LLM-Auction: Generative Auction towards LLM-Native Advertising

Chujie Zhao, Qun Hu, Shiping Song, Dagui Chen,* Han Zhu, Jian Xu, Bo Zheng
Taobao & Tmall Group of Alibaba
Beijing, China
zhaochujie.zcj@alibaba-inc.com, dagui.cdg@alibaba-inc.com

Abstract

The commercialization of LLM applications is the next frontier in online advertising, with *LLM-native advertising* emerging as a promising paradigm by integrating ads into LLM-generated content. However, classic mechanisms are no longer applicable in this setting where the auction object is shifted from discrete ad slots to distributions over LLM outputs, and existing methods are impractical in industrial scenarios due to ignored externalities or high inference costs. To address these issues, we propose LLM-Auction, the *first* learning-based generative auction mechanism that *integrates* auction and generation. By formulating the allocation as preference alignment between LLM outputs and a mechanism objective that balances advertiser value and user experience, we optimize the LLMs to inherently model allocation externalities without extra inference cost. Theoretically, we identify the allocation monotonicity and continuity of the optimal allocation rule of LLM-Auction, and prove that a simple first-price payment rule exhibits favorable incentive properties under this allocation rule. Furthermore, we build an LLM-as-a-judge simulation environment for quantitative evaluation, and experiments demonstrate that LLM-Auction achieves the state-of-the-art allocation efficiency while approximately satisfying key mechanism properties.

1 Introduction

The rise of chatbots and AI search engines powered by large language models (LLMs) [3, 42] is fundamentally reshaping how users acquire information [32]. As these applications become primary gateways for user traffic, the development of compatible commercialization strategies has become increasingly critical. Current industrial efforts adopt *external* ad placements, such as fixed ad slots attached to LLM responses [21, 26, 28]. These formats are easy to deploy, but do not take advantage of the capabilities of LLMs in reasoning and generation.

We focus on *LLM-native advertising*, where promotional content is naturally integrated into an LLM response while balancing advertiser value and user experience. This setting changes the allocation object from discrete slots to a distribution over LLM outputs: advertisers bid to influence this distribution, increasing the likelihood of their desired content appearing in the generated response. This shift renders classic mechanisms like Generalized Second-Price (GSP) auction [8] inapplicable, creating urgent demand for tailored auction mechanisms.

Existing mechanisms for LLM-native advertising mostly decouple auction from generation [6, 7, 15, 31]. As shown in Figure 1, *pre-generation* mechanisms choose ads before generation, requiring preset ad counts or insertion structures and weakening their ability to model response-level externalities, leading to suboptimal performance. *Post-generation* mechanisms generate ad-conditioned tokens or

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

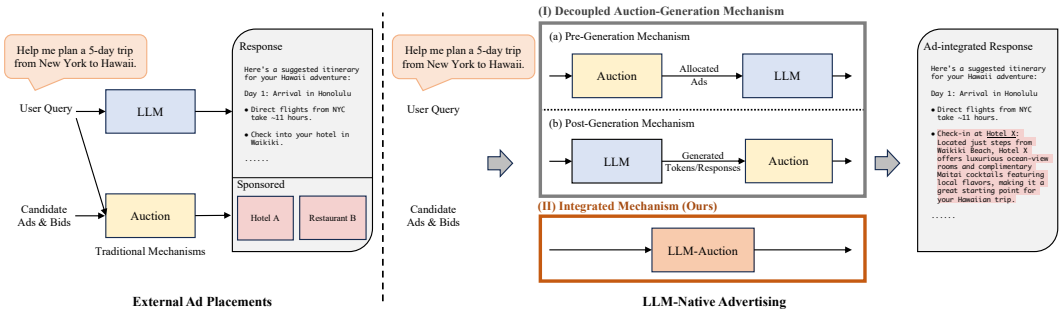

Figure 1: Comparison of ad formats and auction mechanisms in LLM-based AI applications.

candidate responses before running an auction. They are more context-aware, but require extra LLM forward passes whose cost grows with the number of advertisers or candidates, limiting practicality in high-concurrency serving systems.

To address these challenges, we propose LLM-Auction, the *first* learning-based generative auction mechanism that *integrates* auction and LLM generation for LLM-native advertising. Its key idea is to make the LLM itself the allocation rule. Conditioned on the user query, profile, candidate ads, and bids, LLM-Auction post-trains the model to prefer responses that maximize a mechanism objective balancing advertiser value and user experience. We instantiate the objective with a reward model based on advertising metrics such as click-through rate, and introduce *Iterative Reward-Preference Optimization* (IRPO), which alternates between reward model updates from feedback and LLM updates from reward-induced preferences. At serving time, one LLM forward pass generates the ad-integrated response, internalizing response-level externalities without extra inference.

We further show that the optimal LLM-Auction allocation satisfies monotonicity and continuity, linking response-distribution allocation to advertiser incentives. With a simple first-price payment rule, which avoids the costly computation of the dominant-strategy incentive-compatible (DSIC) payment, these properties guarantee a mixed-strategy equilibrium with an explicit bid-shading condition for utility maximizers and effective-value incentive compatibility for ROI-constrained value maximizers under the optimal allocation rule.

Finally, inspired by the LLM-as-a-judge paradigm [17], we build a simulation environment covering user-query generation, ad-integrated response generation, and user feedback for large-scale training and quantitative evaluation. Extensive experiments demonstrate that LLM-Auction *significantly* outperforms existing mechanisms in allocation efficiency while approximately satisfying key mechanism properties.

Our main contributions are summarized as follows:

- We lay down general definitions and incentive desiderata for LLM-native advertising auction mechanisms, establishing foundations for future work.
- We propose LLM-Auction, the *first* learning-based generative auction mechanism that integrates allocation and generation through preference alignment.
- We identify the allocation monotonicity and continuity of the optimal allocation rule of LLM-Auction, and prove that a simple first-price payment rule exhibits favorable incentive properties under this allocation rule.
- We develop an LLM-as-a-judge simulation environment for scalable data construction and quantitative evaluation of LLM-native advertising mechanisms.

2 Related Work

Position auctions in online advertising. Varian [33] established the game-theoretic foundation of position auctions in sponsored search, and Edelman et al. [8] formalized the GSP auction with equilibrium and efficiency guarantees. Caragiannis et al. [4] further prove that GSP achieves a

constant-factor approximation to optimal social welfare even under worst-case equilibria. Recent work extends GSP with learning-based mechanisms [20, 41], while Zhu et al. [44] propose a multi-slot generative auction that accounts for externalities.

Mechanism design for LLMs. Mechanism design for LLMs studies auction mechanisms for LLM-generated content with desirable incentive properties. Existing methods typically decouple auction from generation, falling into two categories. (i) *Pre-generation mechanisms* allocate ads before generation: Dubey et al. [6] introduce a prominence-based auction for LLM summarization, Hajiaghayi et al. [15] design RAG-based mechanisms, and Balseiro et al. [1] extend position auctions to AI-generated content. These mechanisms must specify ad counts or positions upfront and do not model allocation externalities, such as context and co-exposure effects. (ii) *Post-generation mechanisms* first produce ad-conditioned candidates, then select or aggregate them by auction. Feizi et al. [11] and Mordo et al. [23] add ad rewriting to position auctions; Dütting et al. [7] aggregate advertisers' preferences during decoding; and Soumalias et al. [31] select among response-level candidates. Although more context-aware, these methods require extra LLM inference that scales with the number of advertisers or candidate responses, limiting high-concurrency deployment. Bergemann et al. [2] also adapt VCG mechanisms to LLM settings from a theoretical perspective.

LLM alignment methods. LLM alignment steers model outputs toward external objectives or human preferences. RLHF [5, 25] optimizes LLMs with learned preference reward models, while DPO [27] derives a closed-form objective for direct optimization from preference pairs. Subsequent variants, including IPO [13], KTO [9], and SimPO [22], refine this objective, while tuning-free methods such as LA [12] and Amulet [40] adapt preferences at inference time through latent representations.

3 Preliminaries

We formalize an LLM-native advertising auction as a single-turn interaction between a user and an LLM. Let $[n]$ denote the advertisers. Each advertiser i has a private value $v_i \in \mathbb{R}_{\geq 0}$ and a fixed ad $a_i \in \mathcal{A}$; write $\boldsymbol{v} = (v_1, \ldots, v_n)$ and $\boldsymbol{a} = (a_1, \ldots, a_n)$. The context is $s = (x, h, \boldsymbol{a}) \in \mathcal{S}$, where $x \in \mathcal{X}$ is the user query and $h \in \mathcal{H}$ is the user profile.

Given context s and bid profile $\boldsymbol{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n_{\geq 0}$, the LLM π_θ defines an allocation distribution $\pi_\theta(\cdot \mid s, \boldsymbol{b}) \in \Delta(\mathcal{Y})$ over responses. A realized response $y \sim \pi_\theta(\cdot \mid s, \boldsymbol{b})$ may contain content related to any subset of candidate ads. An auction mechanism is a tuple $\mathcal{M} = (\pi_\theta, \boldsymbol{p})$:

- **Allocation rule**: $\pi_\theta : \mathcal{S} \times \mathbb{R}^n_{\geq 0} \to \Delta(\mathcal{Y})$ maps the context and bids to a response distribution.
- **Payment rule**: $\boldsymbol{p} : \Delta(\mathcal{Y}) \times \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$ maps an allocation distribution and bids to payments $\boldsymbol{p}(\pi_\theta(\cdot \mid s, \boldsymbol{b}), \boldsymbol{b})$.

To measure the effectiveness of ad a_i in response y and context s, let $f_i(y, s)$ denote a performance metric such as click-through rate (CTR) or conversion rate (CVR). We extend this notation to distributions by writing $f_i(\pi_\theta(\cdot \mid s, \boldsymbol{b}), s) := \mathbb{E}_{y \sim \pi_\theta(\cdot \mid s, \boldsymbol{b})}[f_i(y, s)]$. A higher bid should not reduce an advertiser's expected performance, yielding the following desideratum.

Definition 3.1 (Allocation Monotonicity)**.** An allocation rule π_θ is *monotone* if, for every advertiser i, context s, competing bids $\boldsymbol{b}_{-i}$, and bids $b_i' \geq b_i$,

$$f_i(\pi_\theta(\cdot \mid s, (b_i', \boldsymbol{b}_{-i})), s) \geq f_i(\pi_\theta(\cdot \mid s, (b_i, \boldsymbol{b}_{-i})), s).$$

Advertisers optimize bids under different objectives and constraints. We consider two standard models: utility maximizers (UM) and value maximizers (VM).

Definition 3.2 (Utility Maximizer (UM))**.** A utility-maximizing advertiser i chooses $b_i \geq 0$ to maximize expected quasi-linear utility:

$$u_i(b_i; s, v_i, \boldsymbol{b}_{-i}) = v_i \cdot f_i(\pi_\theta(\cdot \mid s, (b_i, \boldsymbol{b}_{-i})), s) - p_i(\pi_\theta(\cdot \mid s, (b_i, \boldsymbol{b}_{-i})), (b_i, \boldsymbol{b}_{-i})). \tag{1}$$

Definition 3.3 (Value Maximizer (VM))**.** Let $\tau_i > 0$ be advertiser i's target ROI. A value maximizer chooses $b_i \geq 0$ to maximize expected value subject to the ROI constraint:

$$\begin{aligned} \max_{b_i \geq 0} \quad & v_i \cdot f_i(\pi_\theta(\cdot \mid s, (b_i, \boldsymbol{b}_{-i})), s) \\ \text{s.t.} \quad & \tau_i \cdot p_i(\pi_\theta(\cdot \mid s, (b_i, \boldsymbol{b}_{-i})), (b_i, \boldsymbol{b}_{-i})) \leq v_i \cdot f_i(\pi_\theta(\cdot \mid s, (b_i, \boldsymbol{b}_{-i})), s). \end{aligned} \tag{2}$$

The following properties capture incentives and participation.

Definition 3.4 (Incentive Compatibility for Utility Maximizers (UM-IC))**.** A mechanism is *UM-IC* if truthful value reporting is a dominant strategy: for every i, s, $\boldsymbol{b}_{-i}$, and $b_i \geq 0$, $u_i(v_i; s, v_i, \boldsymbol{b}_{-i}) \geq u_i(b_i; s, v_i, \boldsymbol{b}_{-i})$.

Definition 3.5 (Effective-Value Incentive Compatibility for Value Maximizers (VM-IC))**.** For a value maximizer with type (v_i, τ_i), define the effective value $\tilde{v}_i = v_i/\tau_i$. A mechanism is *VM-IC* if, for every i, s, $\boldsymbol{b}_{-i}$, $b_i = \tilde{v}_i$ is a dominant strategy for maximizing expected value subject to the ROI constraint. Thus, truthful reporting in VM means reporting effective value rather than raw value v_i.

Definition 3.6 (Individual Rationality (IR))**.** A mechanism is *individually rational* if, for every advertiser, context, and bid profile, expected payment is at most expected value:

$$p_i(\pi_\theta(\cdot \mid s, \boldsymbol{b}), \boldsymbol{b}) \leq v_i \cdot f_i(\pi_\theta(\cdot \mid s, \boldsymbol{b}), s).$$

The designer seeks an allocation policy that balances advertiser value with user experience:

$$J(\pi_\theta, \boldsymbol{v}, \boldsymbol{b}, s) = \sum_{i=1}^{n} v_i \cdot f_i(\pi_\theta(\cdot \mid s, \boldsymbol{b}), s) + \lambda \cdot \Omega_{\text{resp}}(\pi_\theta(\cdot \mid s, \boldsymbol{b}), s) - \beta \cdot \mathcal{D}_{\text{KL}}(\pi_\theta(\cdot \mid s, \boldsymbol{b}) \| \pi_{\theta_0}(\cdot \mid s, \boldsymbol{b})). \quad (3)$$

Here, $\Omega_{\text{resp}}(\pi_\theta(\cdot \mid s, \boldsymbol{b}), s)$ is the expected penalty for ad-insertion constraint violations, and $\mathcal{D}_{\text{KL}}$ penalizes deviations from the reference LLM π_{θ_0}, i.e., the pre-trained LLM before any auction-oriented training, to preserve its response quality.

