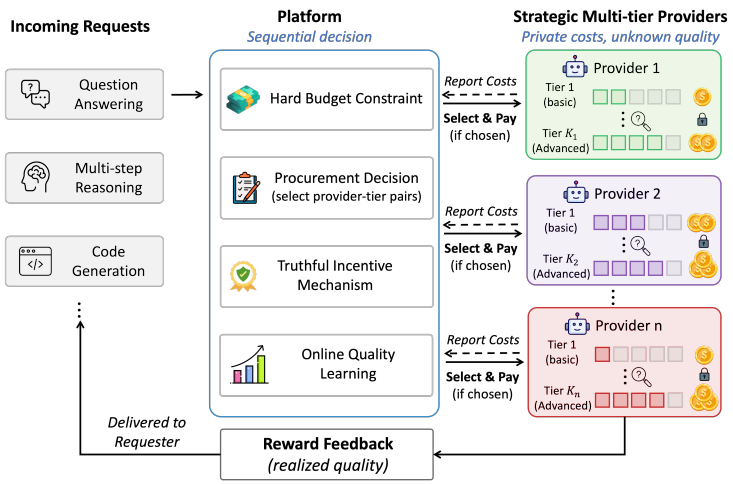


Figure 1: Overview of sequential procurement from multi-tier LLM providers.

request, equivalently a procurement instance, with budget B_t and interacts with $[n]$ providers. Each provider i offers a finite set G_i of service tiers (computational configurations), and the platform may procure at most one tier from that provider in each round.

Private costs and bids. Provider i incurs a private cost $c_i(\theta)$ when delivering tier $\theta \in G_i$. In each round t, providers submit bids $b_{i,t}(\theta)$, which may differ from their costs because of strategic behavior. Let $\mathbf{b}_{i,t} = (b_{i,t}(\theta))_{\theta \in G_i}$ be the vector of bids for provider i at round t. To rule out an unbounded report space and to capture the finite granularity of provider contracts, we work with a bounded cost and bid domain. Specifically, there is a known constant $\hat{c} > 0$ such that $0 \le c_i(\theta) \le \hat{c}$ and $0 \le b_{i,t}(\theta) \le \hat{c}$ for every provider i, tier $\theta \in G_i$, and round t. Thus, each provider–tier contract has bounded cost, and providers report within a known admissible range. The tier sets, contexts, and reward observations are nonstrategic; only the cost vector is privately reported. In particular, the online feature map and the quality estimate supplied to the procurement rule are independent of current-round bids.

Contextual and unknown quality. We consider a contextual stochastic procurement environment over T rounds. Before procurement in round t, the platform observes a context (task feature) $f_t \in \mathcal{F}$ that describes the current buyer request. Selecting arm $a = (i, \theta)$, where i is a provider and $\theta \in G_i$ is one of its tiers, yields an arm-level stochastic reward $r_{t,a} \in [0, 1]$ with unknown conditional mean $\mu_{i,\theta}(f_t) = \mathbb{E}[r_{t,a} \mid a = (i, \theta), f_t]$. The platform does not know $\mu_{i,\theta}(\cdot)$ a priori and must learn these context-dependent qualities online from bandit feedback observed after procurement. For each provider i, index the tiers by increasing service level as $\theta_{i,1}, \ldots, \theta_{i,L_i}$. We assume that a higher tier has weakly greater expected value under every context:

$$\mu_{i,\theta_{i,1}}(f) \le \mu_{i,\theta_{i,2}}(f) \le \cdots \le \mu_{i,\theta_{i,L_i}}(f), \forall f \in \mathcal{F}. \tag{1}$$

This monotonicity reflects the semantics of service tiers: a higher tier represents a weakly stronger computational configuration, such as a more capable model, greater compute, or stricter service guarantees. It is therefore natural to model its expected utility contribution for a fixed request as no smaller than that of a lower tier.

Allocation, payments, and budget. Given the submitted bids and the realized context f_t, the mechanism selects a feasible set S_t of provider–tier pairs, with at most one tier selected from each provider, and specifies nonnegative payments $\{p_i(t)\}_{i=1}^n$, with $p_i(t) = 0$ for every unselected provider. The resulting total payment is $P_t = \sum_{i=1}^n p_i(t)$. The mechanism must satisfy per-round budget feasibility: $P_t \le B_t$.

Objective. The platform aims to maximize cumulative expected contextual value over the sequence of buyer requests, subject to the per-round budget constraints: $\max \mathbb{E}\Big[\sum_{t=1}^T v(S_t, f_t)\Big]$, s.t. $\forall t \in \{1, \ldots, T\}$, $P_t \le B_t$, where we focus throughout on the additive objective $v(S_t, f_t) = \sum_{(i,\theta) \in S_t} \mu_{i,\theta}(f_t)$, which is the objective used by the mechanisms and all subsequent proofs.

Algorithm 1 T-BF: BF-based threshold payments

Require: Providers $[n]$, tier sets $\{G_i\}_{i=1}^n$, bids $\{b_i(\theta)\}$, mean rewards $\{\mu_{i,\theta}\}$, budget B
Ensure: Payments $\{p_i(\theta)\}_{i\in[n],\theta\in G_i}$ and leave-one-out cutoffs $\{m_{-i}\}_{i\in[n]}$
1: Compute $(\theta_{i,\min}, \rho_i)$ for all $i \in [n]$
2: **for** each provider $i \in [n]$ **do**
3: Obtain the leave-one-out BF order π_{-i} and cutoff m_{-i} under budget B
4: **for** each tier $\theta \in G_i$ **do**
5: Compute $p_i(\theta)$ using the threshold rule in Eq. (6)
6: **return** $\big(\{p_i(\theta)\}_{i\in[n],\theta\in G_i}, \{m_{-i}\}_{i\in[n]}\big)$

Design goal. We seek a mechanism $\mathcal{M} = (\mathcal{A}, \{p_i\}_{i=1}^n)$ that satisfies (i) budget feasibility, (ii) individual rationality, and (iii) truthfulness, while achieving (iv) sublinear regret relative to a tractable *contextual approximate benchmark*. In round t, let $x_i(t) \in \{0,1\}$ indicate whether provider i is selected, and let $\theta_i(t)$ denote its allocated tier when $x_i(t) = 1$. The resulting utility of provider i is $u_i(t) = x_i(t)(p_i(t) - c_i(\theta_i(t)))$. We require the following properties to hold for each round t:

$$\text{(Budget feasibility)} \qquad P_t = \sum_{i=1}^{n} p_i(t) \le B_t. \tag{2}$$

$$\text{(Individual rationality)} \qquad u_i(t) \ge 0, \qquad \forall i \in [n]. \tag{3}$$

$$\text{(Truthfulness)} \qquad \mathbb{E}\big[u_i(t; c_i(\cdot), \mathbf{b}_{-i,t}) \mid \mathcal{H}_{t-1}, f_t\big] \ge \mathbb{E}\big[u_i(t; b'_{i,t}(\cdot), \mathbf{b}_{-i,t}) \mid \mathcal{H}_{t-1}, f_t\big], \tag{4}$$

for every provider i, every admissible alternative report $b'_{i,t}(\cdot)$, and every realized history $\mathcal{H}_{t-1}$ and context f_t; the expectation is over the mechanism's internal randomness, when applicable. Because the per-round offline procurement problem remains NP-hard under a knapsack-type budget even after f_t is observed, the online analysis is stated for an oracle with approximation factor $\gamma \in (0,1]$. For TRIP, such a factor is certified only on rounds satisfying the rate and realized-utilization conditions of Theorem 2; absent those conditions, the online result is an oracle-conditional guarantee rather than an unconditional guarantee for TRIP-MC.

Let $\mathrm{OPT}_t(f_t)$ denote the optimal value among allocations selecting at most one tier per provider and having total truthful cost at most B_t, assuming the conditional means $\{\mu_{i,\theta}(f_t)\}$ are known. Define $\mathrm{OPT}_t^{(\gamma)}(f_t) = \gamma\,\mathrm{OPT}_t(f_t)$ as the corresponding γ-approximate benchmark. Our goal is sublinear expected contextual γ-regret: $R_\gamma(T) = \mathbb{E}\left[\sum_{t=1}^{T}\left(\mathrm{OPT}_t^{(\gamma)}(f_t) - v(S_t, f_t)\right)\right]$, where the expectation is taken over the context process and, when applicable, the mechanism's internal randomness. Reward noise affects this pseudo-regret through the learned history, while $v(S_t, f_t)$ itself denotes conditional mean value.

3 Offline Benchmark Mechanism

We first consider the offline setting, in which the platform observes the realized context and knows the corresponding expected rewards $\mu_{i,\theta}(f_t)$. Let B denote the budget available for a single procurement round. The online mechanism introduced later uses this offline mechanism as its allocation oracle. Because the analysis concerns one round and the mechanism depends on the context only through the induced reward values, we suppress the context argument and write $\mu_{i,\theta}(f_t)$ as $\mu_{i,\theta}$ throughout this section. Tiers with $\mu_{i,\theta} = 0$ are discarded before running the mechanism: with nonnegative costs they add no welfare, and the ratio rules below are undefined for them. Providers with no positive-value tier are likewise inactive. Hence all ratios in this section are taken over $\mu_{i,\theta} > 0$.

We begin with T-BF, a sub-mechanism used by the main mechanism. Given the reported bids and expected rewards, T-BF computes a threshold payment $p_i(\theta)$ for every provider–tier pair. The leave-one-out construction of these payments is the basis for the incentive properties of TRIP.

Sub-mechanism T-BF For each provider $i \in [n]$, let $\theta_{i,\min} \in \arg\min_{\theta\in G_i} \frac{b_i(\theta)}{\mu_{i,\theta}}$ denote a tier with the smallest bid-to-reward ratio.