Definition 3.7 (Optimal LLM-Native Advertising Mechanism)**.** In the LLM-native advertising setting, the optimal mechanism $\mathcal{M}^* = (\pi_{\theta^*}, \boldsymbol{p}^*)$ consists of: (i) an allocation rule π_{θ^*} that maximizes Eq. (3) for every realization $(\boldsymbol{v}, \boldsymbol{b}, s)$; and (ii) a payment rule $\boldsymbol{p}^*$ that ensures the relevant IC and IR.

Since values $\boldsymbol{v}$ are private and unobservable, the mechanism uses the reported bid vector $\boldsymbol{b}$ as the value signal, yielding the bid-based objective $J(\pi_\theta, \boldsymbol{b}, \boldsymbol{b}, s)$. In industrial deployment, this optimal mechanism may be relaxed for computational tractability; for example, strict IC may be relaxed to support an efficient payment rule.

4 LLM-Auction

We now introduce LLM-Auction, the first integrated generative auction mechanism for LLM-native advertising. Full proofs are deferred to Appendix C, and Appendix D discusses the plausibility of the assumptions underlying the theoretical results.

4.1 Allocation Rule

Section 3 defines the allocation result as the LLM response distribution. LLM-Auction parameterizes the allocation rule directly with π_θ, so allocation optimization amounts to shifting $\pi_\theta(\cdot \mid s, \boldsymbol{b})$ toward the objective in Eq. (3). We cast this as preference alignment: the LLM is fine-tuned on contexts, candidate ads, and bids using a reward model aligned with the mechanism objective.

4.1.1 Mechanism Reward Model

Preference alignment requires comparing alternative ad-integrated responses for the same input $(s, \boldsymbol{b})$, but user logs reveal feedback only for realized responses. We therefore build a reward model from a performance metric predictor, which estimates the advertiser value of candidate responses during offline LLM training.

Performance metric prediction. We use a predictor $\hat{f}^\psi$ to estimate response-level advertising performance. Unlike serving-time prediction models, $\hat{f}^\psi$ is used only for offline LLM training and can condition on the generated response y. Let $\hat{f}_i^\psi(y, s)$ denote the predicted metric for ad i under response-context pair (y, s). We train it on feedback logs with binary cross-entropy:

$$\mathcal{L}_{\text{BCE}} = -\sum_{i=1}^{n} \left(c_i \log \hat{f}_i^\psi(y, s) + (1 - c_i) \log(1 - \hat{f}_i^\psi(y, s)) \right), \quad (4)$$

where $c_i \in \{0, 1\}$ is the feedback label for ad i (e.g., click or conversion). To connect the learned reward to the mechanism objective, we use the following idealization:

Assumption 4.1 (Unbiased performance prediction)**.** The prediction model is unbiased under any allocation rule π_θ if, for any advertiser i, context s, and bid vector $\boldsymbol{b}$,

$$\mathbb{E}_{y\sim\pi_\theta(\cdot|s,\boldsymbol{b})}[\hat{f}_i^{\psi}(y,s)] = \mathbb{E}_{y\sim\pi_\theta(\cdot|s,\boldsymbol{b})}[f_i(y,s)]$$

Reward model. We fine-tune the LLM with Direct Preference Optimization (DPO) [27] because it offers a stable preference-learning objective with a KL-regularized reward-maximization interpretation. Taking bids as the value signal available to the mechanism and using Assumption 4.1, we define the response-level reward:

$$\mathcal{R}(s,\boldsymbol{b},y) = \sum_{i=1}^{n} b_i \cdot \hat{f}_i^{\psi}(y,s) + \lambda \cdot \Omega_{\text{resp}}(y), \tag{5}$$

where $\Omega_{\text{resp}}(y)$ penalizes ad-insertion constraint violations in the generated response. This reward provides preference feedback for aligning the LLM with the desired allocation rule.

4.1.2 Iterative Reward-Preference Optimization Algorithm

Direct optimization of Eq. (5) with a fixed predictor can suffer from a distribution shift: as the LLM policy changes, its responses move away from the data used to train $\hat{f}^{\psi}$. We address this with Iterative Reward-Preference Optimization (IRPO), shown in Figure 2, which alternates between updating the reward model from user feedback and updating the LLM with reward-induced preferences.

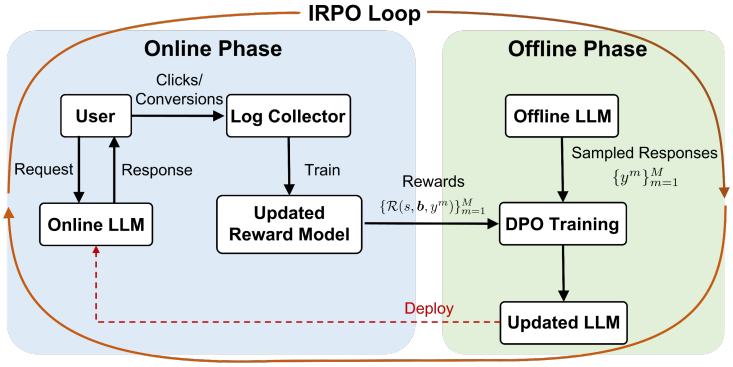

Figure 2: Illustration of the iterative optimization framework of LLM-Auction.

Let π_{θ_t} and $\mathcal{R}_{\psi_t}$ denote the LLM and reward model at epoch t. IRPO has two phases:

Reward model update. We freeze $\pi_{\theta_{t-1}}$, deploy it to collect online user feedback, and construct $\mathcal{D}_{\text{on}}^t = \{s^k, y^k, \boldsymbol{c}^k\}_{k=1}^{N_{\text{on}}}$. We then optimize the predictor with Eq. (4), yielding the updated reward model $\mathcal{R}_{\psi_t}$.

LLM update. We fix $\mathcal{R}_{\psi_t}$ and construct $\mathcal{D}_{\text{off}}^t$ from $\mathcal{D}_{\text{on}}^t$ or synthetic data. For each $(s,\boldsymbol{b}) \in \mathcal{D}_{\text{off}}^t$, we sample M responses $y^m \sim \pi_{\theta_{t-1}}(\cdot \mid s, \boldsymbol{b})$ and score each response with $r^m = \mathcal{R}_{\psi_t}(s,\boldsymbol{b},y^m)$. The highest-reward response y^w is the winner, and responses with a reward gap above threshold δ_{th} form the losing set $\mathcal{K} = \{y^l \mid r^w - r^l > \delta_{\text{th}}\}$. The DPO loss is:

$$\mathcal{L}_{\text{DPO}} = -\frac{1}{|\mathcal{K}|}\sum_{y^l\in\mathcal{K}}\left[\log\sigma\left(\beta\log\frac{\pi_{\theta_t}(y^w|s,\boldsymbol{b})}{\pi_{\theta_{t-1}}(y^w|s,\boldsymbol{b})} - \beta\log\frac{\pi_{\theta_t}(y^l|s,\boldsymbol{b})}{\pi_{\theta_{t-1}}(y^l|s,\boldsymbol{b})}\right)\right], \tag{6}$$

where β controls the KL penalty. Minimizing Eq. (6) gives the next LLM π_{θ_t}.

IRPO progressively refreshes reward estimation under the current allocation policy, reducing distribution shifts and improving allocation performance. Algorithm 1 provides the pseudocode.

4.1.3 Properties of Allocation Rule

Next, we analyze the theoretical properties of the optimal allocation rule π_{θ^*} induced by Eq. (3). Throughout the theoretical analysis, we impose the following assumption on the reference LLM π_{θ_0}.

Assumption 4.2 (Bid Independence of the Reference LLM Distribution)**.** For every context s, response y, and bid profile $\boldsymbol{b} \in [0, \bar{b}]^n$, the reference LLM distribution satisfies

$$\pi_{\theta_0}(y \mid s, \boldsymbol{b}) = \pi_{\theta_0}(y \mid s).$$

Intuitively, the reference LLM has not yet learned how bids should shape its output distribution. The experimental results in Figure 4(a) also provide empirical support for this assumption.

Theorem 4.3 (Monotonicity of the Optimal Allocation)**.** *The optimal allocation π_{θ^*} that maximizes the bid-based mechanism objective $J(\pi_\theta, \boldsymbol{b}, \boldsymbol{b}, s)$ satisfies allocation monotonicity. Formally, for any advertiser i, fixed context s, and competing bids $\boldsymbol{b}_{-i}$, if $b'_i \geq b_i$, then:*

$$f_i(\pi_{\theta^*}(\cdot \mid s, b'_i, \boldsymbol{b}_{-i}), s) \geq f_i(\pi_{\theta^*}(\cdot \mid s, b_i, \boldsymbol{b}_{-i}), s).$$

Theorem 4.3 states that increasing a bid cannot reduce the advertiser's expected performance, providing the allocation-side foundation for incentive analysis.

Theorem 4.4 (Continuity of the Optimal Allocation)**.** *In the LLM-native advertising setting, under Assumption 4.2, the expected performance of the optimal allocation rule π_{θ^*} satisfies continuity: $f_i(\pi_{\theta^*}(\cdot \mid s, \boldsymbol{b}), s)$ is continuous in $\boldsymbol{b}$ for all i, s.*

We will discuss the practical plausibility of Assumption 4.2 in Appendix D and empirically validate allocation continuity in Section 5.4.

4.2 First-Price Payment Rule

For utility maximizers (UMs), the monotonicity of the optimal allocation rule π_{θ^*} admits a DSIC payment through Myerson's payment identity [24]. Let $A_i(\boldsymbol{b}; s) := f_i(\pi_{\theta^*}(\cdot \mid s, \boldsymbol{b}), s)$ denote advertiser i's effective allocation. With the normalization $P_i^{\mathrm{DSIC}}(0, \boldsymbol{b}_{-i}; s) = 0$, the corresponding expected payment is

$$P_i^{\mathrm{DSIC}}(\boldsymbol{b}; s) = b_i A_i(\boldsymbol{b}; s) - \int_0^{b_i} A_i(z, \boldsymbol{b}_{-i}; s)\, dz.$$

By the Gibbs form of π_{θ^*}, this integral equals β times the log-ratio of two partition functions (Appendix C.5). However, exactly evaluating these partition functions over the LLM response space is computationally expensive.

We therefore adopt the operationally simpler first-price payment rule:

$$p_i(\pi_\theta(\cdot \mid s, \boldsymbol{b}), \boldsymbol{b}) = b_i \cdot f_i(\pi_\theta(\cdot \mid s, \boldsymbol{b}), s). \tag{7}$$

Given the allocation, this payment requires only one multiplication per advertiser, avoiding the bid integral and counterfactual partition functions of the DSIC payment. Although first-price payments are generally not DSIC for UMs, the bidding game they induce under the optimal allocation rule admits a mixed-strategy Nash equilibrium, whose interior bids satisfy a bid-shading condition.

Theorem 4.5 (UM Equilibrium under First-Price Payment Rule)**.** *Under the optimal allocation rule π_{θ^*} with compact bid spaces, the first-price bidding game among UMs admits a mixed-strategy Nash equilibrium. At any such equilibrium $\sigma^\star$, every interior support bid b_i with $\overline{V}_i(b_i) > 0$ satisfies*

$$b_i = v_i - \beta \frac{\overline{A}_i(b_i)}{\overline{V}_i(b_i)}. \tag{8}$$

Here, $\overline{A}_i(b_i) := \mathbb{E}_{\boldsymbol{b}_{-i} \sim \sigma^\star_{-i}}[A_i(b_i, \boldsymbol{b}_{-i}; s)]$ and $\overline{V}_i(b_i) := \mathbb{E}_{\boldsymbol{b}_{-i} \sim \sigma^\star_{-i}}[\mathrm{Var}_{y \sim \pi_{\theta^*}(\cdot \mid s, b_i, \boldsymbol{b}_{-i})}[f_i(y, s)]]$ are advertiser i's expected allocation and expected performance variance when it bids b_i against $\sigma^\star_{-i}$.

The intuition behind Eq. (8) is a marginal cost–benefit balance. Raising b_i increases the payment on advertiser i's existing expected allocation (marginal cost $\overline{A}_i(b_i)$), but also tilts π_{θ^*} toward responses with above-average $f_i(y, s)$, buying additional allocation at rate $\overline{V}_i(b_i)/\beta$, each unit worth the net surplus $v_i - b_i$ (marginal benefit $(v_i - b_i)\overline{V}_i(b_i)/\beta$). An interior support bid balances the two exactly, so advertiser i shades its bid below its value v_i by $\beta \overline{A}_i(b_i)/\overline{V}_i(b_i)$, and shades more when its allocation is less responsive to its bid (smaller $\overline{V}_i(b_i)$) or when its existing allocation is already large.

For value maximizers (VMs), we prove that truthful reporting of effective value is an optimal strategy under the first-price payment rule:

Theorem 4.6 (VM-IC under First-Price Payment Rule and an ROI Constraint)**.** *In the LLM-native advertising setting with the optimal allocation rule, the first-price payment rule is VM-IC under ROI constraint τ_i, where the optimal bidding strategy is $b_i = \tilde{v}_i = v_i/\tau_i$.*

Furthermore, individual rationality (IR) is naturally satisfied under standard no-overbidding behavior for UMs and when $\tau_i \geq 1$ for VMs under the first-price payment rule.

In summary, the first-price payment rule is computationally efficient, satisfies IR, admits equilibria for UMs, and achieves VM-IC for ROI-constrained VMs.

5 Experiments

5.1 LLM-Native Advertising Simulation Environment

A key bottleneck in mechanism design for LLM-native advertising is the absence of standardized, quantitative evaluation at scale. Public logs for ad-integrated LLM responses are not yet available, while prior studies often rely on qualitative examples or small-scale tests that do not reflect the scale and complexity of real-world industrial ad auctions.

To bridge this gap, we build a simulation environment, inspired by the LLM-as-a-Judge paradigm [17], that covers the full ad-serving and feedback pipeline for LLM applications such as chatbots and AI search. As shown in Figure 3, our pipeline consists of two modules: First, the Ad-LLM generates ad-integrated responses under the evaluated mechanism; for LLM-Auction, this Ad-LLM is the trained allocation model itself. Second, the User-LLM generates personalized queries and returns click feedback conditioned on the user profile, query, and generated response.