Algorithm 2 TRIP: Main offline mechanism

Require: Providers $[n]$, tier sets $\{G_i\}_{i=1}^n$, bids $\{b_i(\theta)\}$, mean rewards $\{\mu_{i,\theta}\}$, total budget B
Ensure: Winners W, selected tiers $\{\theta_i^\star\}_{i\in W}$, payments $\{p_i^\star\}_{i\in W}$
Compute $(\theta_{i,\min}, \rho_i)$, the BF cutoff k, and $C \leftarrow \{1, \dots, \min\{k+1, n\}\}$
$(\{p_i(\theta)\}, \{m_{-i}\}) \leftarrow$ T-BF$(\{b_i(\theta)\}, \{\mu_{i,\theta}\}, B)$
$W \leftarrow \varnothing$ and $\pi \leftarrow$ PUBLICORDER(C)
for each provider i in the order π **do**
    Set $q_i \leftarrow B/(m_{-i}+2)$ and construct the selected cap menu $\{\widehat{p}_i^{(v)}(\theta)\}_{\theta\in G_i}$ by Eq. (7)
    Form $\mathcal{A}_i^{\mathrm{LOO},(v)}$ by Eq. (8)
    if $\mathcal{A}_i^{\mathrm{LOO},(v)} \neq \varnothing$ **then**
        Select $\theta_i^\star$ by Eq. (9)
        if $\widehat{p}_i^{(v)}(\theta_i^\star) - b_i(\theta_i^\star) \geq 0$ **then**
            Add i to W and set $p_i^\star \leftarrow \widehat{p}_i^{(v)}(\theta_i^\star)$
return $(W, \{\theta_i^\star\}_{i\in W}, \{p_i^\star\}_{i\in W})$

Baseline Budget Feasible (BF) Allocation. Let $\rho_i = \frac{b_i(\theta_{i,\min})}{\mu_{i,\theta_{i,\min}}}$ denote provider i's minimum bid-to-reward ratio. Sort the providers in nondecreasing order of ρ_i, break ties by a fixed public order, and relabel them so that $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_n$. We define the *BF cutoff index*, k, as the largest index satisfying

$$\rho_k \;\leq\; \frac{B}{\sum_{i=1}^{k} \mu_{i,\theta_{i,\min}}}, \quad \rho_{k+1} \;>\; \frac{B}{\sum_{i=1}^{k+1} \mu_{i,\theta_{i,\min}}}. \tag{5}$$

When $k = n$, the second inequality is omitted. If the first provider fails the inequality, we set $k = 0$ and the BF prefix is empty. The baseline BF allocation identifies the first k providers and their minimum-ratio tiers, $\{(i, \theta_{i,\min})\}_{i=1}^k$, as candidate, or virtual-winning, provider–tier pairs.

Payment Computation in T-BF. Fix a provider i and a tier $\theta \in G_i$. To compute the payment for this pair, remove i and sort the remaining providers by their ratios ρ_j, yielding the sequence $\pi_{-i}(1), \dots, \pi_{-i}(n-1)$; here, $\pi_{-i}(j)$ is the jth provider in this order. Applying the BF allocation rule to this reduced market under budget B gives a leave-one-out cutoff index m_{-i}. Equivalently, m_{-i} is the largest prefix length that satisfies the condition in Eq. (5) after provider i has been removed. We use the convention $m_{-i} = 0$ when no nonempty prefix is feasible. If $m_{-i} = n-1$, the index $j = m_{-i}+1 = n$ below refers to a dummy boundary competitor with ratio $+\infty$; the prefix sum then contains all $n-1$ real competitors. This convention also covers a one-provider market. For each $j \in \{1, \dots, m_{-i}+1\}$, define two critical values. The first, $b_{i,j}(\theta) = \mu_{i,\theta} \cdot b_{\pi_{-i}(j)}(\theta_{\pi_{-i}(j),\min}) / \mu_{\pi_{-i}(j),\theta_{\pi_{-i}(j),\min}}$, scales the ratio of the jth competing provider to tier θ of provider i. The second is $\beta_{i,j}(\theta) = \mu_{i,\theta} \cdot B / (\sum_{\ell=1}^{j-1} \mu_{\pi_{-i}(\ell),\theta_{\pi_{-i}(\ell),\min}} + \mu_{i,\theta})$. The threshold payment for tier θ is the largest of the resulting critical values:

$$p_i(\theta) \;=\; \max_{j\in\{1,\dots,m_{-i}+1\}} \min\left\{ b_{i,j}(\theta),\, \beta_{i,j}(\theta) \right\}. \tag{6}$$

The Main Mechanism (TRIP). Algorithm 2 uses the full budget to construct a candidate pool comprising the first $\min\{k+1, n\}$ providers in the BF order. For each candidate $i \in C$, it first invokes T-BF to obtain the leave-one-out cutoff m_{-i} and payment menu $\{p_i(\theta)\}_{\theta\in G_i}$. It assigns the leave-one-out cap $q_i := \frac{B}{m_{-i}+2}$ and converts the threshold menu to one of three common posted-menu variants. Write $\lambda_i^0 := \min_{\theta\in G_i} \frac{p_i(\theta)}{\mu_{i,\theta}}$. The variants share the same candidate pool, leave-one-out thresholds, and cap q_i. Define the public maximum $M_i := \max_{\vartheta\in G_i} \mu_{i,\vartheta}$.

$$\begin{aligned} &\widehat{p}_i^{(A)}(\theta) := \lambda_i^{(A)} \mu_{i,\theta}, && \lambda_i^{(A)} := \min\left\{ \lambda_i^0, \frac{q_i}{M_i} \right\}, \\ &\widehat{p}_i^{(B)}(\theta) := \min\{p_i(\theta), q_i\}, && \widehat{p}_i^{(C)}(\theta) := \min\{\lambda_i^0 \mu_{i,\theta}, q_i\}. \end{aligned} \tag{7}$$

Variant A is the current proportional menu, variant B truncates each threshold payment at the cap, and variant C applies the cap tier-by-tier to the uncapped proportional menu. The admissible set for

the selected variant $v \in \{A, B, C\}$ is

$$\mathcal{A}_i^{\mathrm{LOO},(v)} = \{\theta \in G_i : \widehat{p}_i^{(v)}(\theta) \le q_i\}. \tag{8}$$

All three menus satisfy $\widehat{p}_i^{(v)}(\theta) \le q_i$, so the admissible set contains every active tier. Both q_i and every posted menu are independent of i's own report. The public order π may be retained for implementation, but neither the admissible set nor the allocation of any provider depends on the budget left by earlier candidates. Provider i chooses the tier that maximizes its reported surplus:

$$\theta_i^\star \in \arg\max_{\theta \in \mathcal{A}_i^{\mathrm{LOO},(v)}} \{\widehat{p}_i^{(v)}(\theta) - b_i(\theta)\}. \tag{9}$$

If the resulting maximum surplus is nonnegative, the mechanism selects provider i at tier $\theta_i^\star$ and pays $\widehat{p}_i^{(v)}(\theta_i^\star)$; otherwise it skips i. The denominator $m_{-i} + 2$ is chosen so that the individual caps sum to less than B over the global candidate pool, as established below. All BF-order and surplus-maximization ties are resolved by fixed public rules.

3.1 Theoretical Analysis

Let OPT denote the maximum additive value of a set that selects at most one tier per provider and whose total truthful cost is at most B. We establish the mechanism properties, its approximation analysis, and a corresponding lower bound.

Theorem 1. TRIP *is deterministic, truthful, individually rational, and budget feasible*[1].

We next give an approximation guarantee under a realized-utilization condition. The certificate below applies to variants A and C; variant B remains a legal posted-menu comparator but does not inherit the same proportional-payment identity.

Theorem 2. *Assume $k \ge 1$. Let $V_{\min} := \sum_{i=1}^{k} \mu_{i,\theta_{i,\min}}$ and $ALG := \sum_{i \in W} \mu_{i,\theta_i^\star}$. Further, let $\mu_{\min} := \min_{i,\theta} \mu_{i,\theta} > 0$, $\mu_{\max} := \max_{i,\theta} \mu_{i,\theta}$, and $\alpha := \mu_{\min}/\mu_{\max}$. Then $OPT \le (1 + 2/\alpha)V_{\min}$. Moreover, let $P := \sum_{i \in W} p_i^\star$ and define the realized winner rates $\bar{\lambda}_i := p_i^\star/\mu_{i,\theta_i^\star}$. For variant A, $\bar{\lambda}_i$ is its menu rate; for variant C it is the tier-specific capped rate. Fix a utilization slack $\eta \in (0, 1)$. If $P \ge (1 - \eta)B$ and $\bar{\lambda}_i \le B/V_{\min}$ for every $i \in W$, then $ALG \ge \frac{1-\eta}{1+2/\alpha} OPT$.*

Crucially, the leave-one-out caps do not alter the budget allocated for constructing the BF prefix or threshold menus. Instead, they influence the certificate solely via the realized payment P, without introducing any additional budget-compression factor. A sufficient condition for the cap to preserve the uncapped proportional menu is

$$\lambda_i^0 \max_{\theta \in G_i} \mu_{i,\theta} \le \frac{B}{m_{-i} + 2}, \qquad \text{for every } i \in C. \tag{10}$$

Under Eq. (10), variant A coincides with the uncapped rule, and variant C weakly dominates variant A pointwise: $\widehat{p}_i^{(C)}(\theta) \ge \widehat{p}_i^{(A)}(\theta)$. Both variants preserve the proportional-payment identity required by the certificate, while variant B serves purely as an ablation for feasibility and incentives. Furthermore, the cap-sum argument in Theorem 1 renders the prior sequential remaining-budget check redundant. Thus, under Eq. (10), this modification incurs no allocation or approximation loss. We conclude by establishing a lower bound for the unrestricted vector-cost domain.

Theorem 3. *Consider a market with one provider, two tiers $\theta_{\min}$ and $\theta_{\max}$, and budget $B > 0$, with $0 < \mu(\theta_{\min}) = \mu_{\min} < \mu_{\max} = \mu(\theta_{\max})$. For every deterministic, individually rational, budget-feasible, DSIC mechanism, there is a type on which $\frac{ALG}{OPT} \le \frac{\mu_{\min}}{\mu_{\max}}$.*

4 Online Learning Mechanism

We consider the sequential procurement problem introduced in Section 2. The contextual reward model, history-conditioned stage-wise truthfulness, and contextual γ-regret benchmark remain unchanged. In each round, the platform learns the unknown contextual qualities from semi-bandit feedback and supplies an estimated quality vector $q_t \in [0, 1]^M$ to the offline mechanism, together with the current bids and budget B_t.

[1] Due to the space limit, the proofs can be referred in Appendix B.