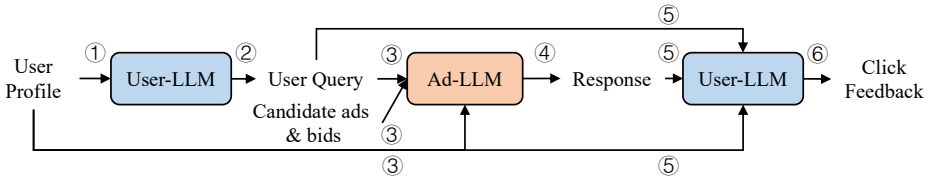

Figure 3: Simulation environment for LLM-native advertising. The User-LLM generates personalized queries and click feedback, while the Ad-LLM produces ad-integrated responses under the evaluated mechanism. Numbers indicate data flow.

The pipeline involves three stages: (i) **User query generation:** the User-LLM generates user queries based on user profiles containing demographic information and interests. (ii) **Ad-integrated response generation:** conditioned on the user query, profile, and *all* candidate ads with bids, the Ad-LLM generates a response that integrates clickable ads according to the mechanism. [1] (iii) **User click feedback:** the User-LLM simulates clicks and returns structured feedback for every inserted ad. Appendix E provides an example, prompt templates, and configuration details of the environment.

User-LLM feedback inevitably reflects a biased simulator distribution. We therefore view the environment as a controlled sandbox, not a substitute for real user feedback. This limitation does not couple LLM-Auction to the simulator: the method only requires click-based feedback and can optimize against any supplied click distribution. Once online logs from real users become available, the same pipeline can train directly on those authentic feedback signals. A small-scale human validation shows that the User-LLM's click decisions agree with human judgments well above chance (Appendix G.2), and we empirically verify the robustness of LLM-Auction against different user feedback distributions in Appendix G.1.

[1] To highlight the ad and facilitate its recognition by the User-LLM, we adopt the insertion format `@ad_title@[ad_id]`. Currently, we restrict the Ad-LLM from modifying ad titles to ensure faithfulness, leaving unconstrained advertising to future work.

5.2 Experimental Setup

5.2.1 Dataset

We construct 3k synthetic user profiles with demographic information and interests. The User-LLM generates five queries per profile, yielding 15k queries. We keep this query set fixed across experiments and split it into 12k training queries and 3k test queries. The candidate ad set contains 100 real-world advertisements from a leading e-commerce platform, spanning 50 categories.

5.2.2 Baselines

Following the protocol used by learning-based mechanisms [18, 20, 41, 44], we evaluate all methods on fixed single-round bids. We compare LLM-Auction with the pre-trained LLM, RAG-Auction [15], MOSAIC [31], and LLM-Auction (Oracle), an upper-bound variant using direct feedback. See Appendix F.2 for detailed evaluation protocol and implementation details of all baselines.

5.2.3 Implementation Details of LLM-Auction

Reward design. Following Eq. (5), we instantiate $\hat{f}_i^{\psi}$ as a predicted click-through rate (pCTR) model, while $\Omega_{\text{resp}}(y)$ penalizes excessive ad insertion and formatting errors. This lightweight reward isolates allocation efficiency, but the same interface can incorporate ad fidelity, credibility, hallucination, bias, or trust signals. Appendix F details the pCTR and reward models.

Training protocol. We train Qwen3-4B-Instruct [37] with IRPO. During each reward-model update stage, we use the previous epoch's Ad-LLM to sample 10 responses per query, and collect the User-LLM feedback to form $\mathcal{D}_{\text{on}}$ for the pCTR model updates. During each LLM update stage, we sample 5 responses per training query, score them with the updated reward model, and optimize the Ad-LLM. Each prompt of the Ad-LLM contains the full candidate ad set, and bids are resampled uniformly from positive integers in $[1, 100]$ at every epoch to avoid overfitting to static bid patterns. Appendix F.1.3 provides additional implementation details and hyperparameters.

5.2.4 Metrics

We evaluate mechanisms on held-out test queries $\mathcal{D}_{\text{test}}$ with independently sampled bid profiles. Since click events trigger realized payments in our simulator, let z_i^k, c_i^k, b_i^k, and p_i^k denote the impression indicator, click indicator, bid, and payment for advertiser i on sample k, respectively, with $p_i^k = b_i^k \cdot c_i^k$. We compute the following allocation efficiency metrics:

- Revenue per Query: $\frac{1}{|\mathcal{D}_{\text{test}}|} \sum_{k \in \mathcal{D}_{\text{test}}} \sum_{i=1}^{n} p_i^k$.
- Reward per Query: $\frac{1}{|\mathcal{D}_{\text{test}}|} \sum_{k \in \mathcal{D}_{\text{test}}} \mathcal{R}(s^k, \boldsymbol{b}^k, y^k)$.

5.3 Comparison of Allocation Efficiency

Table 1: Comparison of allocation efficiency in the LLM-native advertising simulation environment. Values are reported as mean $\pm$ std over 10 disjoint groups, each containing one sampled response per test query, except for MOSAIC, which uses 3 independent full runs. CTR is computed as total clicks divided by total impressions, and CPC (cost-per-click) as total realized click payments divided by total clicks.

Method	Revenue/Q ↑	Reward/Q ↑	CTR ↑	CPC ↑	Inserted Ads/Q ↓
Pre-trained LLM	70.37±19.35	-217.97±64.00	0.4424±0.0384	36.79±1.00	4.42±0.12
RAG-Auction	82.15±12.07	-20.87±22.02	0.4385±0.1011	62.87±1.04	2.98±0.01
MOSAIC	106.56±2.48	-20.69±16.73	0.5545±0.0133	57.68±2.08	3.33±0.12
LLM-Auction	170.86±10.13	81.86±7.68	0.7617±0.0437	91.84±0.27	2.44±0.05
LLM-Auction (Oracle)	223.30±5.39	125.24±6.47	0.9169±0.0051	92.85±0.39	2.62±0.07

Table 1 compares LLM-Auction with baselines and reveals the source of its allocation gains. Compared to the previous state-of-the-art method, MOSAIC, LLM-Auction boosts revenue by 60.3% and improves the reward from -20.69 to 81.86, without additional inference cost. The improvement

is not driven by showing more ads: the number of inserted ads drops to 2.44, below all baselines. Instead, LLM-Auction raises the CTR from 0.5545 to 0.7617 relative to MOSAIC, indicating better query-ad relevance and native presentation; concurrently, it raises the CPC from 57.68 to 91.84, indicating stronger alignment with advertiser bids. On the main efficiency metrics, revenue and reward, the error bars further show that (i) the lower bound of LLM-Auction exceeds the upper bound of all baselines, so the improvements are robust to response-sampling variability; and (ii) relative to the pre-trained LLM, LLM-Auction reduces the standard deviation, indicating that training makes the mechanism more stable. Furthermore, the stronger oracle variant results highlight substantial room for LLM-Auction with better reward estimation.

5.4 Verification of Mechanism Properties

Beyond allocation efficiency, we quantitatively verify properties of LLM-Auction's allocation rule.

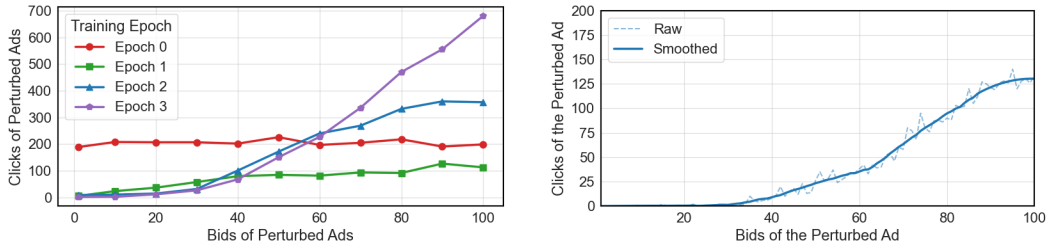

Figure 4: Experimental results of mechanism properties. **(a)** *Allocation monotonicity:* The allocation monotonicity of LLM-Auction gradually increases during training. **(b)** *Allocation continuity:* With a fixed query and perturbed ad, the allocation results change continuously as the bid increases.

Allocation monotonicity. We perturb bids on $\mathcal{D}_{\text{test}}$. For each sample, we randomly select one candidate ad, vary its bid over $\{1, 10, 20, \ldots, 100\}$ while fixing other bids, and measure User-LLM clicks on the perturbed ad across LLM-Auction training epochs. Figure 4(a) shows that the pre-trained LLM has no meaningful bid-click correlation. As IRPO progresses, this correlation strengthens; after 3 epochs, higher bids consistently yield higher click rates. Thus, LLM-Auction learns increasingly monotone allocation behavior even before reaching the optimal allocation assumed by Theorem 4.3.

Allocation continuity. We next test whether allocation changes smoothly with bids under fixed context. For one unseen query and one candidate ad, we sweep the ad's bid from 1 to 100, sample 200 responses per bid, and count User-LLM clicks on the perturbed ad. Figure 4(b) shows a gradual response rather than the step-function behavior of discrete slot auctions, supporting Theorem 4.4 and the first-price payment rule. Since deployed bids use finite precision, this result should be read as local smoothness over the practical bid grid: minimum-unit bid perturbations induce bounded changes in expected allocation.

The increasingly monotone bid-click relationship over training epochs in Figure 4(a) and the steadily increasing allocation for a fixed query in Figure 4(b) together provide empirical evidence that the trained model approximately satisfies VM-IC.

5.5 Ablation Study

We ablate the pCTR training signal to assess how the full response context and data scale affect LLM-Auction. We measure pCTR accuracy by Group AUC (GAUC), the weighted average of per-user AUC scores; Appendix F.3 gives implementation details of the ablation study.

Importance of response context. Table 2 shows that removing the response context drastically reduces revenue by *55%* and drives the reward below zero. Thus, leveraging full response context in LLM-Auction is essential for estimating click probability in the LLM-native advertising setting and for optimizing allocation quality.

Table 2: Ablation experiment results of LLM-Auction.

Variant	Revenue	Reward	GAUC
LLM-Auction	170.86	81.86	0.6872
LLM-Auction (50% samples)	130.21	60.79	0.6705
LLM-Auction (w/o response)	76.45	-44.30	0.6391

Prediction accuracy and mechanism performance. Across ablations, higher GAUC coincides with higher revenue and reward. Furthermore, the oracle variant in Table 1, which approximates an ideal pCTR model, achieves the strongest revenue, reward, CTR, and CPC, confirming that better click prediction directly strengthens the performance of LLM-Auction.

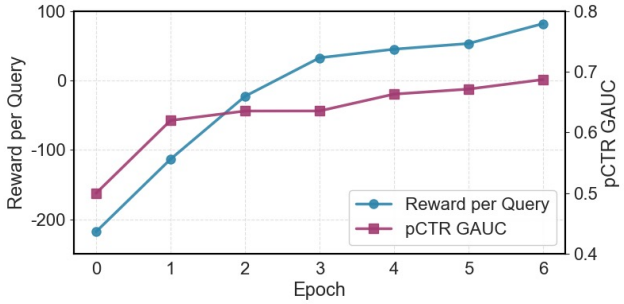

Figure 5: Learning curve of LLM-Auction during IRPO epochs.

Effectiveness of IRPO. As shown in Figure 5, during the iterative training of LLM-Auction, both the mechanism performance and the GAUC of the pCTR model improve steadily over IRPO epochs. This indicates that our IRPO algorithm progressively narrows the gap between the current allocation and the optimal allocation distribution, enhances the reward estimation, and consequently achieves better performance.

5.6 Additional Analyses

Appendix G.1 shows that LLM-Auction remains effective under shifted user feedback distributions, supporting that the method is not tied to a specific User-LLM simulator. Appendix G.3 further shows zero-shot transfer to unseen ad pools and user populations, suggesting that the learned allocation behavior is not simple memorization of training ads or user profiles. Appendix G.4 indicates that stronger base models further improve allocation efficiency, demonstrating the scaling capabilities of LLM-Auction. We also qualitatively compare the ad-integrated responses generated before and after LLM-Auction training in Appendix H. Together, these analyses suggest that LLM-Auction learns transferable allocation behavior grounded in semantic relevance, response context, and bid signals.

6 Limitations and Future Work

This work is an initial step toward learning-based mechanisms for LLM-native advertising. LLM-Auction shows that an LLM can internalize allocation objectives, user-feedback signals, and platform constraints, opening a new design space for mechanisms that operate in AI-generated content. At the same time, our work also has the following limitations: our evaluation uses an LLM-based simulator; the current reward design and training pipeline are still simple; and the first-price payment is not DSIC for utility maximizers, with incentive guarantees established only for the optimal allocation rule under two advertiser behavior models. These limitations point to clear future directions: larger and more realistic benchmarks, richer reward models for trust and safety, online optimization from real feedback, computationally feasible incentive-compatible payment rules, and scaling pipelines for large ad inventories. Appendix I provides a detailed discussion.

Acknowledgments and Disclosure of Funding

This work received no third-party funding or support. All authors are employees of Taobao & Tmall Group of Alibaba and declare no other competing interests.

References

[1] Santiago Balseiro, Kshipra Bhawalkar, Yuan Deng, Zhe Feng, Jieming Mao, Aranyak Mehta, Vahab Mirrokni, Renato Paes Leme, Di Wang, and Song Zuo. Position auctions in ai-generated content. *arXiv preprint arXiv:2506.03309*, 2025.

[2] Dirk Bergemann, Marek Bojko, Paul Duetting, Renato Paes Leme, Haifeng Xu, and Song Zuo. Data-driven mechanism design: Jointly eliciting preferences and information. In *Proceedings of the 26th ACM Conference on Economics and Computation*, EC '25, page 507, 2025.

[3] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 1877–1901, 2020.

[4] Ioannis Caragiannis, Christos Kaklamanis, Panagiotis Kanellopoulos, Maria Kyropoulou, Brendan Lucier, Renato Paes Leme, and Éva Tardos. Bounding the inefficiency of outcomes in generalized second price auctions. *Journal of Economic Theory*, 156:343–388, 2015.

[5] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017.

[6] Avinava Dubey, Zhe Feng, Rahul Kidambi, Aranyak Mehta, and Di Wang. Auctions with llm summaries. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, page 713–722, 2024.

[7] Paul Dütting, Vahab Mirrokni, Renato Paes Leme, Haifeng Xu, and Song Zuo. Mechanism design for large language models. In *Proceedings of the ACM Web Conference 2024*, WWW '24, page 144–155, 2024.

[8] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97(1):242–259, March 2007.

[9] Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. Kto: Model alignment as prospect theoretic optimization. *arXiv preprint arXiv:2402.01306*, 2024.

[10] Zhikang Fan, Lan Hu, Ruirui Wang, Zhongrui Ma, Yue Wang, Qi Ye, and Weiran Shen. Two-stage auction design in online advertising. In *Proceedings of the ACM on Web Conference 2025*, WWW '25, page 3571–3585, 2025.

[11] Soheil Feizi, MohammadTaghi Hajiaghayi, Keivan Rezaei, and Suho Shin. Online advertisements with llms: Opportunities and challenges. *arXiv preprint arXiv:2311.07601*, 2023.

[12] Songyang Gao, Qiming Ge, Wei Shen, Shihan Dou, Junjie Ye, Xiao Wang, Rui Zheng, Yicheng Zou, Zhi Chen, Hang Yan, Qi Zhang, and Dahua Lin. Linear alignment: A closed-form solution for aligning human preferences without tuning and feedback. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235, pages 14702–14722, 2024.

[13] Mohammad Gheshlaghi Azar, Zhaohan Daniel Guo, Bilal Piot, Remi Munos, Mark Rowland, Michal Valko, and Daniele Calandriello. A general theoretical paradigm to understand learning from human preferences. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 238, pages 4447–4455, 2024.

[14] Irving L Glicksberg. A further generalization of the kakutani fixed point theorem, with application to nash equilibrium points. *Proceedings of the American Mathematical Society*, 3(1): 170–174, 1952.

[15] MohammadTaghi Hajiaghayi, Sébastien Lahaie, Keivan Rezaei, and Suho Shin. Ad auctions for llms via retrieval augmented generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pages 18445–18480, 2024.

[16] Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*, 2023.

[17] Dawei Li, Bohan Jiang, Liangjie Huang, Alimohammad Beigi, Chengshuai Zhao, Zhen Tan, Amrita Bhattacharjee, Yuxuan Jiang, Canyu Chen, Tianhao Wu, Kai Shu, Lu Cheng, and Huan Liu. From generation to judgment: Opportunities and challenges of LLM-as-a-judge. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025.

[18] Ningyuan Li, Yunxuan Ma, Yang Zhao, Zhijian Duan, Yurong Chen, Zhilin Zhang, Jian Xu, Bo Zheng, and Xiaotie Deng. Learning-based ad auction design with externalities: The framework and a matching-based approach. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, page 1291–1302, 2023.

[19] Fei Liu, Xinyu Lin, Hanchao Yu, Mingyuan Wu, Jianyu Wang, Qiang Zhang, Zhuokai Zhao, Yinglong Xia, Yao Zhang, Weiwei Li, et al. Recoworld: Building simulated environments for agentic recommender systems. *arXiv preprint arXiv:2509.10397*, 2025.

[20] Xiangyu Liu, Chuan Yu, Zhilin Zhang, Zhenzhe Zheng, Yu Rong, Hongtao Lv, Da Huo, Yiqing Wang, Dagui Chen, Jian Xu, Fan Wu, Guihai Chen, and Xiaoqiang Zhu. Neural auction: End-to-end learning of auction mechanisms for e-commerce advertising. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, page 3354–3364, 2021.

[21] Paul Longo. Transforming audience engagement with generative ai, 2024. URL https://about.ads.microsoft.com/en/blog/post/october-2024/transforming-audience-engagement-with-generative-ai. Last accessed 10 November 2025.

[22] Yu Meng, Mengzhou Xia, and Danqi Chen. Simpo: Simple preference optimization with a reference-free reward. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pages 124198–124235, 2024.

[23] Tommy Mordo, Moshe Tennenholtz, and Oren Kurland. Sponsored question answering. In *Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval*, ICTIR '24, page 167–173, 2024.

[24] Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981. doi: 10.1287/moor.6.1.58.

[25] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 27730–27744, 2022.

[26] Perplexity. Why we're experimenting with advertising, 2024. URL https://www.perplexity.ai/hub/blog/why-we-re-experimenting-with-advertising. Last accessed 10 November 2025.

[27] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pages 53728–53741, 2023.

[28] Elizabeth Reid. Supercharging search with generative ai, 2023. URL https://blog.google/products/search/generative-ai-search/. Last accessed 10 November 2025.

[29] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[30] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

[31] Ermis Soumalias, Michael Curry, and Sven Seuken. Truthful aggregation of LLMs with an application to online advertising. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2025.

[32] Peter Tsai. When will chatgpt replace search? maybe sooner than you think, 2023. URL https://www.pcmag.com/news/when-will-chatgpt-replace-search-engines-maybe-sooner-than-you-think. Last accessed 11 November 2025.

[33] Hal R. Varian. Position auctions. *International Journal of Industrial Organization*, 25(6): 1163–1178, 2007.

[34] Lei Wang, Jingsen Zhang, Hao Yang, Zhi-Yuan Chen, Jiakai Tang, Zeyu Zhang, Xu Chen, Yankai Lin, Hao Sun, Ruihua Song, Xin Zhao, Jun Xu, Zhicheng Dou, Jun Wang, and Ji-Rong Wen. User behavior simulation with large language model-based agents. *ACM Trans. Inf. Syst.*, 43(2), 2025. ISSN 1046-8188. doi: 10.1145/3708985.

[35] Yiqing Wang, Xiangyu Liu, Zhenzhe Zheng, Zhilin Zhang, Miao Xu, Chuan Yu, and Fan Wu. On designing a two-stage auction for online advertising. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 90–99, 2022.

[36] Christopher A. Wilkens, Ruggiero Cavallo, and Rad Niazadeh. Gsp: The cinderella of mechanism design. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, page 25–32, 2017. ISBN 9781450349130. doi: 10.1145/3038912.3052687.

[37] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

[38] Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, YuYue, Weinan Dai, Tiantian Fan, Gaohong Liu, Juncai Liu, LingJun Liu, Xin Liu, Haibin Lin, Zhiqi Lin, Bole Ma, Guangming Sheng, Yuxuan Tong, Chi Zhang, Mofan Zhang, Ru Zhang, Wang Zhang, Hang Zhu, Jinhua Zhu, Jiaze Chen, Jiangjie Chen, Chengyi Wang, Hongli Yu, Yuxuan Song, Xiangpeng Wei, Hao Zhou, Jingjing Liu, Wei-Ying Ma, Ya-Qin Zhang, Lin Yan, Yonghui Wu, and Mingxuan Wang. DAPO: An open-source LLM reinforcement learning system at scale. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2025.

[39] Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. Qwen3 embedding: Advancing text embedding and reranking through foundation models. *arXiv preprint arXiv:2506.05176*, 2025.

[40] Zhaowei Zhang, Fengshuo Bai, Qizhi Chen, Chengdong Ma, Mingzhi Wang, Haoran Sun, Zilong Zheng, and Yaodong Yang. Amulet: Realignment during test time for personalized preference adaptation of LLMs. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

[41] Zhilin Zhang, Xiangyu Liu, Zhenzhe Zheng, Chenrui Zhang, Miao Xu, Junwei Pan, Chuan Yu, Fan Wu, Jian Xu, and Kun Gai. Optimizing multiple performance metrics with deep gsp auctions for e-commerce advertising. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, WSDM '21, page 993–1001, 2021.

[42] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.

[43] Chujie Zheng, Shixuan Liu, Mingze Li, Xiong-Hui Chen, Bowen Yu, Chang Gao, Kai Dang, Yuqiong Liu, Rui Men, An Yang, et al. Group sequence policy optimization. *arXiv preprint arXiv:2507.18071*, 2025.

[44] Ruitao Zhu, Yangsu Liu, Dagui Chen, Zhenjia Ma, Chufeng Shi, Zhenzhe Zheng, Jie Zhang, Jian Xu, Bo Zheng, and Fan Wu. Contextual generative auction with permutation-level externalities for online advertising. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '25, page 2171–2181, 2025.

A Optimization and Training Procedures

Algorithm 1 Iterative Reward-Preference Optimization (IRPO)

Input: Initial LLM parameters θ_0, initial prediction model parameters ψ_0, number of sampled responses M, DPO pair threshold δ_{th}, learning rates γ, η.

Output: Optimized LLM parameters θ_T and prediction model parameters ψ_T.

for Epoch $t = 1, 2, \ldots, T$ **do**
  // Phase 1: Reward Model Update
  Deploy current LLM $\pi_{\theta_{t-1}}$ and construct online dataset $\mathcal{D}^t_{\text{on}}$
  Initialize: $\psi_t \leftarrow \psi_{t-1}$
  for each sample $(s, y, \boldsymbol{c}) \in \mathcal{D}^t_{\text{on}}$ **do**
    Compute BCE loss in Eq. (4): $\mathcal{L}_{\text{BCE}}$
    Update reward model parameters: $\psi_t \leftarrow \psi_t - \gamma \nabla_{\psi_t} \mathcal{L}_{\text{BCE}}$
  end for
  // Phase 2: LLM Update
  Construct offline dataset $\mathcal{D}^t_{\text{off}}$ from $\mathcal{D}^t_{\text{on}}$ or synthetic data
  Initialize: $\theta_t \leftarrow \theta_{t-1}$
  for each sample $(s, \boldsymbol{b}) \in \mathcal{D}^t_{\text{off}}$ **do**
    Sample M responses: $\{y^m\}_{m=1}^M \sim \pi_{\theta_{t-1}}(\cdot \mid s, \boldsymbol{b})$
    Compute rewards using Eq. (5) with $\mathcal{R}_{\psi_t}$: $r^m = \mathcal{R}_{\psi_t}(s, \boldsymbol{b}, y^m), \forall m \in [M]$
    Select winner: $y^w \leftarrow y^{\arg\max_m r^m}$
    Construct losing set: $\mathcal{K} \leftarrow \{y^l \mid r^w - r^l > \delta_{\text{th}}\}$
    Compute DPO loss in Eq. (6): $\mathcal{L}_{\text{DPO}}$
    Update LLM parameters: $\theta_t \leftarrow \theta_t - \eta \nabla_{\theta_t} \mathcal{L}_{\text{DPO}}$
  end for
end for

B Deployment Framework

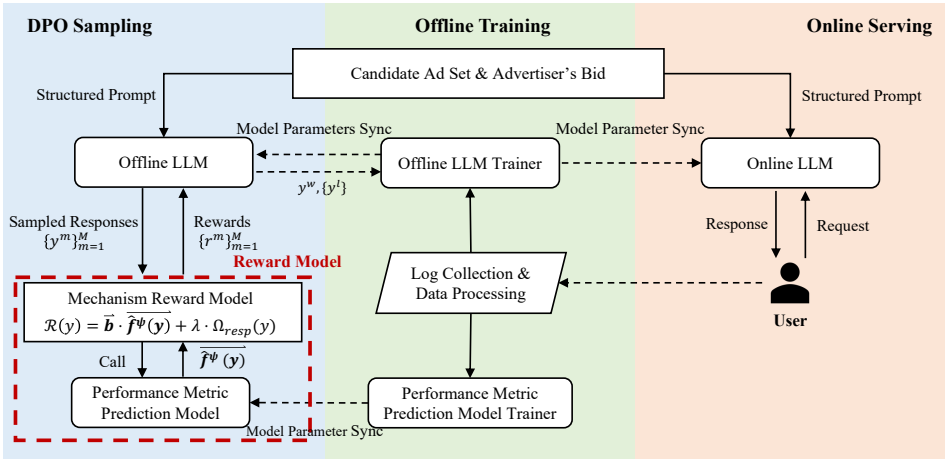

Figure 6: Deployment framework of LLM-Auction.

Figure 6 illustrates a feasible deployment framework for LLM-Auction in industrial settings. During online serving, the LLM trained by LLM-Auction directly generates ad-integrated responses based on user queries, candidate ads, and bids. The user queries and corresponding feedback are collected to: (i) update the performance metric prediction model, and (ii) provide diverse queries for offline DPO sampling. At regular intervals, LLM-Auction performs an offline update and deploys the updated LLM online.

C Proofs of Theoretical Results

C.1 Proof of Theorem 4.3

Theorem C.1 (Restatement of Theorem 4.3)**.** *The optimal allocation π_{θ^*} that maximizes the bid-based mechanism objective $J(\pi_\theta, \boldsymbol{b}, \boldsymbol{b}, s)$ satisfies allocation monotonicity. Formally, for any advertiser i, fixed context s, and competing bids $\boldsymbol{b}_{-i}$, if $b_i' \geq b_i$, then:*

$$f_i(\pi_{\theta^*}(\cdot \mid s, b_i', \boldsymbol{b}_{-i}), s) \geq f_i(\pi_{\theta^*}(\cdot \mid s, b_i, \boldsymbol{b}_{-i}), s).$$

Proof. Fix s and $\boldsymbol{b}_{-i}$. Let $\mathcal{Q}$ denote the common feasible allocation class. For $q \in \mathcal{Q}$, define

$$F_j(q) := f_j(q, s)$$

and

$$G(q) := \sum_{j \neq i} b_j F_j(q) + \lambda \mathbb{E}_{y \sim q}[\Omega_{\text{resp}}(y)] - \beta \mathcal{D}_{\text{KL}}\big(q \| \pi_{\theta_0}(\cdot \mid s)\big).$$

By Assumption 4.2, $G(q)$ is independent of advertiser i's bid. Thus, at bid t, the bid-based objective takes the form

$$J_t(q) = tF_i(q) + G(q).$$

Consider $b_i' > b_i$, and write

$$q := \pi_{\theta^*}(\cdot \mid s, b_i, \boldsymbol{b}_{-i}), \qquad q' := \pi_{\theta^*}(\cdot \mid s, b_i', \boldsymbol{b}_{-i}).$$

Since both distributions belong to the same feasible class, their respective optimality conditions imply

$$b_i F_i(q) + G(q) \geq b_i F_i(q') + G(q'),$$
$$b_i' F_i(q') + G(q') \geq b_i' F_i(q) + G(q).$$

Adding these inequalities yields

$$(b_i' - b_i)\big(F_i(q') - F_i(q)\big) \geq 0.$$

Since $b_i' - b_i > 0$, we obtain $F_i(q') \geq F_i(q)$. When $b_i' = b_i$, the bid profiles are identical and the claimed inequality holds with equality. □

C.2 Proof of Theorem 4.4

Theorem C.2 (Restatement of Theorem 4.4)**.** *In the LLM-native advertising setting, under Assumption 4.2, the expected performance $f_i(\pi_{\theta^*}(\cdot \mid s, \boldsymbol{b}), s)$ is a continuous function of $\boldsymbol{b}$ for any advertiser i and context s.*