Algorithm 3 TRIP-MC

Require: Horizon T, budgets $\{B_t\}_{t=1}^T$, feature map ϕ, parameters $\lambda, S, \sigma, \delta$, offline oracle $\textsc{Oracle}_\gamma(\cdot)$

$V \leftarrow \lambda I_d$, $y \leftarrow 0$, $\widehat{\vartheta} \leftarrow 0$
for $t = 1$ to T **do**
  Observe context f_t and bids $\{b_{i,t}(\theta)\}$
  Compute α_t and $\{L_a(t), U_a(t)\}_{a\in\mathcal{A}}$ as defined in Eq. (12) and Eq. (13)
  Construct $\widetilde{q}_t$ by applying the backward sweep to each $\mathcal{C}_i(t)$
  $(S_t, \{p_i(t)\}) \leftarrow \textsc{Oracle}_\gamma(\widetilde{q}_t, B_t)$
  Execute S_t and observe rewards $\{r_{t,a}\}_{a\in S_t}$
  Update $(V, y, \widehat{\vartheta})$ using the observed semi-bandit feedback

4.1 TRIP-MC Mechanism

Contextual linear model. Let $\mathcal{A} = \{(i,\theta) : i \in [n],\ \theta \in G_i\}$ be the set of provider–tier arms. For arm $a = (i,\theta)$ in round t, let $z_{t,a} = \phi(f_t, a) \in \mathbb{R}^d$, where ϕ is a prespecified, bid-independent feature map that jointly represents the observable request and the provider–tier arm. In the repaired implementation, the provider block is a deterministic one-hot vector indexed by the fixed episode provider list; this is injective and collision-free. The request block has dimension 10, the provider block has dimension n, and tier plus interaction blocks contribute six coordinates, so the standard $n = 7$ setting has $d = 23$. A signed hash provider block is retained only as an explicit feature-representation ablation and is not the default. The analysis requires only that $\|z_{t,a}\|_2 \le L$ for all (t,a). We posit a shared linear model: there exists an unknown $\vartheta^\star \in \mathbb{R}^d$ with $\|\vartheta^\star\|_2 \le S$ such that

$$\mu_a(f_t) = \mu_{i,\theta}(f_t) = \langle \vartheta^\star, z_{t,a}\rangle \in [0,1]. \tag{11}$$

Whenever $a \in S_t$ is selected, semi-bandit feedback reveals $r_{t,a} \in [0,1]$ with conditional mean $\mu_a(f_t)$. The within-round noise vector is assumed conditionally independent across selected arms (the weaker joint sub-Gaussian martingale condition also suffices), with each coordinate conditionally σ-sub-Gaussian.

LinUCB statistics and confidence bounds. Fix $\lambda > 0$. After round $t-1$, the ridge statistics are $V_{t-1} = \lambda I_d + \sum_{s=1}^{t-1}\sum_{a\in S_s} z_{s,a}z_{s,a}^\top, y_{t-1} = \sum_{s=1}^{t-1}\sum_{a\in S_s} r_{s,a}z_{s,a}, \widehat{\vartheta}_{t-1} = V_{t-1}^{-1}y_{t-1}$. Use the standard self-normalized radius

$$\alpha_t = \sigma\sqrt{2\log\Big(\frac{\det(V_{t-1})^{1/2}\det(\lambda I_d)^{-1/2}}{\delta}\Big)} + \sqrt{\lambda}\,S. \tag{12}$$

For $a \in \mathcal{A}$, set $\beta_a(t) = \alpha_t\|z_{t,a}\|_{V_{t-1}^{-1}}$. The clipped confidence bounds are

$$L_a(t) = \Pi_{[0,1]}\Big(\langle\widehat{\vartheta}_{t-1}, z_{t,a}\rangle - \beta_a(t)\Big), U_a(t) \quad = \Pi_{[0,1]}\Big(\langle\widehat{\vartheta}_{t-1}, z_{t,a}\rangle + \beta_a(t)\Big), \tag{13}$$

where $\|z\|_{V^{-1}} := \sqrt{z^\top V^{-1} z}$ and $\Pi_{[0,1]}$ denotes projection onto $[0,1]$.

Ordered-tier monotonicity and structured optimism. For each provider i, index the tiers in increasing service level as $\theta_{i,1}, \dots, \theta_{i,L_i}$ and assume $\mu_{i,\theta_{i,1}}(f) \le \mu_{i,\theta_{i,2}}(f) \le \cdots \le \mu_{i,\theta_{i,L_i}}(f), \forall f \in \mathcal{F}$. We therefore define the monotone confidence set $\mathcal{C}_i(t) = \{q_i \in [0,1]^{L_i} : L_{i,\ell}(t) \le q_{i,\ell} \le U_{i,\ell}(t), \forall \ell, q_{i,1} \le \cdots \le q_{i,L_i}\}$. We construct an optimistic representative $\widetilde{q}_i(t)$ by a backward sweep $\widetilde{q}_{i,L_i}(t) = U_{i,L_i}(t)$, and $\widetilde{q}_{i,\ell}(t) = \min\{U_{i,\ell}(t), \widetilde{q}_{i,\ell+1}(t)\}, \ell = L_i - 1, \dots, 1$. The sweep returns the largest vector satisfying the monotonicity constraint and the upper confidence bound. Lemma 1 shows that it also satisfies the lower bounds on the confidence event.

Oracle call and update. Concatenating the provider-wise representatives yields $\widetilde{q}_t = \{\widetilde{q}_{i,\theta}(t)\}_{i,\theta}$. Zero-quality tiers are treated as inactive before the offline ratio rule is evaluated. The mechanism calls $\textsc{Oracle}_\gamma(\widetilde{q}_t, B_t)$, executes the returned set S_t, observes the corresponding semi-bandit rewards, and updates: $V_t = V_{t-1} + \sum_{a\in S_t} z_{t,a}z_{t,a}^\top, y_t = y_{t-1} + \sum_{a\in S_t} r_{t,a}z_{t,a}, \widehat{\vartheta}_t = V_t^{-1}y_t$. Thus, the online layer modifies only the quality vector supplied to the offline mechanism; the allocation and payment rules of the offline mechanism are unchanged. Algorithm 3 summarizes the procedure.

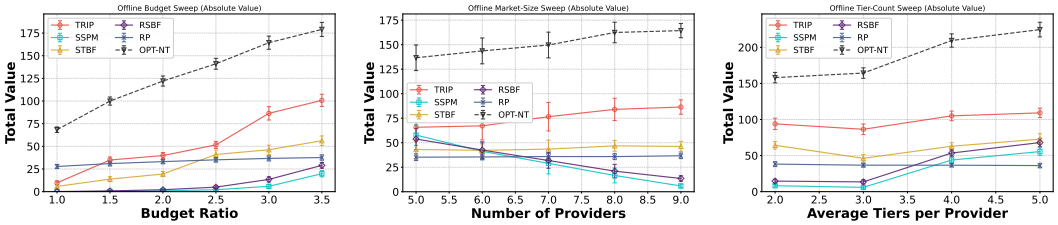


Figure 2: Offline total value under budget, market-size, and tier-richness sweeps.

4.2 Theoretical Analysis

For notational simplicity, write $\mu_{t,a} := \mu_a(f_t)$ for the true contextual mean of arm a in round t.

Lemma 1 (Structured optimism of MonConf)**.** *For each round t, define the history-measurable event $\mathcal{E}_t = \left\{\forall a \in \mathcal{A} : \ |\langle \widehat{\vartheta}_{t-1} - \vartheta^\star, z_{t,a}\rangle| \le \beta_a(t)\right\}$, and let $\mathcal{E} = \bigcap_{t=1}^{T} \mathcal{E}_t$. On $\mathcal{E}_t$, the backward sweep returns the componentwise maximum of the nonempty set $\mathcal{C}_i(t)$ for every provider i. Consequently, for $\mu_t := \{\mu_{t,a}\}_{a\in\mathcal{A}}$, $\mu_t \le \widetilde{q}_t$ and $0 \le \widetilde{q}_a(t) - \mu_{t,a} \le 2\beta_a(t)$, $\forall a \in \mathcal{A}$.*

Theorem 4. *Under truthful current-round reports, the linear contextual model, ordered-tier monotonicity, Eq. (41), $\|z_{t,a}\|_2 \le L$, and $|S_t| \le m$, where $L > 0$ and $m \in \mathbb{N}_+$. Let $\kappa_{m,L,\lambda} := \frac{mL^2/\lambda}{\log(1+mL^2/\lambda)}$ and $\bar{\alpha}_T := \sigma\sqrt{2\log\left[\frac{(1+mTL^2/(\lambda d))^{d/2}}{\delta}\right]} + \sqrt{\lambda}\,S$. Then $R_\gamma(T) \le 2\bar{\alpha}_T\sqrt{\kappa_{m,L,\lambda}\, d\, mT \log\left(1 + \frac{mTL^2}{\lambda d}\right)} + mT\delta$. Setting $\delta = 1/T$ replaces $mT\delta$ by m. With* TRIP*, Theorem 1 also yields stage-wise truthfulness.*

5 Experiments

We evaluate the offline mechanism TRIP and the online mechanism TRIP-MC in trace-backed LLM procurement markets. The experiments address three questions: **(Q1)** how much value TRIP retains relative to a full-information, non-truthful oracle while satisfying the mechanism properties established in Section 3; **(Q2)** whether TRIP-MC improves cumulative reward under semi-bandit feedback; and **(Q3)** how much explicit multi-tier modeling contributes relative to single-tier reductions and simpler posted menus.

Trace-driven setup. We construct offline instances and online rounds using MMLU rollouts from nine LLMs, where questions serve as requests, models as providers, and expected best-of-k scores as tier qualities Kang et al. (2026). Costs incorporate the request workload, provider-specific factors, and monotonically increasing tier multipliers Snell et al. (2025). Budgets are defined as $B_t = \rho C_{\mathrm{ref}}(t)$, where $C_{\mathrm{ref}}(t)$ denotes the median candidate cost, with a default ratio of $\rho = 3$. Our primary evaluation employs TRIP-MC under a standard setting of $n = 7$ providers, $L = 3$ tiers, and 50 random seeds; all figures report the mean and one standard deviation. Performance is measured by total realized value offline and final cumulative reward online. Note that gaps to the OPT-NT ceiling serve as descriptive metrics rather than empirical γ-regret bounds. In the offline setting, we benchmark TRIP against STBF, RSBF, SSPM, and RP, alongside the non-truthful OPT-NT ceiling. For the online setting, TRIP-MC is compared with STBF-PM, STBF-TS, O-RSBF, FM-BwK, and O-RP. These baselines comprehensively cover single-tier reductions, posted-menu procurement, and budget-aware exploration. Formal experimental settings, definitions and theoretical incentive properties for all baselines are provided in Appendix C.