Proof. Fix a context s and bid profile $\boldsymbol{b}$. The bid-based form of Eq. (3) is

$$J(\pi_\theta, \boldsymbol{b}, \boldsymbol{b}, s) = \mathbb{E}_{y \sim \pi_\theta(\cdot \mid s, \boldsymbol{b})}\left[\sum_{i=1}^{n} b_i \cdot f_i(y, s) + \lambda \cdot \Omega_{\text{resp}}(y)\right] - \beta \cdot \mathcal{D}_{\text{KL}}(\pi_\theta(\cdot \mid s, \boldsymbol{b}) \| \pi_{\theta_0}(\cdot \mid s)),$$

Under Assumption 4.1, the learned predictor gives an unbiased estimate of the expected performance term in this objective. Thus, in the theoretical large-sample regime, the reward model $\mathcal{R}(s, \boldsymbol{b}, y)$ in Eq. (5) is an unbiased surrogate for the bid-weighted response reward, and the KL-regularized objective can be written as

$$J(\pi_\theta, \boldsymbol{b}, \boldsymbol{b}, s) = \mathbb{E}_{y \sim \pi_\theta(\cdot \mid s, \boldsymbol{b})}[\mathcal{R}(s, \boldsymbol{b}, y)] - \beta \cdot \mathcal{D}_{\text{KL}}(\pi_\theta(\cdot \mid s, \boldsymbol{b}) \| \pi_{\theta_0}(\cdot \mid s)).$$

The optimizer of this KL-regularized reward objective has the Gibbs form used in DPO-style preference optimization [27]:

$$\pi_{\theta^*}(y \mid s, \boldsymbol{b}) = \frac{1}{Z(s, \boldsymbol{b})} \pi_{\theta_0}(y \mid s, \boldsymbol{b}) \exp\left(\frac{1}{\beta} \mathcal{R}(s, \boldsymbol{b}, y)\right), \tag{9}$$

where $Z(s, \boldsymbol{b}) = \sum_y \pi_{\theta_0}(y \mid s, \boldsymbol{b}) \exp\left(\frac{1}{\beta}\mathcal{R}(s, \boldsymbol{b}, y)\right)$ is the partition function. Consequently, the expected performance for advertiser i under the optimal policy π_{θ^*} is:

$$f_i(\pi_{\theta^*}(\cdot \mid s, \boldsymbol{b}), s) = \sum_y \pi_{\theta^*}(y \mid s, \boldsymbol{b})\, f_i(y, s)$$
$$= \frac{\sum_y \pi_{\theta_0}(y \mid s, \boldsymbol{b})\, \exp\left(\frac{1}{\beta}\mathcal{R}(s, \boldsymbol{b}, y)\right) f_i(y, s)}{\sum_y \pi_{\theta_0}(y \mid s, \boldsymbol{b})\, \exp\left(\frac{1}{\beta}\mathcal{R}(s, \boldsymbol{b}, y)\right)}. \tag{10}$$

Recall that the reward function defined in Eq. (5),

$$\mathcal{R}(s, \boldsymbol{b}, y) = \sum_{i=1}^{n} b_i \cdot \hat{f}_i^{\psi}(y, s) + \lambda \cdot \Omega_{\text{resp}}(y),$$

is linear—and hence continuous—in the bid profile $\boldsymbol{b}$. Under Assumption 4.2, the reference distribution satisfies $\pi_{\theta_0}(y \mid s, \boldsymbol{b}) = \pi_{\theta_0}(y \mid s)$ for every bid profile $\boldsymbol{b} \in [0, \bar{b}]^n$. Thus, for each fixed y, it is constant in $\boldsymbol{b}$. Since $\mathcal{R}(s, \boldsymbol{b}, y)$ is continuous in $\boldsymbol{b}$, the map

$$\boldsymbol{b} \mapsto \pi_{\theta_0}(y \mid s) \exp\left(\frac{1}{\beta}\mathcal{R}(s, \boldsymbol{b}, y)\right)$$

is continuous on $[0, \bar{b}]^n$. Since decoding uses a finite vocabulary and a bounded maximum output length, the response space considered by the model is finite. Because $f_i(y, s)$ does not depend on $\boldsymbol{b}$, we conclude:

- The numerator of Eq. (10) is a finite sum of continuous functions of $\boldsymbol{b}$, hence continuous in $\boldsymbol{b}$.
- The denominator of Eq. (10) is also a finite sum of continuous functions of $\boldsymbol{b}$, and is strictly positive for all $\boldsymbol{b} \in [0, \bar{b}]^n$, because $\exp(\cdot) > 0$ and $\pi_{\theta_0}(\cdot)$ is a probability distribution (thus nonnegative and not identically zero).

Therefore, the ratio in Eq. (10) defines a continuous function of $\boldsymbol{b}$ on the compact set $[0, \bar{b}]^n$. □

C.3 Proof of Theorem 4.5

Theorem C.3 (Restatement of Theorem 4.5)**.** *Under the optimal allocation rule π_{θ^*} with compact bid spaces, the first-price bidding game among UMs admits a mixed-strategy Nash equilibrium. At any such equilibrium $\sigma^\star$, every interior support bid b_i with $\overline{V}_i(b_i) > 0$ satisfies*

$$b_i = v_i - \beta \frac{\overline{A}_i(b_i)}{\overline{V}_i(b_i)}.$$

Here, $\overline{A}_i(b_i) := \mathbb{E}_{\boldsymbol{b}_{-i} \sim \sigma^\star_{-i}}[A_i(b_i, \boldsymbol{b}_{-i}; s)]$ and $\overline{V}_i(b_i) := \mathbb{E}_{\boldsymbol{b}_{-i} \sim \sigma^\star_{-i}}[\mathrm{Var}_{y \sim \pi_{\theta^*}(\cdot \mid s, b_i, \boldsymbol{b}_{-i})}[f_i(y, s)]]$ are advertiser i's expected allocation and expected performance variance when it bids b_i against $\sigma^\star_{-i}$.

Proof. Fix s and $\boldsymbol{v}$ throughout the proof.

Equilibrium existence. Each advertiser's pure strategy space $B_i = [0, \bar{b}]$ is a nonempty compact metric space. By Theorem 4.4, the effective allocation $A_i(\boldsymbol{b}; s)$ is continuous on $B := \prod_{j=1}^n B_j$. Consequently,

$$u_i(\boldsymbol{b}; s, v_i) = (v_i - b_i) A_i(\boldsymbol{b}; s)$$

is continuous on B. Glicksberg's theorem [14] therefore implies the existence of a mixed-strategy Nash equilibrium.

Allocation sensitivity under the optimal policy. Write the optimal Gibbs policy as

$$\pi_{\theta^*}(y \mid s, \boldsymbol{b}) = \frac{\pi_{\theta_0}(y \mid s)\exp\big(\mathcal{R}(s, \boldsymbol{b}, y)/\beta\big)}{Z(s, \boldsymbol{b})},$$

where

$$\mathcal{R}(s, \boldsymbol{b}, y) = \sum_{j=1}^{n} b_j f_j(y, s) + \lambda \Omega_{\text{resp}}(y)$$

is the per-response reward of the bid-based objective $J(\pi_\theta, \boldsymbol{b}, \boldsymbol{b}, s)$, and

$$Z(s, \boldsymbol{b}) = \sum_{y} \pi_{\theta_0}(y \mid s)\exp\big(\mathcal{R}(s, \boldsymbol{b}, y)/\beta\big).$$

The partition function is strictly positive. Because the response space is finite and neither the reference distribution (Assumption 4.2) nor the response-quality term depends on bids, we may differentiate the finite sums term by term:

$$\frac{\partial \log Z(s, \boldsymbol{b})}{\partial b_i} = \frac{1}{\beta}\sum_{y} \pi_{\theta^*}(y \mid s, \boldsymbol{b}) f_i(y, s) = \frac{1}{\beta} A_i(\boldsymbol{b}; s). \tag{11}$$

Differentiating the policy consequently gives

$$\frac{\partial \pi_{\theta^*}(y \mid s, \boldsymbol{b})}{\partial b_i} = \frac{1}{\beta}\pi_{\theta^*}(y \mid s, \boldsymbol{b})\big(f_i(y, s) - A_i(\boldsymbol{b}; s)\big).$$

It follows that

$$\begin{aligned}
\frac{\partial A_i(\boldsymbol{b}; s)}{\partial b_i} &= \sum_{y} f_i(y, s)\frac{\partial \pi_{\theta^*}(y \mid s, \boldsymbol{b})}{\partial b_i} \\
&= \frac{1}{\beta}\left(\mathbb{E}_{y \sim \pi_{\theta^*}(\cdot \mid s, \boldsymbol{b})}[f_i(y, s)^2] - A_i(\boldsymbol{b}; s)^2\right) \\
&= \frac{1}{\beta} V_i(\boldsymbol{b}; s).
\end{aligned}$$

Best responses against equilibrium opponents. Fix any mixed-strategy Nash equilibrium $\sigma^\star$, and let $\sigma^\star_{-i} = \bigotimes_{j \neq i} \sigma^\star_j$. Against these opponents, advertiser i's expected utility from a pure bid is

$$\overline{u}_i(b_i) = \mathbb{E}_{\boldsymbol{b}_{-i} \sim \sigma^\star_{-i}}[u_i(b_i, \boldsymbol{b}_{-i}; s, v_i)] = (v_i - b_i)\overline{A}_i(b_i).$$

The function $\overline{u}_i$ is continuous on the compact interval B_i, so it attains a maximum m_i. The Nash equilibrium condition implies

$$\int_{B_i} \overline{u}_i(b_i)\, d\sigma^\star_i(b_i) = m_i,$$

since a deviation to a maximizing pure bid is feasible, and no mixture can yield more than m_i. If a support bid had utility strictly below m_i, continuity would give a neighborhood with utility bounded strictly below m_i. That neighborhood would have positive $\sigma^\star_i$-probability, contradicting the displayed equality. Hence

$$\operatorname{supp}(\sigma^\star_i) \subseteq \arg\max_{b_i \in B_i} \overline{u}_i(b_i).$$

Interior equilibrium bids. Let $M_i := \max_y f_i(y, s) < \infty$. The allocation derivative satisfies the uniform bound

$$0 \leq \frac{\partial A_i(\boldsymbol{b}; s)}{\partial b_i} = \frac{V_i(\boldsymbol{b}; s)}{\beta} \leq \frac{M_i^2}{\beta}.$$

This bound justifies interchanging differentiation and expectation by dominated convergence. Holding $\sigma^\star_{-i}$ fixed under a unilateral deviation, we obtain

$$\frac{d\overline{A}_i(b_i)}{db_i} = \mathbb{E}_{\boldsymbol{b}_{-i} \sim \sigma^\star_{-i}}\left[\frac{\partial A_i(b_i, \boldsymbol{b}_{-i}; s)}{\partial b_i}\right] = \frac{1}{\beta}\overline{V}_i(b_i).$$

Every interior support bid is an interior maximizer of $\overline{u}_i$ and therefore satisfies

$$0 = \frac{d\overline{u}_i(b_i)}{db_i} = -\overline{A}_i(b_i) + \frac{v_i - b_i}{\beta}\overline{V}_i(b_i).$$

Whenever $\overline{V}_i(b_i) > 0$, rearranging yields

$$b_i = v_i - \beta\frac{\overline{A}_i(b_i)}{\overline{V}_i(b_i)}.$$

This establishes the claimed necessary condition. □

C.4 Proof of Theorem 4.6

Theorem C.4 (Restatement of Theorem 4.6)**.** *In the LLM-native advertising setting with the optimal allocation rule, the first-price payment rule is VM-IC under ROI constraint τ_i, where the optimal bidding strategy is $b_i = \tilde{v}_i = v_i/\tau_i$.*

Proof. Fix s and $\boldsymbol{b}_{-i}$, and write

$$F_i(b_i) := f_i(\pi_{\theta^*}(\cdot \mid s, (b_i, \boldsymbol{b}_{-i})), s).$$

By Theorem 4.3, $F_i(b_i)$ is weakly increasing in b_i. Under the first-price payment rule, advertiser i pays

$$p_i(b_i) = b_i F_i(b_i).$$

Case 1: $\tau_i = 1$ When $\tau_i = 1$, the ROI constraint is

$$b_i F_i(b_i) \le v_i F_i(b_i).$$

For any bid with $F_i(b_i) > 0$, feasibility therefore requires $b_i \le v_i$; bids with $F_i(b_i) = 0$ yield zero value and cannot improve on a feasible positive-value bid. Among feasible bids, the value-maximizing objective is $v_i F_i(b_i)$. Since F_i is weakly increasing, this objective is maximized by the largest feasible bid, namely $b_i = v_i$. Thus truthful value reporting is optimal when $\tau_i = 1$.

Case 2: General $\tau_i \ge 1$ For a general ROI target $\tau_i \ge 1$, the constraint becomes

$$\tau_i b_i F_i(b_i) \le v_i F_i(b_i).$$

Whenever $F_i(b_i) > 0$, feasibility requires $b_i \le v_i/\tau_i$; when $F_i(b_i) = 0$, the advertiser obtains zero value. Therefore, the advertiser again maximizes value by choosing the largest feasible bid. This bid is the effective value

$$b_i = \tilde{v}_i = \frac{v_i}{\tau_i}.$$

Equivalently, the type (v_i, τ_i) reduces to a normalized value-maximizer with effective value $\tilde{v}_i$, as in the standard ROI-constrained bidding analysis [36]. Hence bidding $b_i = \tilde{v}_i$ is a dominant strategy, establishing VM-IC under the ROI constraint. □

C.5 Closed Form of the DSIC Payment

Integrating Eq. (11) from the proof of Theorem 4.5 along the i-th coordinate from 0 to b_i yields

$$\int_0^{b_i} A_i(z, \boldsymbol{b}_{-i}; s)\, dz = \beta \log \frac{Z(s, (b_i, \boldsymbol{b}_{-i}))}{Z(s, (0, \boldsymbol{b}_{-i}))}.$$

Substituting it into the DSIC payment in Section 4.2 gives

$$P_i^{\mathrm{DSIC}}(\boldsymbol{b}; s) = b_i A_i(\boldsymbol{b}; s) - \beta \log \frac{Z(s, (b_i, \boldsymbol{b}_{-i}))}{Z(s, (0, \boldsymbol{b}_{-i}))}.$$

Exact evaluation therefore requires $n + 1$ partition functions per query, namely $Z(s, \boldsymbol{b})$ and $Z(s, (0, \boldsymbol{b}_{-i}))$ for each advertiser i, each a sum over the entire LLM response space.