Offline Results. Figure 2 reports the offline total value under budget, market-size, and tier-richness sweeps. The repaired main evaluation employs the tier-wise capped menu (TRIP-MC). At the standard budget ratio $\rho = 3$, it achieves a total value of 86.41, outperforming the strongest implementable baseline (46.22) by 87.0%. The budget sweep demonstrates that TRIP-MC yields consistent improvements at moderate-to-large budgets (ranging from 12.6% to 87.0% for $\rho \ge 1.5$), though it remains sensitive and underperforms the baseline under the most restrictive budget ($\rho = 1.0$). The provider sweep consistently favors TRIP-MC, with relative gains scaling monotonically from 13.7% to 87.0% as the market size n increases from 5 to 9. Furthermore, under the cross-subset

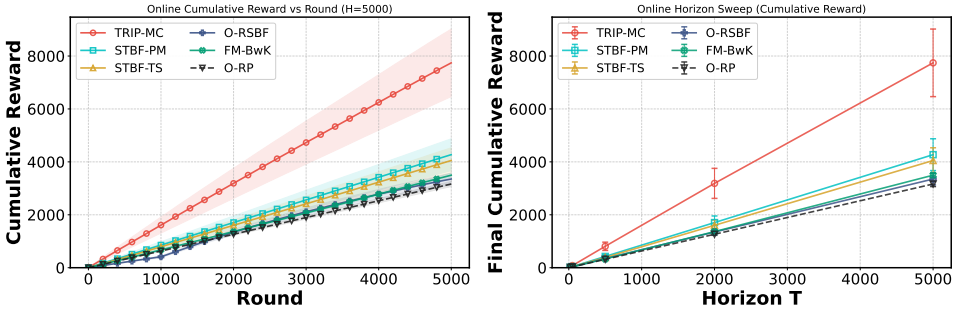


Figure 3: Online learning performance.

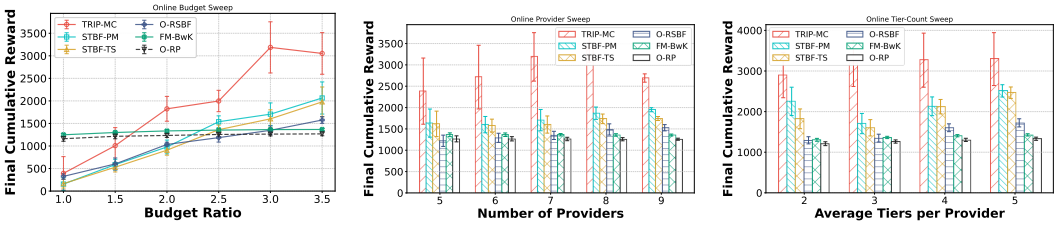


Figure 4: Online robustness under budget, provider-count, and tier-count sweeps.

RSBF pricing rule, the mechanism maintains substantial advantages across varying tier counts ($L \in \{2, 3, 4, 5\}$), yielding improvements between 46.2% and 87.0%. Finally, the offline cap ablation at the standard setting results in total values of 70.16, 88.57, and 87.77 for designs A, B, and C, respectively. Design C is selected as the primary configuration because it recovers nearly all of B's value while preserving the proportional conditional certificate.

Online Results. Figure 3 reports the round-wise cumulative reward and the horizon sweep. With the repaired common-prefix protocol, the tuned TRIP-MC-C consistently outperforms the strongest implementable baseline across all evaluated time horizons ($T \in \{20, \ldots, 5000\}$). It maintains a substantial lead, ranging from 81.2% at $T = 5000$ to 97.7% in earlier rounds. The repaired cap and tuned confidence configuration effectively eliminate the previously observed long-horizon reversal in the standard setting. Note that this remains an empirical comparison, separate from the oracle-conditional theorem. Figure 4 tests whether the advantage persists beyond the standard setting. The budget sweep demonstrates relative gains between 29.9% and 86.9% for moderate-to-large budgets ($\rho \geq 2.0$), though performance drops at tighter budgets. Similarly, the provider sweep exhibits strong positive gains across all tested market sizes (ranging from 38.1% to 86.9% for $n \in \{5, 6, 7, 8, 9\}$). The tier sweep also remains strictly favorable across varying tier counts ($L \in \{2, 3, 4, 5\}$), securing improvements between 28.7% and 86.9%. Thus, the online advantage is robust at the standard setting and remains positive across these stress sweeps.

6 Conclusion

We study truthful procurement from strategic multi-tier LLM providers with hard budgets and unknown service quality. TRIP is a deterministic mechanism that is truthful, individually rational, and budget-feasible, with an approximation certificate under explicit realized-utilization and rate conditions. Its online extension, TRIP-MC, learns contextual tier qualities from semi-bandit feedback while preserving conditional current-round truthfulness; its regret bound is conditional on a per-round approximate-oracle guarantee. On trace-backed markets constructed from MMLU rollouts, we evaluate the framework against truthful baselines and non-truthful references, especially in markets with richer service menus. Future work can extend this framework to combinatorial service bundles, nonstationary provider quality, and longer-horizon strategic participation.

AI USE STATEMENT

Generative AI tools assisted with manuscript organization and drafting, language revision, LaTeX editing, clarification of algorithm descriptions and proof explanations, and interpretation of reported experimental results. AI also assisted with executing experimental code, and the authors manually checked this AI-assisted experimental work. The authors take responsibility for the final text, technical claims, and research artifacts.

ETHICS STATEMENT

This work studies LLM procurement through theoretical analysis and benchmark-based simulations. Practical deployment should account for biases in model outputs and the limitations of benchmark-based quality estimates.

REPRODUCIBILITY STATEMENT

The appendix provides proofs and detailed experimental settings, including baseline definitions and random seeds. We also submit the experimental code as supplementary material to support reproducibility.

REFERENCES

Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. Characterizing truthful multi-armed bandit mechanisms. *SIAM Journal on Computing*, 43(1):194–230, 2014.

Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. *Journal of the ACM (JACM)*, 65(3):1–55, 2018.

Xiaohui Bei, Ning Chen, Nick Gravin, and Pinyan Lu. Budget feasible mechanism design: from prior-free to bayesian. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pp. 449–458, 2012.

Satyanath Bhat, Shweta Jain, Sujit Gujar, and Yadati Narahari. An optimal bidimensional multi-armed bandit auction for multi-unit procurement. *Annals of Mathematics and Artificial Intelligence*, 85(1):1–19, 2019.

Arpita Biswas, Shweta Jain, Debmalya Mandal, and Y Narahari. A truthful budget feasible multi-armed bandit mechanism for crowdsourcing time critical tasks. In *AAMAS*, pp. 1101–1109, 2015.

Mark Braverman, Jieming Mao, Jon Schneider, and S Matthew Weinberg. Multi-armed bandit problems with strategic arms. In *Conference on Learning Theory*, pp. 383–416. PMLR, 2019.

Hau Chan and Jing Chen. Truthful multi-unit procurements with budgets. In *International Conference on Web and Internet Economics*, pp. 89–105. Springer, 2014.

Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. 24, 2011.

Lingjiao Chen, Matei Zaharia, and James Zou. Frugalgpt: How to use large language models while reducing cost and improving performance. *arXiv preprint arXiv:2305.05176*, 2023.

Ning Chen, Nick Gravin, and Pinyan Lu. On the approximability of budget feasible mechanisms. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pp. 685–699. SIAM, 2011.

Dujian Ding, Ankur Mallick, Chi Wang, Robert Sim, Subhabrata Mukherjee, Victor Rühle, Laks Lakshmanan, and Ahmed H Awadallah. Hybrid llm: Cost-efficient and quality-aware query routing. In *International Conference on Learning Representations*, volume 2024, pp. 41348–41366, 2024.

Shangbin Feng, Wenxuan Ding, Alisa Liu, Zifeng Wang, Weijia Shi, Yike Wang, Shannon Zejiang Shen, Xiaochuang Han, Hunter Lang, Chen-Yu Lee, et al. When one llm drools, multi-llm collaboration rules. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 17048–17063, 2026.

Jonas Geiping, Sean McLeish, Neel Jain, John Kirchenbauer, Siddharth Singh, Brian Bartoldson, Bhavya Kailkhura, Abhinav Bhatele, and Tom Goldstein. Scaling up test-time compute with latent reasoning: A recurrent depth approach. *Advances in Neural Information Processing Systems*, 38: 41340–41391, 2026.

Nick Gravin, Yaonan Jin, Pinyan Lu, and Chenhao Zhang. Optimal budget-feasible mechanisms for additive valuations. *ACM Transactions on Economics and Computation (TEAC)*, 8(4):1–15, 2020.

Jiawei Gu, Xuhui Jiang, Zhichao Shi, Hexiang Tan, Xuehao Zhai, Chengjin Xu, Wei Li, Yinghan Shen, Shengjie Ma, Honghao Liu, et al. A survey on llm-as-a-judge. *The Innovation*, 7(6), 2026.

Qitian Jason Hu, Jacob Bieker, Xiuyu Li, Nan Jiang, Benjamin Keigwin, Gaurav Ranganath, Kurt Keutzer, and Shriyash Kaustubh Upadhyay. Routerbench: A benchmark for multi-llm routing system. *arXiv preprint arXiv:2403.12031*, 2024.

Nicole Immorlica, Karthik Sankararaman, Robert Schapire, and Aleksandrs Slivkins. Adversarial bandits with knapsacks. *Journal of the ACM*, 69(6):1–47, 2022.

Zhewei Kang, Xuandong Zhao, and Dawn Song. Scalable best-of-n selection for large language models via self-certainty. *Advances in neural information processing systems*, 38:19720–19745, 2026.

Xiang Liu, Hau Chan, Minming Li, and Weiwei Wu. Budget feasible mechanisms: A survey. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 2024.

Xiang Liu, Hau Chan, Minming Li, Weiwei Wu, and Long Tran-Thanh. Non-stochastic budgeted online pricing with semi-bandit feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18978–18986, 2025.

Isaac Ong, Amjad Almahairi, Vincent Wu, Wei-Lin Chiang, Tianhao Wu, Joseph E Gonzalez, Mohammed Kadous, and Ion Stoica. Routellm: Learning to route llms from preference data. In *International Conference on Learning Representations*, volume 2025, pp. 34433–34448, 2025.

Zhihong Pan, Kai Zhang, Yuze Zhao, and Yupeng Han. Adaptive model and strategy routing for cost-efficient llm services. In *Proceedings of the ACM Web Conference 2026*, pp. 5568–5578, 2026.

Manhin Poon, XiangXiang Dai, Xutong Liu, Fang Kong, John CS Lui, and Jinhang Zuo. Online multi-llm selection via contextual bandits under unstructured context evolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 24855–24863, 2026.

Reza Shirkavand, Shangqian Gao, Peiran Yu, and Heng Huang. Cost-aware contrastive routing for llms. *Advances in Neural Information Processing Systems*, 38:155627–155659, 2026.

Yaron Singer. Budget feasible mechanisms. In *2010 IEEE 51st Annual Symposium on foundations of computer science*, pp. 765–774. IEEE, 2010.

Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling llm test-time compute optimally can be more effective than scaling parameters for reasoning. In *International Conference on Learning Representations*, volume 2025, pp. 10131–10165, 2025.

Zihe Wang and Pingzhong Tang. Optimal mechanisms with simple menus. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pp. 227–240, 2014.

S Matthew Weinberg and Zixin Zhou. Optimal multi-dimensional mechanisms are not locally-implementable. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pp. 875–896, 2022.

Jun Wu, Yuan Zhang, Yu Qiao, Lei Zhang, Chongjun Wang, and Junyuan Xie. Multi-unit budget feasible mechanisms for cellular traffic offloading. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1693–1701, 2019.

Yu Xia, Fang Kong, Tong Yu, Liya Guo, Ryan A Rossi, Sungchul Kim, and Shuai Li. Which llm to play? convergence-aware online model selection with time-increasing bandits. In *Proceedings of the ACM Web Conference 2024*, pp. 4059–4070, 2024.

A Related Work (Full Version)

Our work connects three lines of research: budget-feasible procurement, online learning with strategic agents, and cost-aware routing across LLM services. Each line captures one component of our setting, but their modeling assumptions separate the payment constraint, the learning problem, and the provider's strategic type. We study these components jointly: a platform repeatedly procures heterogeneous service tiers from providers with private vector-valued costs, learns their context-dependent qualities from feedback, and must keep its realized payments within a hard budget.

Budget feasibility. Prior work studies reverse auctions in which a buyer maximizes a known valuation while the total payment to strategic sellers cannot exceed a fixed budget Singer (2010); Liu et al. (2024). Classical results and subsequent approximation improvements focus predominantly on single-parameter environments: each seller controls one item or service and reports one private cost Bei et al. (2012); Chen et al. (2011); Gravin et al. (2020). A smaller multi-unit literature allows a seller to supply several units of one item at a common private per-unit cost Chan & Chen (2014); Wu et al. (2019). Such multi-unit supply differs from our multi-tier model. Units offered by one seller are interchangeable in the former, whereas a provider in our setting offers mutually exclusive, heterogeneous tiers and privately knows a separate cost for each tier. Moreover, budget-feasible mechanisms are generally formulated as static, full-information allocation problems in which item values are known. Our platform must instead learn contextual provider–tier quality online while preserving truthfulness and budget feasibility.

Online learning with strategic agents. Mechanism design has also been integrated with multi-armed bandits. Early work on sponsored-search auctions shows that truthful bandit mechanisms may need to separate exploration from exploitation when click-through rates are unknown and advertisers have private per-click values Babaioff et al. (2014). Other strategic-bandit models allow arms to retain or manipulate rewards rather than report procurement costs Braverman et al. (2019). Resource-constrained bandit formulations, including bandits with knapsacks, impose consumption budgets but treat costs as nonstrategic problem data Badanidiyuru et al. (2018); Immorlica et al. (2022). Procurement-oriented bandit mechanisms are closer to our setting. In particular, budget-feasible crowdsourcing mechanisms learn worker quality while eliciting one scalar cost per worker Biswas et al. (2015), multi-unit procurement auctions learn agent quality while eliciting a per-unit cost and a capacity Bhat et al. (2019), and post-pricing mechanisms learn sellers' private costs Liu et al. (2025). Even these formulations associate an agent with a single service whose repeated units share one quality and per-unit cost. They therefore do not capture a provider that can strategically report a cost vector over heterogeneous tiers, nor the resulting interaction between contextual tier learning and a hard per-round payment budget.

Cost-aware LLM routing. LLM routing and cascading systems select among models to improve the quality–cost trade-off. Representative approaches learn cascades over multiple APIs Chen et al. (2023); Poon et al. (2026); Xia et al. (2024), route between smaller and larger models using predicted difficulty or preference data Ding et al. (2024); Ong et al. (2025), or benchmark routers across a pool of models and operating points Hu et al. (2024). These systems establish the practical value of query-dependent model selection, but they take the available endpoints and their API prices as exogenous. Even when several models, providers, or cascade stages are available, each endpoint is treated as a fixed alternative rather than as part of a strategic provider's privately costed service menu. Consequently, this literature does not design payments, enforce incentive compatibility or individual rationality, or distinguish a hard budget on mechanism payments from an average inference-cost target.

Our setting lies at the intersection of these literatures. Relative to budget-feasible procurement, it replaces scalar seller types and known item values with provider-level cost vectors and unknown contextual tier qualities. Relative to strategic bandits, it combines learning with a truthful, per-round budget-feasible procurement rule over heterogeneous services. Relative to LLM routing, it endogenizes reported service costs and the payments induced by allocation. This intersection motivates the separation in our framework between a truthful posted-menu procurement layer and a bid-independent contextual learning layer.

B PROOFS

B.1 PROOF OF THEOREM 1

Proof. **Truthfulness.** Fix a cap variant $v \in \{A, B, C\}$. Fix a provider i, its true cost vector c_i, and the reports b_{-i} of the other providers. The leave-one-out market excludes i, so its cutoff m_{-i} and the menu $\{p_i(\theta)\}_{\theta \in G_i}$ are independent of b_i. It follows that the selected posted menu $\{\widehat{p}_i^{(v)}(\theta)\}_{\theta \in G_i}$, the cap $q_i = B/(m_{-i}+2)$, and the admissible set $\mathcal{A}_i^{\mathrm{LOO},(v)}$ are all independent of b_i.

We first show that a provider outside the truthful candidate pool cannot profit by underbidding to enter it. Relabel providers by their minimum ratios under the profile (c_i, b_{-i}), using the fixed public tie-breaking rule, and suppose provider r has rank $r \geq k+2$. Removing r leaves both the feasibility of prefix k and the failure of prefix $k+1$ unchanged. Every longer prefix also fails because its left-hand ratio is weakly larger while its budget-per-reward bound is smaller. Hence the leave-one-out cutoff is $m_{-r} = k$, and, for every tier $\theta \in G_r$,

$$\frac{p_r(\theta)}{\mu_{r,\theta}} = \max_{1 \leq j \leq k+1} \min \left\{ \rho_j, \frac{B}{\sum_{\ell=1}^{j-1} \mu_{\ell,\theta_{\ell,\min}} + \mu_{r,\theta}} \right\} \leq \rho_{k+1}. \tag{14}$$

For every variant, $\widehat{p}_r^{(v)}(\theta) \leq p_r(\theta)$ and $p_r(\theta)/\mu_{r,\theta} \leq \rho_{k+1}$. On the other hand, because $r \geq k+2$ under truthful reporting,

$$\frac{c_r(\theta)}{\mu_{r,\theta}} \geq \rho_r \geq \rho_{k+1} \qquad \text{for every } \theta \in G_r. \tag{15}$$

Thus every tier in r's fixed posted menu satisfies

$$\widehat{p}_r^{(v)}(\theta) \leq c_r(\theta). \tag{16}$$

Truthful reporting leaves r outside the candidate pool and gives utility zero. Any deviation that still leaves r outside also gives zero, while a deviation that moves r into the candidate pool gives nonpositive true utility if trade occurs and zero otherwise. Such a provider therefore has no profitable deviation. The inequality can be an equality under ties, so the correct conclusion is nonpositive, rather than necessarily negative, utility.

It remains to consider a provider i in the truthful candidate pool. For every deviation that keeps i in the pool, the mechanism offers the same fixed set $\mathcal{A}_i^{\mathrm{LOO},(v)}$ and chooses a tier maximizing reported surplus over that set. Whatever report is made, the provider's true utility can therefore be no larger than

$$\max\Big\{0, \max_{\theta \in \mathcal{A}_i^{\mathrm{LOO},(v)}} \big(\widehat{p}_i^{(v)}(\theta) - c_i(\theta)\big)\Big\}. \tag{17}$$

Truthful reporting attains this bound. A deviation that removes i from the candidate pool gives utility zero, which cannot improve on the nonnegative truthful utility. Hence truth-telling is a dominant strategy. Notice that no shared remaining budget enters $\mathcal{A}_i^{\mathrm{LOO},(v)}$; thus, although b_i may affect the menus of other candidates, it cannot change the set offered to i.

Individual rationality. Under truthful reporting, an accepted provider i is assigned a tier

$$\theta_i^\star \in \arg\max_{\theta \in \mathcal{A}_i^{\mathrm{LOO},(v)}} \{\widehat{p}_i^{(v)}(\theta) - c_i(\theta)\} \tag{18}$$

only when the displayed maximum is nonnegative. Its realized utility is then

$$p_i^\star - c_i(\theta_i^\star) = \widehat{p}_i^{(v)}(\theta_i^\star) - c_i(\theta_i^\star) \geq 0. \tag{19}$$

If i is not accepted, its utility is zero. Thus the rule is ex-post individually rational.