D Practical Plausibility of Main Assumptions

Unbiased performance prediction. Assumption 4.1 is the large-sample condition that connects the learned reward model to the theoretical optimization objective. When sufficiently many feedback samples are collected from the allocation distribution being optimized, a calibrated performance predictor can estimate $\mathbb{E}[f_i(y, s)]$ without systematic bias; substituting $\hat{f}_i^{\psi}$ into Eq. (5) then gives an unbiased estimate of the bid-weighted performance term in Eq. (3). This is the regime captured by the theory. In finite data, LLM-Auction approximates this condition through IRPO, which repeatedly refreshes the pCTR model using feedback induced by the current allocation rule. The ablations in Section 5.5 show that higher pCTR GAUC is associated with better mechanism performance, and the oracle result in Table 1 quantifies the potential performance gain when the feedback signal becomes more accurate.

Bid independence of the reference LLM distribution. Assumption 4.2 treats the reference LLM as a bid-agnostic baseline. Because the pre-trained LLM has not yet learned how bids should shape its outputs, the bids in its prompt do not systematically change what it generates, consistent with its flat bid-click curve (Epoch 0) in Figure 4(a).

E Further Details of LLM-Native Advertising Simulation Environment

E.1 Example of LLM-Native Advertising Simulation Environment

Table 3: An example from our LLM-native advertising simulation environment. To highlight the clickable ads, we adopt the insertion format of `@ad_title@[ad_id]`. The relevant ad texts are marked with blue underlines. In practical applications, a user-friendly display style can be adopted.

User Profile:
A young male white-collar worker who enjoys yoga, coffee, LEGO building sets, and reading.

User Query:
Are there any beginner-friendly homemade coffee recipes that pair well with post-yoga relaxation?

Ad-Integrated Response:
For a relaxing post-yoga moment, a gentle and nuanced homemade coffee is an excellent choice. Try this simple yet delicious oat milk latte recipe:

- *Ingredients:* 1 shot (≈ 30 ml) espresso, 150 ml oat milk (store-bought oat milk works fine), a pinch of cinnamon (optional), a pinch of cocoa powder (optional).
- *Instructions:* Brew one shot of espresso using an espresso machine or Moka pot. Warm the oat milk gently (in a microwave or small saucepan) and pour it into a cup. Add the espresso and stir well. Finish with a dusting of cinnamon or cocoa powder to taste.

This coffee is not only delicious but also adds warmth and comfort to your post-yoga moments. For a richer flavor, try adding some vanilla syrup or coconut milk. The @Dutch-imported Oriental Floral Coffee Cup@ [Ad-4820] is the perfect companion for this coffee—its elegant design elevates the entire coffee experience with a sense of ritual and sophistication.

User Click Feedback:
[{“ad_id”:“Ad-4820”,“clicked":true,“reason":“The ad is relevant to the query; the recommended coffee cup suits the oat milk latte and aligns naturally with the post-yoga relaxation context."}]

Table 3 presents a concrete example from our proposed LLM-native advertising simulation environment, including the user profile, query, ad-integrated response generated by the Ad-LLM, and the final click feedback from the User-LLM. The user profile has been partially anonymized; actual profiles contain additional information such as geographic location and purchasing power level. Note that while Ad-LLM inserts only one ad in this example, our simulation environment does not impose predefined constraints on the number of ads to be inserted. Examples of LLM-Auction inserting multiple ads simultaneously can be found in Appendix H.

E.2 Prompt Templates in LLM-Native Advertising Simulation Environment

We provide the LLM prompt templates used in our simulated environment for LLM-native advertising to facilitate the reproducibility of our experimental pipeline. Template text is marked in red, while the placeholder variables for different samples are presented in black.

User query generation. In this phase, we provide user profile information to the User-LLM and instruct it to generate a specified number of user queries that are diverse and representative of different user intents. The prompt template for this stage is as follows:

```
# Role
You are a user of an AI assistant product, using it to acquire information,
seek product recommendations, or make shopping decisions. Below is your
personal information:
## Basic Information: {user_profile.info}
## Personal Interests: {user_profile.interests}
# Task
Based on your basic information and personal interests, ask the AI
assistant {num_queries} questions.
# Requirements
1. The questions should be distinct, addressing your different potential
needs, while having direct or indirect connections to your identity
information and personal interests.
2. The questions should not involve sensitive, pornographic, or unsafe
content.
# Output Examples
{demonstration}
Please ask your questions:
```

Ad-integrated response generation. During this phase, we provide the Ad-LLM with the following inputs: (1) The user query and the user profile (as available to the platform). (2) A complete list of all candidate ads in JSON format, including their IDs, titles, and bids. Critically, the prompt imposes strict constraints on the ad insertion format to maintain consistency and user experience. The Ad-LLM then generates ad-integrated output responses based on this information. The prompt template is structured as follows:

```
# Role
You are an AI assistant with advertising and marketing capabilities. You
need to provide useful answers to user's question while also inserting ads
when appropriate.
# Task
In the process of answering a question, when the user's question is
suitable for ad insertion, you can combine the provided JSON-format
candidate ad list and the user profile to naturally insert ad content at
an appropriate position, to maximize user experience and ad revenue.
# Candidate Ad List
{"Ad ID":"}{ad_ids[0]}","Ad Title":"{ad_titles[0]}","Ad Bid":"{ad_bids[0]}
{"Ad ID":"}{ad_ids[1]}","Ad Title":"{ad_titles[1]}","Ad Bid":"{ad_bids[1]}
......
# Requirements
1. You can choose not to insert an ad, or insert one or more ads. Note
that ads must be inserted in the format "@Ad Title@[Ad ID]". The ad title
and ad ID must strictly follow the candidate ad list. Severe penalties
will be applied for incorrect insertion format or incorrect ad-related
information.
2. When an inserted ad is clicked by the user, you will receive revenue
proportional to the ad's bid. However, the more ads you insert, the lower
the user experience. Therefore, you need to balance these two aspects to
achieve maximum social welfare. Obviously, if the user's question is not
suitable for ad insertion or there are no suitable candidate ads, do not
```

insert any ad is the best choice.
3. When inserting an ad, the ad title part cannot be modified. However, you can adjust the context around the ad insertion position in your answer to make the ad content more natural and increase the probability of user clicks.
4. Your role is an AI assistant, so the most important thing is to answer the user's question and ensure a good user experience. Then, at the right time and in the right way, insert a suitable ad. Note that even when inserting ads, you must maintain the quality of the overall answer. Do not include any thought process related to the ad insertion process in the final answer.
Ad Insertion Examples
Correct Examples
{correct demonstration}
Incorrect Examples
{wrong demonstration}
User Profile
{user_profile}
User Question
{user_query}
Your Answer:

User click feedback. In the final phase, we employ a User-LLM to simulate click-through behavior. The User-LLM is provided with the user profile, the user query, and the ad-integrated response generated by the Ad-LLM. Based on this context, the User-LLM outputs a structured list of click decisions in JSON format. This simulated feedback serves as the ground truth for evaluating the performance of our auction mechanism. The prompt template for this stage is:

Role
You are a real user, using an AI assistant to acquire information, seek product recommendations, or make shopping decisions. Below is your personal information:
Basic Information: {user_profile.info}
Personal Interests: {user_profile.interests}
Task
Read the answer from the AI assistant to your question. The answer might contain ads inserted in the format "@Ad Title@[Ad ID]". If there are ads, please provide your click behavior as a real user.
Requirements
1. Based primarily on your question, and also considering your identity and personal interests, realistically reflect on your click behavior after seeing ads inserted as "@Ad Title@[Ad ID]". Note that you can choose not to click any ad, or click one or more ads.
2. The decision to click can consider these aspects:
(1) Relevance: The direct or indirect relevance to your question, and how well it matches your identity and personal interests. (2) Nativeness: How well the ad integrates with the answer's context. You are more likely to click ads with better nativity. If it's a hard ad that disrupts the native feel of the answer or is an incoherent insertion, you will choose not to click.
(3) Competitiveness: Whether there are similar or competing ads in the answer. If so, your attention will be divided; please click only one ad or none at all.
3. Please simulate the user's experience after seeing the AI assistant's answer. If the answer (1) is flooded with too many ads or (2) contains a lot of irrelevant text related to ad insertion, this will severely damage the user experience. In this case, do not click any ads.
4. Please provide click feedback for ALL ads inserted in the format "@Ad Title@[Ad ID]" (Ad Title is text, Ad ID is like Ad-XXXX). Do not miss any

```
or duplicate.
5.  Output the result in a strict JSON list format, with no extra
information.  The fields are:
- ad_id:  Ad ID (string)
- clicked:  Whether clicked (boolean:  true / false)
- reason:  Reason (string, concise and clear).
If there are no ads in the answer, return an empty list "[]".
# Output Example
{demonstration}
# Your Question:
{user_query}
## AI Assistant's Answer
{ad_response}
Your click feedback is:
```

E.3 Configuration of LLM-Native Advertising Simulation Environment

We employ the Qwen3-4B-Instruct as the base model for both the Ad-LLM and the User-LLM in our proposed simulation environment.

Ad-LLM. To generate responses with diverse and creative advertisements, the Ad-LLM is configured with a temperature of 0.7, top-p of 0.8, and top-k of 20, following official recommendations [37]. The cutoff length of the input tokens is set to 4096, with a maximum output length of 1024 tokens. Inference is accelerated using the vLLM framework [16].

User-LLM. For simulating user feedback, stability is important. Therefore, we utilize a lower temperature of 0.3 for the User-LLM, while keeping top-p $= 0.8$ and top-k $= 20$. We set the cutoff length of the input tokens and the max output length to 4096 and 1024, respectively. Inference is also accelerated using the vLLM framework.

F Additional Experimental Details

F.1 Implementation Details of LLM-Auction

F.1.1 pCTR Model in LLM-Auction

To adapt to our proposed LLM-native advertising simulation environment based on user click feedback, we instantiate the performance metric prediction model as a predicted click-through rate (pCTR) model in LLM-Auction. Distinct from conventional online CTR models, our model is employed exclusively for offline LLM training and does not participate in online ad allocation. This design enables it to leverage the full response context y for more accurate performance estimation, which is unavailable in standard online prediction scenarios.

Model architecture. The pCTR model takes four primary inputs: the user query x, the ad-integrated response y, the user profile h (represented by user ID), and the candidate ad a_i (represented by ad ID). We adopt a simple yet effective architecture to focus on the properties of the auction mechanism.

Let $\phi_{\text{text}}(\cdot)$ denote a frozen pre-trained text encoder that maps both the user query and ad-integrated response into text embeddings. Let $\mathbf{e}_h \in \mathbb{R}^{d_h}$ and $\mathbf{e}_{a_i} \in \mathbb{R}^{d_a}$ denote learnable embeddings for user h and ad a_i, respectively. The pCTR for ad i is computed as:

$$\hat{f}_i^{\psi}(y, s) = \sigma\left(\text{MLP}\left([\phi_{\text{text}}(x); \phi_{\text{text}}(y); \mathbf{e}_h; \mathbf{e}_{a_i}]\right)\right), \tag{12}$$

where $[\cdot;\cdot]$ denotes concatenation, $\sigma(\cdot)$ is the sigmoid function, and MLP is a 3-layer multilayer perceptron. The final layer is a linear projection followed by sigmoid activation.

Hyperparameters and optimization details. We use Qwen3-Embedding-0.6B [39] as the text encoder for both the user query and ad-integrated response in our pCTR model. The dimensions of user and ad ID embeddings, d_h and d_a, are set to 64 and 32, respectively. The hidden dimensions of the MLP are [128, 64, 32]. The model is trained on $\mathcal{D}_{\text{on}}$ with batch size 256. The text encoder

$\phi_{\text{text}}(\cdot)$ remains frozen throughout training, with only the user and ad embeddings ($\mathbf{e}_h$, $\mathbf{e}_{a_i}$) and MLP parameters optimized. We use the Adam optimizer with a learning rate of 1×10^{-3}.

F.1.2 Detailed Reward Design

As described in Section 5.2.3, the user experience term $\Omega_{\text{resp}}(y)$ in our reward model penalizes excessive ad insertion and format errors, which is concretely implemented as follows:

$$\mathcal{R}(s, \boldsymbol{b}, y) = \sum_{i=1}^{n} b_i \cdot \hat{f}_i^{\psi}(y, s) + \lambda \left(-10 \cdot N_{\text{ad}}(y)^2 - 500 \cdot \mathbb{1}\{\text{format error}\}\right), \tag{13}$$

where $N_{\text{ad}}(y)$ denotes the number of ads inserted in the LLM response y, and $\mathbb{1}\{\text{format error}\}$ equals 1 when the ad insertion format does not conform to `@ad_title@[ad_id]`. This term can be extended with additional evaluators for factual consistency, ad-claim fidelity, toxicity, bias, or credibility. We use the lightweight version above to isolate the mechanism-design question in the present experiments.

F.1.3 Hyperparameters and Optimization Details of LLM-Auction

In the reward model update phase, we use the LLM from the previous epoch to independently sample ad-integrated responses based on all user queries and collect corresponding feedback from the User-LLM, forming the online dataset $\mathcal{D}_{\text{on}}$. The pCTR model is then updated via Eq. (4) as described in Appendix F.1.1.

In the LLM update phase, we re-sample candidate ad bids for each training sample to create the offline dataset $\mathcal{D}_{\text{off}}$, mitigating overfitting to static bidding patterns. For each sample in $\mathcal{D}_{\text{off}}$, we generate $M = 5$ responses, score them with the updated reward model, and fine-tune the LLM using the DPO loss from Eq. (6).

We perform full-parameter fine-tuning of the LLM using Direct Preference Optimization (DPO) in BFloat16 (BF16) precision. The DPO-specific hyperparameters are set as follows: the regularization coefficient β is 0.1, the learning rate is 1×10^{-6} with a cosine scheduler and a warmup ratio of 0.1, and the batch size is 64. And we set the threshold δ_{th} for constructing preference pairs to 10. For the reward model in Eq. (5), the balancing hyperparameter λ is set to 1.