Budget feasibility. Fix an arbitrary reported profile and let k be its BF cutoff. First suppose $k < n$, so $C = \{1, \ldots, k+1\}$. For each $i \leq k$, removing i leaves the other $k-1$ providers from the feasible BF prefix. The largest ratio in this reduced prefix is at most ρ_k, and its total reward is at most $\sum_{\ell=1}^{k} \mu_{\ell,\theta_{\ell,\min}}$. The BF feasibility of the original prefix therefore implies that this reduced prefix is feasible, so

$$m_{-i} \geq k-1 \quad \text{and hence} \quad q_i \leq \frac{B}{k+1}, \qquad i \leq k. \tag{20}$$

Removing the boundary provider $k+1$ leaves the original feasible prefix of length k, so

$$m_{-(k+1)} \geq k \quad \text{and hence} \quad q_{k+1} \leq \frac{B}{k+2}. \tag{21}$$

Every winner is a candidate and satisfies $p_i^\star \leq q_i$ for every cap variant. Consequently,

$$\sum_{i \in W} p_i^\star \leq \sum_{i \in C} q_i \tag{22}$$

$$\leq \frac{kB}{k+1} + \frac{B}{k+2} = B - \frac{B}{(k+1)(k+2)} < B. \tag{23}$$

This calculation also covers $k = 0$, under the convention that the empty reduced prefix has cutoff zero.

If $k = n$, then $C = [n]$. Removing any i leaves a feasible prefix of length $n-1$, so $m_{-i} \geq n-1$ and $q_i \leq B/(n+1)$. Thus

$$\sum_{i \in W} p_i^\star \leq \sum_{i=1}^{n} q_i \leq \frac{nB}{n+1} < B. \tag{24}$$

The rule is therefore ex-post budget feasible for every reported profile. □

B.2 Proof of Theorem 2

Proof. Let $K = \{1, \ldots, k\}$ denote the BF prefix; this set is distinct from the realized winner set W. Splitting an optimal allocation according to K gives

$$OPT \leq OPT(K) + OPT([n] \setminus K). \tag{25}$$

For every $i \in K$, the reward of any one tier is at most $\mu_{\max} \leq \mu_{i,\theta_{i,\min}}/\alpha$. Since at most one tier per provider can be selected,

$$OPT(K) \leq \frac{1}{\alpha} V_{\min}. \tag{26}$$

If $k < n$, then for $j \notin K$ and every $\theta \in G_j$, $b_j(\theta)/\mu_{j,\theta} \geq \rho_j \geq \rho_{k+1}$. Consequently any budget-feasible allocation using only providers outside K has value at most B/ρ_{k+1}. The BF cutoff condition equation 5 then gives

$$OPT([n] \setminus K) \leq \frac{B}{\rho_{k+1}} < \sum_{i=1}^{k+1} \mu_{i,\theta_{i,\min}} = V_{\min} + \mu_{k+1,\theta_{k+1,\min}} \leq \Big(1 + \frac{1}{\alpha}\Big) V_{\min}. \tag{27}$$

If $k = n$, the outside term is zero and satisfies the same upper bound. Therefore,

$$OPT \leq \Big(1 + \frac{2}{\alpha}\Big) V_{\min}. \tag{28}$$

For the second claim, put $\tau = B/V_{\min}$ and let $P = \sum_{i \in W} p_i^\star$. For variants A and C, define $\bar{\lambda}_i := p_i^\star/\mu_{i,\theta_i^\star}$. The payment identity and the assumed winner-rate bound imply

$$P = \sum_{i \in W} \bar{\lambda}_i \mu_{i,\theta_i^\star} \leq \tau \sum_{i \in W} \mu_{i,\theta_i^\star} = \tau\, ALG, \tag{29}$$

whereas the assumed utilization bound gives $P \geq (1-\eta)B = (1-\eta)\tau V_{\min}$. Hence $ALG \geq (1-\eta)V_{\min}$. Combining this with the benchmark upper bound yields

$$OPT \leq \Big(1 + \frac{2}{\alpha}\Big) V_{\min} \leq \frac{1 + 2/\alpha}{1-\eta}\, ALG. \tag{30}$$

□

B.3 Proof of Theorem 3

Proof. Choose any $\eta > 0$ and consider the two types

$$\begin{aligned} c^{(1)}(\theta_{\min}) &= 0, \quad c^{(1)}(\theta_{\max}) = B, \\ c^{(2)}(\theta_{\min}) &= B, \quad c^{(2)}(\theta_{\max}) = B + \eta. \end{aligned} \tag{31}$$

Under type $c^{(2)}$, tier $\theta_{\max}$ cannot be allocated by an individually rational and budget-feasible mechanism: individual rationality would require a payment at least $B+\eta$, while budget feasibility requires a payment at most B. The offline optimum selects $\theta_{\min}$ and obtains $\mu_{\min} > 0$.

If the mechanism allocates nothing at $c^{(2)}$, its ratio on that type is zero and the result follows. Otherwise it must allocate $\theta_{\min}$. Its payment $p(c^{(2)})$ then satisfies

$$p(c^{(2)}) \geq c^{(2)}(\theta_{\min}) = B \quad \text{and} \quad p(c^{(2)}) \leq B, \tag{32}$$

so $p(c^{(2)}) = B$.

Suppose, for contradiction, that the mechanism allocates $\theta_{\max}$ at $c^{(1)}$. Budget feasibility gives $p(c^{(1)}) \leq B$, and thus truthful utility is

$$u(c^{(1)}; c^{(1)}) = p(c^{(1)}) - c^{(1)}(\theta_{\max}) \leq B - B = 0. \tag{33}$$

Reporting $c^{(2)}$ instead causes allocation of $\theta_{\min}$ and payment B, giving true utility

$$u(c^{(1)}; c^{(2)}) = p(c^{(2)}) - c^{(1)}(\theta_{\min}) = B > 0. \tag{34}$$

This contradicts DSIC. Therefore at $c^{(1)}$ the mechanism allocates $\theta_{\min}$ or nothing, and obtains value at most $\mu_{\min}$.

The offline optimum at $c^{(1)}$ allocates $\theta_{\max}$ at cost B and obtains $\mu_{\max}$. The ratio on this type is therefore at most $\mu_{\min}/\mu_{\max}$. □

B.4 Proof of Lemma 1

Proof. Fix a round t and a provider i. Write $\bar{\mu}_a(t) = \langle \widehat{\vartheta}_{t-1}, z_{t,a} \rangle$. The linear-model assumption gives $\mu_{t,a} = \langle \vartheta^\star, z_{t,a} \rangle \in [0,1]$. On $\mathcal{E}_t$,

$$\bar{\mu}_a(t) - \beta_a(t) \leq \mu_{t,a} \leq \bar{\mu}_a(t) + \beta_a(t). \tag{35}$$

Projection onto $[0,1]$ is monotone and fixes $\mu_{t,a}$, so applying it to the two outer terms gives $L_a(t) \leq \mu_{t,a} \leq U_a(t)$. Define

$$\mu_i(t) := (\mu_{t,(i,\theta_{i,1})}, \ldots, \mu_{t,(i,\theta_{i,L_i})}). \tag{36}$$

By ordered-tier monotonicity, $\mu_i(t)$ is nondecreasing across tiers; together with the coordinate-wise bounds above, this gives $\mu_i(t) \in \mathcal{C}_i(t)$. In particular, $\mathcal{C}_i(t)$ is nonempty.

We next prove that $\widetilde{q}_i(t) \in \mathcal{C}_i(t)$. By construction, $\widetilde{q}_{i,\ell}(t) \leq U_{i,\ell}(t)$ for all ℓ, and the backward sweep yields $\widetilde{q}_{i,1}(t) \leq \cdots \leq \widetilde{q}_{i,L_i}(t)$. It therefore remains to show the lower bounds. Since $\mu_i(t) \in \mathcal{C}_i(t)$, we have $\mu_{i,L_i}(t) \leq U_{i,L_i}(t) = \widetilde{q}_{i,L_i}(t)$. If inductively $\mu_{i,\ell+1}(t) \leq \widetilde{q}_{i,\ell+1}(t)$, then

$$\begin{aligned} \mu_{i,\ell}(t) &\leq \min\{U_{i,\ell}(t), \mu_{i,\ell+1}(t)\} \\ &\leq \min\{U_{i,\ell}(t), \widetilde{q}_{i,\ell+1}(t)\} = \widetilde{q}_{i,\ell}(t). \end{aligned} \tag{37}$$

Backward induction yields $\mu_i(t) \leq \widetilde{q}_i(t)$ componentwise. Since also $L_{i,\ell}(t) \leq \mu_{i,\ell}(t)$, we obtain $L_{i,\ell}(t) \leq \widetilde{q}_{i,\ell}(t) \leq U_{i,\ell}(t)$ for every ℓ, proving $\widetilde{q}_i(t) \in \mathcal{C}_i(t)$.

Now let $q_i \in \mathcal{C}_i(t)$ be arbitrary. Then $q_{i,L_i} \leq U_{i,L_i}(t) = \widetilde{q}_{i,L_i}(t)$. If $q_{i,\ell+1} \leq \widetilde{q}_{i,\ell+1}(t)$, feasibility of q_i implies

$$\begin{aligned} q_{i,\ell} &\leq \min\{U_{i,\ell}(t), q_{i,\ell+1}\} \\ &\leq \min\{U_{i,\ell}(t), \widetilde{q}_{i,\ell+1}(t)\} = \widetilde{q}_{i,\ell}(t). \end{aligned} \tag{38}$$

Backward induction gives $q_i \leq \widetilde{q}_i(t)$ componentwise for every feasible $q_i \in \mathcal{C}_i(t)$. Applying this to $q_i = \mu_i(t)$ for each provider i yields $\mu_t \leq \widetilde{q}_t$ componentwise. Finally, for every arm a,

$$\begin{aligned} 0 \leq \widetilde{q}_a(t) - \mu_{t,a} &\leq U_a(t) - \mu_{t,a} \\ &\leq \max\{0, \bar{\mu}_a(t) + \beta_a(t) - \mu_{t,a}\} \\ &\leq |\bar{\mu}_a(t) - \mu_{t,a}| + \beta_a(t) \leq \beta_a(t) + \beta_a(t) = 2\beta_a(t), \end{aligned} \tag{39}$$

where the first inequality in the second line uses $\Pi_{[0,1]}(x) - y \leq \max\{0, x - y\}$ for $y \in [0,1]$. This proves the lemma. □

B.5 Proof of Theorem 4

Proof. First consider incentives when the oracle is TRIP. Conditional on the realized history and context, $\widetilde{q}_t$ is computed before current-round reports are used. The leave-one-out menu, cutoff, and admissibility cap of each provider are therefore independent of that provider's current report. The dominance argument of Theorem 1 applies in every round, giving current-round DSIC conditional on the history and hence history-conditioned, stage-wise truthfulness in expectation. This argument does not establish dynamic incentive compatibility for deviations that trade current utility against changes in future learning states.