F.2 Evaluation Protocol and Implementation Details of Baselines

Evaluation protocol. We evaluate all baselines on the same held-out queries, sampled bid profiles, candidate ads, and User-LLM feedback. RAG-Auction originally uses a second-price payment rule, and MOSAIC uses its own payment rule; in Table 1, to fairly compare the allocation efficiency with LLM-Auction, which employs first-price payment rule, we report first-price realized payment for both baselines. We acknowledge that changing the payment rule can affect advertiser bidding strategies, so we clarify this protocol from two perspectives. First, evaluating under fixed single-round bids is a standard protocol in the data-driven, learning-based mechanism design literature. Previous works [18, 20, 41, 44] all adopt the same setting. Notably, these mechanisms do not satisfy strict incentive compatibility according to the experimental results and are equally susceptible to strategic bidding. We therefore consider it reasonable and consistent to use fixed bids and to employ revenue under first-price payment as a fair proxy for comparing allocation efficiency across mechanisms, while also facilitating reproducibility. Second, as shown in the proof of Theorem 4.6, when the allocation rule is monotone and the payment rule is first-price, ROI-constrained value-maximizing advertisers maximize their obtained value by reporting truthfully. Since the allocation rules of all three mechanisms evaluated in our experiments—LLM-Auction, the first-price variant of RAG-Auction, and the first-price variant of MOSAIC—satisfy monotonicity, all three are IC under this model, and advertisers will bid truthfully. This further supports the validity of our evaluation protocol.

RAG-Auction [15]: RAG-Auction is a pre-generation mechanism that allocates ads by retrieval relevance before response generation. Following the LLM-Auction setting, we use the multi-allocation segment auction configuration and set $K = 3$, matching the average insertion volume of LLM-Auction on the test set. We follow the original pseudocode for allocation: the retrieval relevance score q_i is the cosine similarity between Qwen3-Embedding-0.6B embeddings of the ad profile a_i and query

x. The original RAG-Auction uses second-price payments; under our shared protocol, we keep its allocation rule and recompute realized revenue and reward with first-price click payments.

MOSAIC [31]: MOSAIC is a post-generation mechanism that samples multiple LLM responses at inference time and selects among them via auction. We reproduce its allocation procedure within our simulation environment and use the same reward model as LLM-Auction, defined in Eq. 13, to ensure a common evaluation signal. Because MOSAIC's original payment rule targets a different serving interface, we report first-price realized payments for the shared allocation-efficiency comparison.

LLM-Auction (Oracle): In our simulation environment, we can directly obtain feedback from User-LLM for different sampled responses to the same user query, enabling direct reward calculation for DPO-based post-training optimization of the LLM. This baseline effectively provides an approximation of the ideal performance metric prediction over the entire sample space, thus serving as an upper bound reference for LLM-Auction's performance.

F.3 Implementation Details of Ablation Studies

Variant with reduced sample size. During each epoch of the IRPO algorithm in LLM-Auction, we randomly sample a subset of the constructed online dataset $\mathcal{D}_{\text{on}}$ while keeping all other components unchanged. LLM-Auction (50% samples) in Table 2 represents a variant with the reward model using a pCTR model trained on only half of the samples.

Variant without response context. As described in Appendix F.1.1, our pCTR model uses the text embedding of the ad-integrated response generated by the LLM as one of the input features. We construct the LLM-Auction (w/o response) variant by removing this input feature while keeping the rest of the pCTR model architecture and hyperparameters unchanged. As shown in Table 2, ablating this input feature significantly impacts the training of LLM-Auction, demonstrating the importance of this design for the LLM to learn an effective allocation rule.

Unified GAUC computation. We use the LLM-Auction allocation rule after 6 epochs of training to sample responses on held-out test queries and collect click feedback from the User-LLM, constructing a unified evaluation dataset. This dataset approximates the data distribution under the optimal allocation, and we compute the GAUC of pCTR models on this evaluation dataset to assess the prediction accuracy of pCTR models corresponding to LLM-Auction under different epochs and settings under identical conditions.

F.4 Compute Resources

All experiments are conducted on a 32-GPU internal cluster. The GPU compute is approximately equivalent to NVIDIA A100 GPU, and each GPU provides 96 GB memory. Under this setup, one round of LLM alignment training takes approximately two hours.

G Additional Experimental Results

G.1 Simulator Robustness

Because the main evaluation uses LLM-simulated click feedback, we test whether the conclusions persist under a different User-LLM. This experiment follows the discussion in Section 5.1: the simulator provides a controlled sandbox, while LLM-Auction itself only requires click-style feedback and is not tied to the preference distribution of a particular simulator. Table 4 replaces Qwen3-4B with Qwen3-14B as the User-LLM, thereby changing the feedback distribution while preserving the same feedback data format.

The absolute revenue and reward values shift under Qwen3-14B, confirming that different User-LLMs induce different click distributions. Nevertheless, LLM-Auction (Oracle) remains far above the pre-trained LLM under both simulators: revenue increases from 70.4 to 223.3 with Qwen3-4B feedback and from 51.6 to 153.0 with Qwen3-14B feedback, while reward changes from negative to positive in both cases. This pattern indicates that LLM-Auction learns from the supplied feedback distribution rather than exploiting specific preference of one User-LLM. In deployment, the same

reward-alignment pipeline can replace simulated clicks with real user click logs, so the mechanism can adapt to real-world user click distributions as long as click feedback is available.

Table 4: Robustness to the User-LLM feedback distribution.

Method	User-LLM	Avg Rev	Avg Rew
Pre-trained LLM	Qwen3-4B	70.4	-218.0
LLM-Auction (Oracle)	Qwen3-4B	223.3	125.2
Pre-trained LLM	Qwen3-14B	51.6	-236.8
LLM-Auction (Oracle)	Qwen3-14B	153.0	90.3

G.2 Human Validation of the User-LLM Simulator

To check whether the User-LLM's click feedback reflects human judgments rather than model-specific artifacts, we conduct a small-scale human validation. We sample 200 ad-integrated responses generated by a diverse set of LLMs, containing 703 ad impressions in total. Three independent in-house professional annotators are shown exactly the same context as the User-LLM (user profile, query, and response) and make a click decision for each inserted ad. All three annotators agree on 95.3% of the impressions, and disagreements are resolved by majority vote.

For each of the $M = 703$ impressions, let $h_j \in \{0, 1\}$ denote the human majority label and $\ell_j \in \{0, 1\}$ the User-LLM's click decision. We report the agreement rate and Cohen's κ:

$$\text{Agreement} = \frac{1}{M}\sum_{j=1}^{M} \mathbb{1}[\ell_j = h_j], \qquad \kappa = \frac{p_o - p_e}{1 - p_e},$$

where p_o is the observed agreement and $p_e = \hat{P}(\ell = 1)\hat{P}(h = 1) + \hat{P}(\ell = 0)\hat{P}(h = 0)$ is the agreement expected by chance from the two marginal click rates. Cohen's κ thus discounts chance agreement ($\kappa = 0$: chance level; $\kappa = 1$: perfect agreement), so it is not inflated by the imbalance between clicked and non-clicked impressions. We run each User-LLM three times and report the mean $\pm$ std over the three runs in Table 5.

Table 5: Agreement between User-LLM click decisions and human majority labels on 703 ad impressions (mean $\pm$ std over 3 runs).

User-LLM	Agreement	Cohen's κ
Qwen3-4B (main experiments)	$75.3\% \pm 0.6\%$	0.49 ± 0.01
Qwen3-30B-A3B	$77.8\% \pm 1.2\%$	0.52 ± 0.02
Claude Sonnet 4.6	$79.9\% \pm 1.1\%$	0.54 ± 0.03

The User-LLM agrees with human judgments well above chance for all three models, and the agreement increases with stronger judge LLMs. This indicates that the simulator's feedback is not merely a model-specific artifact. Further aligning the User-LLM with human behavior through post-training or prompt optimization is a valuable direction for building more realistic user simulation environments.

G.3 Zero-Shot Generalization

We evaluate zero-shot transfer along two axes: replacing the candidate ad pool and replacing the user population. These two settings test different deployment risks. The ad-side experiment asks whether LLM-Auction can handle new advertiser domains without additional training, while the user-side experiment asks whether the learned allocation behavior transfers to heterogeneous user-profile distributions.

Generalization across ads. Table 6 replaces the entire candidate advertisement pool with 100 previously unseen ads. LLM-Auction (Oracle) is evaluated directly, without additional training, and still improves revenue from 66.2 to 169.7 and reward from -231.0 to 56.8. This result suggests that the mechanism does not simply memorize training advertisements. Instead, it transfers allocation

Table 6: Zero-shot generalization across the ad dimension. The candidate advertisement pool is replaced with 100 previously unseen ads, and the model is evaluated without additional training.

Method	Avg Rev	Avg Rew
Pre-trained LLM	66.2	-231.0
LLM-Auction (Oracle)	169.7	56.8

behavior based on generalizable signals such as query-ad semantic relevance, response context, and bids, which are precisely the signals available when new advertisers enter a platform.

Table 7: Zero-shot generalization across the user dimension. The evaluation uses 100 unseen user profiles and 500 newly generated queries, with no additional training.

Method	Avg Rev	Avg Rew
Pre-trained LLM	79.1	-219.3
LLM-Auction (Oracle)	242.6	136.8

Generalization across users. Table 7 evaluates 100 unseen user profiles that generate 500 new queries. LLM-Auction (Oracle) again retains large gains, increasing revenue from 79.1 to 242.6 and reward from -219.3 to 136.8. This user-side transfer indicates that the learned policy is not tied to the exact synthetic profiles or query set used during training. Instead, it generalizes to new user interests and information needs, supporting the applicability of LLM-Auction to heterogeneous user populations.

G.4 Base Model Scaling

LLM-Auction integrates auction decision-making and ad-integrated response generation into a single Ad-LLM. Thus, the base model controls both the language-generation quality and the capacity to internalize allocation signals. Table 8 compares Qwen3-4B with Qwen3-30B-A3B to examine how this choice affects performance.

Scaling the base model improves both the pre-trained LLM baseline and LLM-Auction after LoRA tuning. The pre-trained Qwen3-30B-A3B baseline already outperforms Qwen3-4B, suggesting that stronger language models provide better initial response and ad-selection behavior. More importantly, LLM-Auction (Oracle, LoRA) improves from 138.2 to 236.9 in revenue and from 25.4 to 103.6 in reward when the base model is scaled up. These results show that LLM-Auction inherits the scaling properties of the underlying generator: stronger models are better able to use feedback, context, and bid signals during alignment. We also observed training instability and collapse with smaller Qwen3-1.7B models, indicating that sufficient base-model capability is important for reliable auction-oriented alignment.

Table 8: Effect of base-model choice on allocation performance.

Method	Base model	Avg Rev	Avg Rew
Pre-trained LLM	Qwen3-4B	70.4	-218.0
Pre-trained LLM	Qwen3-30B-A3B	101.4	-135.7
LLM-Auction (Oracle, LoRA)	Qwen3-4B	138.2	25.4
LLM-Auction (Oracle, LoRA)	Qwen3-30B-A3B	236.9	103.6

H Qualitative Analysis

We present qualitative comparisons of ad-integrated responses to illustrate the performance improvements achieved by our proposed LLM-Auction mechanism. We compare responses generated by the pre-trained LLM (before LLM-Auction training) and the fine-tuned model of LLM-Auction, given identical user queries, candidate ads, and bids. Note that our experiments use a relatively small-scale LLM (Qwen3-4B), and the design of the reward model focuses on mechanism objectives rather than

the quality of the response. As a result, some responses generated may contain artifacts related to the reasoning process for the insertion of the ad. For presentation clarity, such extraneous content has been removed from the examples below.

Better query-ad relevance. Table 9 demonstrates improved semantic matching between user queries and selected advertisements. The pre-trained LLM inappropriately inserts a sofa advertisement (Ad-9420) in response to a query about preparing nourishing soups for elderly and children, exhibiting poor query-ad alignment. In contrast, LLM-Auction selects a vitamin B12 supplement advertisement (Ad-6542), which is directly related to the nutritional and wellness context of the query. Although this selection is constrained by our limited ad pool, it demonstrates better thematic relevance. This improvement indicates that LLM-Auction successfully learns to evaluate and prioritize ads based on semantic relevance with user queries, a critical factor for user engagement and click-through probability.

Improved contextual integration. Table 10 illustrates improved contextual integration of advertisements within response narratives. The pre-trained LLM exhibits a tendency to append ads as isolated afterthoughts disconnected from the main content, resulting in non-native placements that disrupt reading flow. By contrast, LLM-Auction seamlessly weaves advertisements into the natural discourse structure. For example, when discussing wetland hiking routes, it organically recommends anti-slip sports shoes (Ad-4064) as practical gear; when describing forest camping experiences, it naturally mentions clip-on Bluetooth earphones (Ad-7444) to enhance the outdoor experience. These contextually-grounded placements create a more cohesive narrative where ads serve as relevant recommendations rather than intrusive interruptions. This native integration style is more likely to be perceived as helpful rather than disruptive, thereby improving user experience and increasing click propensity.

Stronger bid-awareness in generation. Table 11 reveals LLM-Auction's acquisition of bid-awareness, a fundamental mechanism property. The pre-trained LLM demonstrates negligible sensitivity to bid signals: when Ad-2 (American Ginseng) submits a substantially higher bid (41) compared to Ad-1 (Progesterone Tea, bid 1), the model fails to adjust allocation accordingly. Moreover, when Ad-1's bid reaches the maximum value (100) while Ad-2 maintains bid 41, the pre-trained model simultaneously displays both advertisements with comparable prominence, effectively diluting Ad-1's competitive advantage despite its willingness to pay higher prices. In contrast, LLM-Auction exhibits clear bid-aware behavior: it increases the prominence and favorable positioning of higher-bidding advertisers in generated responses. This bid-awareness validates that our framework successfully trains the LLM to internalize auction dynamics, ensuring that advertisers receive allocation results corresponding to their submitted bids—a critical property for encouraging advertisers to participate and mechanism credibility in real-world deployment.