We next prove the regret statement. Let $\mathcal{F}_t$ contain the history through round $t-1$, the current context and reports, and all quantities computed before the oracle is called. The event $\mathcal{E}_t$ is $\mathcal{F}_t$-measurable. On $\mathcal{E}_t$, Lemma 1 gives

$$\mu_t \le \widetilde{q}_t, \qquad 0 \le \widetilde{q}_a(t) - \mu_{t,a} \le 2\beta_a(t),\ \forall a. \tag{40}$$

Since the feasible family is independent of the quality vector, $\mathrm{OPT}_t(\cdot)$ is monotone.

Let $\textsc{Oracle}_\gamma(\cdot)$ denote TRIP applied to the estimated quality vector q_t and budget B_t, with the context and truthful bid profile suppressed. We assume the per-round guarantee

$$\mathbb{E}\big[v(\textsc{Oracle}_\gamma(q_t, B_t); q_t)\big] \ \ge\ \gamma\,\mathrm{OPT}_t(q_t), \tag{41}$$

where the expectation is over any oracle randomness. By Theorem 2, this guarantee holds for TRIP on rounds satisfying its rate and budget-utilization conditions; a fixed horizon-wide γ additionally requires a uniform lower bound on the resulting round-specific factors.

The oracle guarantee equation 41 therefore implies

$$\gamma\,\mathrm{OPT}_t(\mu_t) \le \gamma\,\mathrm{OPT}_t(\widetilde{q}_t) \le \mathbb{E}\big[v(S_t; \widetilde{q}_t) \mid \mathcal{F}_t\big]. \tag{42}$$

Thus, for every history in $\mathcal{E}_t$, additivity gives

$$\begin{aligned}\gamma\mathrm{OPT}_t(\mu_t) - \mathbb{E}\big[v(S_t;\mu_t) \mid \mathcal{F}_t\big] \\ \le 2\alpha_t\,\mathbb{E}\Big[\sum_{a\in S_t} \|z_{t,a}\|_{V_{t-1}^{-1}} \,\Big|\, \mathcal{F}_t\Big].\end{aligned}$$

If $\mathcal{E}_t$ fails, the conditional one-round regret is at most m, while the confidence-width term is nonnegative. Therefore, for every realized $\mathcal{F}_t$,

$$\begin{aligned}\gamma\mathrm{OPT}_t(\mu_t) - \mathbb{E}\big[v(S_t;\mu_t) \mid \mathcal{F}_t\big] \\ \le 2\alpha_t\,\mathbb{E}\Big[\sum_{a\in S_t} \|z_{t,a}\|_{V_{t-1}^{-1}} \,\Big|\, \mathcal{F}_t\Big] + m\mathbf{1}\{\mathcal{E}_t^c\}.\end{aligned} \tag{43}$$

It remains to bound the displayed confidence widths. The update is batched: every arm selected in round t is measured in V_{t-1}^{-1}, so the sequential one-vector version of the elliptical potential lemma cannot be applied directly. Define

$$X_t = \sum_{a\in S_t} z_{t,a} z_{t,a}^\top, \qquad A_t = V_{t-1}^{-1/2} X_t V_{t-1}^{-1/2}. \tag{44}$$

Every eigenvalue η of A_t lies in $[0, mL^2/\lambda]$, because $V_{t-1} \succeq \lambda I_d$, $|S_t| \le m$, and $\|z_{t,a}\|_2 \le L$. Put

$$\kappa_{m,L,\lambda} := \frac{mL^2/\lambda}{\log(1 + mL^2/\lambda)}. \tag{45}$$

The function $x/\log(1+x)$ is increasing on $x \ge 0$, so $\eta \le \kappa_{m,L,\lambda}\log(1+\eta)$. Consequently,

$$\begin{aligned}\sum_{t=1}^{T}\sum_{a\in S_t} \|z_{t,a}\|^2_{V_{t-1}^{-1}} &= \sum_{t=1}^{T} \mathrm{tr}(A_t) \\ &\le \kappa_{m,L,\lambda} \sum_{t=1}^{T} \log\det(I_d + A_t) \\ &= \kappa_{m,L,\lambda} \log\frac{\det(V_T)}{\det(\lambda I_d)} \\ &\le \kappa_{m,L,\lambda}\, d \log\left(1 + \frac{N_{\mathrm{tot}}(T) L^2}{\lambda d}\right).\end{aligned} \tag{46}$$

The last step follows from AM–GM applied to the eigenvalues and $\mathrm{tr}(V_T - \lambda I_d) \le N_{\mathrm{tot}}(T)L^2$. By Cauchy–Schwarz and $N_{\mathrm{tot}}(T) \le mT$, every realized action sequence therefore satisfies

$$\sum_{t=1}^{T} \sum_{a \in S_t} \|z_{t,a}\|_{V_{t-1}^{-1}} \le \sqrt{\kappa_{m,L,\lambda}\, d\, mT \log\left(1 + \frac{mTL^2}{\lambda d}\right)}. \tag{47}$$

Also, the determinant bound and equation 12 give the deterministic upper bound

$$\alpha_t \le \bar{\alpha}_T := \sigma \sqrt{2 \log\left(\left(1 + \frac{mTL^2}{\lambda d}\right)^{d/2} \frac{1}{\delta}\right)} + \sqrt{\lambda}\, S. \tag{48}$$

The standard self-normalized inequality used to construct α_t gives $\Pr(\mathcal{E}^c) \le \delta$, and hence $\Pr(\mathcal{E}_t^c) \le \delta$ for each t. Summing equation 43, taking expectations, and using the pathwise width bound above yields

$$R_\gamma(T) \le 2\bar{\alpha}_T \sqrt{\kappa_{m,L,\lambda}\, d\, mT \log\left(1 + \frac{mTL^2}{\lambda d}\right)} + mT\delta. \tag{49}$$

Setting $\delta = 1/T$ gives the asserted bound. □

C Experiment Details

C.1 Experimental Setup

Trace-backed market construction. We construct both offline instances and online rounds from MMLU rollouts of nine candidate LLMs: Mistral-7B-Instruct-v0.3, Qwen-2.5-7B, Llama-3.3-8B-Instruct, Qwen-2.5-1.5B, Phi-2, DeepSeek-R1-Distill-Qwen-7B, Qwen2.5-3B-Instruct, Gemma-2-9B-it, and Phi-3.5-mini-instruct. Each MMLU question is one request, and each model is one provider. For provider i, tier quality is its trace-derived expected best-of-k score Kang et al. (2026). The standard setting activates the first three entries of $\{1, 2, 4, 8, 16\}$, namely $k \in \{1, 2, 4\}$; the tier sweep varies this prefix length over $L \in \{2, 3, 4, 5\}$. Thus, higher tiers weakly improve service quality while incurring higher costs. Provider-tier rewards lie in $[0, 1]$, but total request value may exceed one when selecting multiple provider-tier pairs.

Costs and budgets. We construct provider-tier costs from a request-dependent workload proxy, provider-specific cost factors, and increasing tier multipliers. Thus, longer requests and stronger service tiers are more expensive, while providers retain heterogeneous cost profiles Snell et al. (2025). The request budget is $B_t = \rho\, C_{\mathrm{ref}}(t)$, where $C_{\mathrm{ref}}(t)$ is the median constructed cost over the candidate provider–tier pairs for request t. We use $\rho = 3$ outside the budget sweep, which varies $\rho \in \{1.0, 1.5, 2.0, 2.5, 3.0, 3.5\}$.

Scenario families and reporting protocol. Offline and online main experiments use seeds $\{0, 1, \ldots, 49\}$. The repaired online horizon sweep uses the common control setting $n = 7$, $L = 3$, $\rho = 3$, and $T \in \{20, 50, 500, 2000, 5000\}$. Every shorter episode is a prefix of the same seed-specific shuffled request stream, so horizon comparisons do not change the sampled request order. Online budget, provider, and tier sweeps use $T = 2000$ and vary only the named parameter around the same standard setting. We report mean $\pm$ one standard deviation over 50 seeds; percentage changes compare these means. The main reported configuration is TRIP-C offline and online, with an injective one-hot provider representation, $S = 2$, and $\sigma = 0.25$ for TRIP-MC. The online monotone optimism step uses the coordinate-wise largest nondecreasing envelope that remains below the tier-wise UCBs; this respects the fact that best-of-k quality is nondecreasing in k. Offline results report total realized value while online results report final cumulative reward. Reported gaps to OPT-NT are descriptive full-information oracle gaps; they are not empirical estimates of the theoretical γ-regret because the conditional certificate does not supply a uniform realized γ for every experimental round.

C.2 Baselines

Offline baselines. We compare TRIP with four mechanism-consistent baselines and one non-truthful upper bound.

Single-Tier Budget Feasible (STBF) first collapses each provider to one randomly chosen tier, and then runs a truthful single-parameter budget-feasible mechanism Singer (2010) on the reduced instance. This baseline tests whether most of the gain comes merely from truthful procurement after removing the multi-tier structure.

Random-Sampling Budget-Feasible (RSBF) Bei et al. (2012) randomly partitions the full multi-tier provider set into subsets A and B. Each subset has budget $B/2$; the posted-payment multiplier for A is calibrated from the total reference value of B, and the multiplier for B is calibrated from the total reference value of A. Each subset then procures its feasible provider-tier options under its cross-calibrated posted prices.

Score-Sorted Posted-Menu (SSPM) builds a bid-independent posted menu by scoring each provider-tier option with estimated value-per-cost and then scans options in descending score order until the budget is exhausted Wang & Tang (2014). It serves as a simple truthful posted-pricing baseline without the full oracle allocation step.

Random-Pricing (RP) samples a public posted price in a budget-scaled range and accepts an option only when its reported bid is no larger than that public offer. It is a budget-feasible posted-price heuristic; we make no DSIC claim for it.

OPT-NonTruthful (OPT-NT) solves the offline welfare-maximizing allocation with full knowledge of realized values and costs, without enforcing truthfulness. It is reported only as an evaluation ceiling rather than an implementable mechanism.

Online baselines. We compare TRIP-MC against five online baselines that share the same trace-backed action space and per-round budget process.

Single-Tier Budget-Feasible Posterior Mean (STBF-PM) randomly keeps one tier per provider and then applies the truthful single-parameter budget-feasible rule using posterior-mean reward estimates, isolating exploitation under a reduced single-tier action space.

Single-Tier Budget-Feasible Thompson Sampling (STBF-TS) uses the same single-tier reduction but replaces posterior-mean selection with Thompson sampling Chapelle & Li (2011), testing whether stronger exploration alone can recover the value lost by removing multi-tier choice.

Online Random-Sampling Budget-Feasible (O-RSBF) uses public history to calibrate a posted offer and then procures options under that offer. It is classified as truthful-style conditional on the public calibration, not as a dynamic incentive theorem.

Fixed-Menu BwK (FM-BwK) Badanidiyuru et al. (2018) keeps a frozen posted menu and learns which menu options to activate via budget-aware upper-confidence exploration, separating bandit learning from menu design.

Online Random-Pricing (O-RP) uses public budget-scaled posted offers and random selection, providing a low-information budget-feasible heuristic without a DSIC claim.

C.3 LLM Trace Construction

We construct the trace-backed market from offline rollouts on the MMLU benchmark, which covers multiple disciplines in the humanities, social sciences, STEM, and professional domains through four-way multiple-choice questions. The candidate provider pool contains nine LLMs: Mistral-7B-Instruct-v0.3, Qwen-2.5-7B, Llama-3.3-8B-Instruct, Qwen-2.5-1.5B, Phi-2, DeepSeek-R1-Distill-Qwen-7B, Qwen2.5-3B-Instruct, Gemma-2-9B-it, and Phi-3.5-mini-instruct. For each model and each MMLU question, we independently sample 32 rollouts, yielding approximately $9 \times 14\text{K} \times 32$ model-question interactions in total.

Each rollout uses generation parameters `max_new_tokens=512`, `temperature=0.7`, and `top_p=0.9`. Randomness is introduced through three sources: the sampling seed, the choice among ten prompt templates, and a small random prefix of 0–3 spaces added to the prompt. Concretely, the k-th rollout for a fixed model-question pair uses seed base_seed $+ k$, which determines both the prompt formatting variation and the token-level sampling randomness. This protocol is designed to expose realistic within-model variability while keeping the data-generation process reproducible.

Each JSONL record stores the model identifier, question identifier, rollout index, seed, instantiated prompt, generated response, ground-truth answer, and two reward signals. The first signal is a hard correctness score `score_0_100`$\in\{0,100\}$, which records whether the final extracted answer matches the ground truth. The second is a dense soft reward `y_norm`$\in[0,1]$, defined as the softmax probability assigned to the correct option among {A, B, C, D}; we also record `reward_log`=log(max(`y_norm`,10^{-10})) as an auxiliary dense transformation. In the procurement environment, `y_norm` serves as the realized utility signal because it is substantially denser than the binary correctness indicator and therefore supports more stable value estimation and bandit learning.

C.4 Parameter Settings

Table 1 summarizes the main configuration choices used in the experiments. Main sweeps use 50 random seeds $\{0, 1, \ldots, 49\}$ and report mean $\pm$ standard deviation; controlled ablations retain their original five-seed configuration. The standard budget and provider experiments use the best-of-k tiers $\{1, 2, 4\}$. The tier-richness sweep uses prefixes of $\{1, 2, 4, 8, 16, 32\}$, with $L \in \{2, 3, 4, 5\}$ in the reported main sweep. Tier values are defined by the expected best-of-k `y_norm` reward induced by the trace-backed rollouts. Per-round budgets follow $B_t := \rho\, C_{\mathrm{ref}}(t)$, where $C_{\mathrm{ref}}(t)$ is the median constructed cost across candidate provider–tier options for the current request, with default $\rho = 3$ and sweep values $\{1.0, 1.5, 2.0, 2.5, 3.0, 3.5\}$.

For request t, let ℓ_t^{prompt} and $\ell_t^{\mathrm{question}}$ denote its prompt and question lengths. We use the request-cost anchor

$$a_t = \max\left\{\frac{\ell_t^{\mathrm{prompt}} + 0.5\ell_t^{\mathrm{question}}}{1000},\ 0.05\right\}, \tag{50}$$

and assign provider i at tier k the cost $c_{i,k}(t) = a_t s_i m_k$. Here, $s_i \in [0.35, 1.10]$ is a fixed model-specific cost multiplier, and the tier multipliers are $m_k \in \{0.12, 0.20, 0.32, 0.50, 0.80, 1.20\}$ for $k \in \{1, 2, 4, 8, 16, 32\}$, respectively. This construction makes costs increase with both request length and service tier while retaining cost heterogeneity across providers. The request budget is

$$B_t = \rho\, C_{\mathrm{ref}}(t), \tag{51}$$

where $C_{\mathrm{ref}}(t)$ is the median constructed cost over the candidate provider–tier pairs for request t. We use $\rho = 3$ outside the budget sweep, which varies $\rho \in \{1.0, 1.5, 2.0, 2.5, 3.0, 3.5\}$.

For completeness, the code registry also fixes the main-paper comparison sets. The offline experiments use TRIP, STBF, RSBF, SSPM, RP, and the non-truthful ceiling OPT-NT; the plotted online experiments use TRIP-MC, STBF-PM, STBF-TS, O-RSBF, FM-BwK, and O-RP.

C.5 More Results

C.5.1 Offline Results

For completeness, we also report the offline value-to-oracle ratio in Figure 5. This view complements the main-paper total-value plots by normalizing each method against OPT-NT and makes the relative efficiency gap across sweeps easier to compare. The normalized curves complement the main-paper total-value plots by separating changes in absolute market value from changes in the fraction of the OPT-NT ceiling attained by each mechanism. Because OPT-NT pays true costs and does not enforce incentive compatibility, these ratios are descriptive evaluation quantities rather than approximation certificates.

Table 1: Reproducibility-oriented parameter settings extracted from the code configuration.

Item	Setting
Random seeds	Main sweeps use $\{0, 1, \ldots, 49\}$; report mean $\pm$ standard deviation across 50 seeds. Controlled ablations retain their original five-seed configuration.
Tier sets	Standard budget/provider setting: $\{1, 2, 4\}$; tier sweep: prefixes of $\{1, 2, 4, 8, 16, 32\}$ with reported $L \in \{2, 3, 4, 5\}$.
Tier-value definition	Expected best-of-k `y_norm` value computed from trace-backed rollouts.
Budget rule	$B_t := \rho\, C_{\mathrm{ref}}(t)$ with default $\rho = 3$.
Budget sweeps	$\rho \in \{1.0, 1.5, 2.0, 2.5, 3.0, 3.5\}$ for offline, online, and stress-test sweeps.
Tier cost multipliers	$c_k / c_{\mathrm{base}} \in \{0.12, 0.20, 0.32, 0.50, 0.80, 1.20\}$ for $k \in \{1, 2, 4, 8, 16, 32\}$.
Model-specific base costs	Mistral-7B: 1.0; Qwen-2.5-7B: 0.9; Llama-3.3-8B: 1.0; Qwen-2.5-1.5B: 0.4; Phi-2: 0.35; DeepSeek-R1-Distill-Qwen-7B: 1.1; Qwen2.5-3B: 0.55; Gemma-2-9B-it: 1.0; Phi-3.5-mini: 0.6.
Provider heterogeneity	Fixed model-specific multipliers listed above; the $[0.8, 1.2]$ rank-based fallback is used only when no model-specific mapping is configured.
Request cost anchor	Prompt-length weight 1.0, question-length weight 0.5, minimum anchor cost 0.05.
Main market-size sweep	Number of providers in $\{5, 6, 7, 8, 9\}$.
Main tier-richness sweep	Average tiers per provider in $\{2, 3, 4, 5\}$.
Online horizons	$\{20, 50, 500, 2000, 5000\}$, using common seed-specific request prefixes; standard controls are $n = 7$, $L = 3$, and $\rho = 3$.
TRIP cap variants	A: capped proportional rate; B: threshold truncation; C: tier-wise cap on the proportional menu. Main offline and online runs use C; A/B/C are evaluated as controlled legal ablations on the same episodes.
Provider feature representation	Injective one-hot provider block by default; signed hash block of dimension 4 is an explicit ablation only.
TRIP-MC parameter sweep	$S \in \{0.5, 1, 2\}$ and $\sigma \in \{0, 0.1, 0.25\}$ at fixed $T = 2000$, $n = 7$, $L = 3$, $\rho = 3$; $\lambda = 1$ and $\delta = 1/T$.
Online noise levels	The repaired main run uses $S = 2$ and $\sigma = 0.25$; the S, σ grid is reported as an explicit parameter ablation.

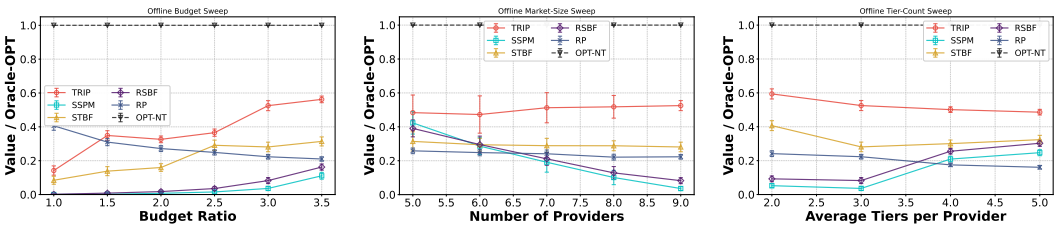


Figure 5: Offline value-to-oracle ratio of TRIP and the mechanism-consistent baselines under budget, market-size, and tier-richness sweeps. From left to right, the panels report the budget sweep, the market-size sweep, and the tier-count sweep. Each curve shows the ratio between realized total value and the value of OPT-NT.