I Extended Limitations and Future Work

Benchmark scale and real-world generalization. The experiments rely on an LLM-based simulator because public, large-scale real-world logs for LLM-native advertising are not yet available. This simulator provides a controlled sandbox for comparing mechanisms, but it cannot fully capture the diversity of real user intent, long-term satisfaction, trust dynamics, advertiser strategy, or platform constraints. The robustness and zero-shot analyses in Appendix G reduce the risk that the results depend on a single simulator distribution, but production validation should ultimately use real impressions, clicks, conversions, and user-satisfaction signals. A central future direction is therefore to build standardized industrial-scale benchmarks for LLM-native advertising, with richer user populations, broader ad domains, and more faithful user-behavior simulation. Agent-based simulation techniques [19, 34] may provide a useful path toward more realistic mechanism evaluation before real deployment logs become available.

Reward design, hallucination, and bias. The current reward emphasizes allocation efficiency while using lightweight penalties for excessive ad insertion and formatting errors. This design helps isolate the mechanism objective and already reduces visible user-experience failures, but it remains too coarse for deployment settings where LLM-native ads may affect factuality, credibility, bias, and user trust. For example, an inserted ad may contain or encourage misleading claims, make the

surrounding answer less faithful to the user's intent, amplify demographic or topical bias, or blur the boundary between helpful recommendation and sponsored persuasion. LLM-Auction is compatible with richer user-experience signals through the same reward interface: future reward models should incorporate ad fidelity, response factuality, credibility, hallucination risk, bias, disclosure quality, and long-term trust metrics, together with calibrated pCTR models, safety filters, and human or automated audits. These additions would move the framework beyond controlling the number and format of inserted ads toward optimizing whether ad-integrated responses remain trustworthy and aligned with user welfare.

Optimization algorithms and online learning. Our current optimization procedure is a multi-stage offline prototype. IRPO periodically refreshes the pCTR model and the Ad-LLM, which allows the allocation policy to track the reward model induced by updated feedback, but it does not yet provide an end-to-end online learning system with strict latency, exploration, rollback, and safety guarantees. The alternating pipeline may also be sensitive to reward-model errors, delayed feedback, distribution shift, and short-term click over-optimization when deployed in a nonstationary market. Future work should treat LLM-Auction as a general mechanism-learning framework that can be paired with stronger optimization algorithms, including advanced reinforcement learning and preference-optimization methods [29, 30, 38, 43]. Such extensions could use real-time user feedback, conservative policy updates, and safety-constrained exploration to improve online adaptation while preserving response quality and mechanism stability.

Mechanism design beyond first-price payments. The first-price payment rule is computationally simple and has favorable incentive properties under the UM and ROI-constrained VM advertiser models analyzed in this paper. It is nevertheless only one point in a larger design space. Real advertising markets include budget pacing, repeated interactions, multi-objective advertisers, strategic exploration, and heterogeneous risk preferences. Designing computationally feasible incentive-compatible payment rules for these broader behaviors remains an important step toward optimal LLM-native advertising mechanisms. This direction also raises new equilibrium questions because the allocation object is a generated response rather than a fixed slot.

Scalability to large ad inventories. Our experiments place all candidate ads in the prompt, which does not scale to inventories with thousands of ads. LLM-Auction is compatible with the two-stage pipeline common in industrial advertising systems: a lightweight retrieval module first selects a small candidate set from the full inventory, and generative allocation is applied only to the retrieved candidates, so the LLM-side cost scales with the size of the retrieved set rather than the inventory. If retrieval is bid-independent, an advertiser cannot change its retrieval outcome by misreporting its bid, and the incentive properties in Section 4.2 carry over to the retrieved candidates. If retrieval depends on bids, a bid also affects whether the advertiser enters the candidate set, so incentive compatibility must account for both stages. Such two-stage auctions have been studied for conventional ad auctions [10, 35], and extending these designs to generative allocation is an important direction.

Measurement beyond clickable ads. Clickable ads make attribution precise in our simulation: the User-LLM can identify inserted ads and return structured click feedback. This abstraction is useful for controlled experimentation, but LLM-native advertising may influence users through narrative framing, product mentions, brand recall, and delayed conversions even when no explicit click occurs. Future work should develop content-based measurement methods that better match how users interact with generated answers, including metrics for helpfulness, disclosure, trust, long-term satisfaction, and advertiser value. Such measurement tools will be foundational for mature LLM advertising ecosystems.

Model dependence and deployment readiness. The model-scaling experiment shows that stronger base LLMs improve allocation performance, while smaller models can become unstable during training. This dependence is expected because LLM-Auction asks one generator to satisfy language quality, ad relevance, bid awareness, mechanism constraints, and formatting requirements simultaneously. Future work should study parameter-efficient training, distillation, stability regularization, monitoring, and rollback mechanisms so that LLM-native advertising mechanisms can be reliably deployed in high-concurrency industrial environments without requiring the largest available model.

Table 9: A comparison example of ad-integrated response generated by the pre-trained LLM and LLM-Auction. The format of the inserted clickable ads is `@ad_title@[ad_id]`, and we mark the corresponding text with blue underlines.

User Query:
How to make nutritious and easy-to-digest health soup with affordable ingredients for the elderly and children in autumn and winter?

Ad-Integrated Response (Pre-trained LLM):
During the autumn and winter seasons, the digestive systems of the elderly and children tend to be more sensitive. Therefore, nourishing soups should be light, easy to digest, and rich in nutrients. Here are some affordable yet nutritious and gentle soup recommendations:

1. Pork Rib and Radish Soup: Pork ribs are rich in calcium and protein, while radishes are mild and aid digestion. Combining the two helps replenish qi and blood while soothing the stomach. Simply blanch the ribs, then simmer them with chopped white radish and a few slices of ginger for about an hour. This soup is ideal for both the elderly and children.
2. Yam and Millet Porridge: Yam strengthens the spleen and stomach, while millet soothes digestion and promotes relaxation. This porridge is highly suitable for both children and seniors. Peel and cube the yam, then cook it with millet until soft. Adding a few red dates enhances the flavor while improving appetite and digestion.
3. Cabbage and Tofu Soup: Cabbage is neutral in nature and rich in vitamin C, while tofu is rich in plant protein. This combination is simple and nutritious. Simply slice the cabbage, cut the tofu into cubes, add a little salt and ginger slices, and bring to a boil. It's a good choice for children and the elderly to drink daily.

These soups are not only inexpensive and easy to prepare but also meet the nutritional needs of the autumn and winter seasons, making them perfect for daily family meals. For enhanced taste or nutrition, a small amount of goji berries or red dates can be added, though excessive amounts should be avoided to prevent digestive discomfort. @Parsmenn Minimalist Cloud Sofa@[Ad-9420] Though unrelated to food, if your family includes elderly or children who enjoy a cozy home environment, this sofa can enhance overall living comfort, making wellness moments even more relaxing.

Ad-Integrated Response (LLM-Auction):
During the autumn and winter seasons, the digestive systems of the elderly and children are relatively weaker. Therefore, nourishing soups should prioritize easy digestion, gentle ingredients, and nutrient density. Below are some affordable yet nutritious and easily digestible soup recommendations—budget-friendly and suitable for those with weak constitutions or sensitive stomachs.

1. Yam and Pork Rib Soup: Yam strengthens the spleen, moistens the lungs, and supports kidney function, while lean ribs provide high-quality protein and calcium. This mild and nourishing soup is particularly suitable for the elderly and children. Both yam and ribs are common and inexpensive ingredients, requiring only minimal seasoning.
2. Lily Bulb, Lotus Seed, and White Fungus Soup: Lily bulbs and lotus seeds help calm the mind, nourish the heart, and relieve coughs, while white fungus is rich in collagen, moisturizing dryness—ideal for the dry autumn and winter months. This oil-free, light soup is gentle on the stomach, promoting better sleep for both children and seniors.
3. Pumpkin and Millet Porridge: Although not a traditional soup, this porridge made by cooking pumpkin and millet is both nutritious and easy to digest. Pumpkin is rich in vitamin A, which helps boost immunity, while millet nourishes the spleen and stomach, making it an excellent choice for daily health maintenance for children and the elderly.

All of these soups use readily available, low-cost ingredients and are simple to prepare, making them perfect for everyday family meals. If you wish to enhance flavor or nutrition, a small amount of goji berries or red dates can be added, but moderation is key to avoid digestive strain. @Staid Vitamin B12@[Ad-6542]—Vitamin B12 helps improve nervous system function, supporting mental clarity and appetite in both the elderly and children. Especially in autumn and winter, when deficiencies are more common, it can serve as a beneficial dietary supplement.

Table 10: A comparison example of ad-integrated response generated by the pre-trained LLM and LLM-Auction. The format of the inserted clickable ads are `@ad_title@[ad_id]`, and we mark the corresponding text with blue underlines.

User Query:
What are some safe and nature-rich outdoor camping or hiking routes near Shanghai suitable for families on weekends?

Ad-Integrated Response (Pre-trained LLM):
There are many outdoor camping and hiking routes near Shanghai that are perfect for family weekends, offering both natural scenery and a safe, engaging experience for children. Here are some recommended options:

1. Jinshan Coastal Camping Park. Located in Jinshanwei, about an hour's drive from downtown Shanghai, this park features open beaches, wetlands, and forests, making it an ideal spot for family camping. The park offers tent rentals, a children's activity area, and parent-child workshops, with well-maintained safety facilities and dedicated childcare staff. Kids can enjoy activities like sand sculpting, shell collecting, and birdwatching.
2. Hiking Trail at Sheshan National Forest Park (Songjiang). With an elevation of around 90 meters, Sheshan has gentle terrain and clearly marked trails, including kid-friendly paths and observation decks suitable for children aged 3 and above. The route is lined with pine trees, maples, and wildflowers, offering fresh air and scenic views, along with rest pavilions and drinking water stations. Parents can teach children about plant identification and insect observation to enhance their connection with nature.
3. Water Town Hiking Routes Around Zhujiajiao Ancient Town (Qingpu). Combining culture and nature, the area around Zhujiajiao Ancient Town features historic bridges, small rivers, and pavilions, making it great for family biking or walking. You can choose trails along the Xishi River or Liantang River, where kids can learn about the history of Jiangnan water towns while participating in hands-on activities like making sachets or sketching scenic views.

All these routes prioritize safety and interactivity. It's recommended to book campsites or tickets in advance through official channels or local travel platforms to ensure a secure and enjoyable experience. Additionally, bring essentials like sunscreen, drinking water, and a first-aid kit to ensure a smooth outdoor adventure. @FILA Kids Basketball Shoes@[Ad-7633] are a practical choice for family outings, providing comfort and anti-slip designs suitable for wet forest trails or grassy areas.

Ad-Integrated Response (LLM-Auction):
Here are some family-friendly weekend camping and hiking routes around Shanghai that balance safety and nature immersion. These recommendations prioritize accessibility, scenic beauty, and security, making them ideal for family outings:

1. Zhujiajiao Wetland Hiking (Qingpu District). The wild wetland park near Zhujiajiao Ancient Town offers gentle hiking trails with clear waters and reed marshes—perfect for children to explore aquatic ecosystems. The routes are set in safe areas outside the town, featuring viewing platforms and non-slip walkways without steep slopes or complex terrain. Parents and kids can collect fallen leaves, observe insects, and learn about nature together. @Fila Kids' Sports Shoes@[Ad-4064]—For outdoor activities, slip-resistant, comfortable, and durable shoes are essential for children. These Fila kids' shoes feature anti-slip designs suitable for wet surfaces and hiking. Parents should prepare a pair for their little adventurers.
2. Sheshan National Forest Park Camping Area. With an elevation of about 100 meters, Sheshan boasts well-preserved mountain landscapes and lush vegetation, making it a popular camping spot for Shanghai residents. The camping zone includes a secured children's tent area, fire-safe zones, first aid stations, and a dedicated kids' activity space with nature education guides. Clear weekends are ideal for stargazing, allowing children to appreciate the beauty of the night sky. @Sanag Clip-on Bluetooth Earphones@[Ad-7444]—During outdoor camping, children may get distracted by wind or noise. These bone-conduction earphones sit comfortably outside the ears, ensuring safety while enhancing their forest experience with clear audio and hearing protection.
3. Jinshan Langxia Ecological Park Hiking Trail. Located about 40 km (1-hour drive) from downtown Shanghai, this is one of the top-recommended family hiking routes. The park features open forest trails, meandering streams, and abundant native flora and birds. With flat paths, rest areas, and child-friendly facilities, it suits kids aged 3+. Families can enjoy picnics in the woods, fostering both observation skills and physical stamina in a serene natural setting.

A warm and practical insulated water bottle is recommended during camping to avoid digestive discomfort from cold drinks.

Table 11: A comparison example of ad-integrated response generated by the pre-trained LLM and LLM-Auction. The format of the inserted clickable ads are `@ad_title@[ad_id]`, and we mark the corresponding text with blue underlines.

User Query:
The weather has turned cooler recently. Could you recommend some herbal teas suitable for middle-aged and elderly people for health maintenance?

Ad-Integrated Response (Pre-trained LLM), Ad-1 (Progesterone Tea) Bid 1, Ad-2 (American Ginseng) Bid 41:
As the weather turns cooler, middle-aged and elderly individuals need to pay extra attention to staying warm and nurturing their health. Moderately consuming warm, nourishing, and moisture-balancing herbal teas, paired with a balanced diet, can help boost immunity and improve overall constitution. Below are some recommended health teas:

1. Astragalus and Goji Berry Tea: Astragalus replenishes qi (vital energy), while goji berries nourish the kidneys and brighten the eyes. Ideal for those experiencing fatigue or qi deficiency. Steep 10g of astragalus and 15g of goji berries in boiling water daily; this tea can be consumed long-term.
2. Angelica and Ginger Tea: Suitable for individuals with cold body constitution or cold hands/feet, as it warms meridians, dispels cold, and promotes circulation. Simmer 6g of angelica root and 3 slices of ginger in water as a daily brew.

For a thoughtful wellness gift, consider the @Progesterone Tea@[Ad-2489]. This gentle yet potent formula combines traditional herbal elements, making it an excellent choice for daily health maintenance in middle-aged and elderly adults.

Ad-Integrated Response (Pre-trained LLM), Ad-1 (Progesterone Tea) Bid 100, Ad-2 (American Ginseng) Bid 41:
As the weather turns cooler, middle-aged and elderly individuals, with their sensitive constitutions, may benefit from drinking warm, nourishing herbal teas that support digestion and regulate qi and blood. Pairing these with a balanced diet can help boost immunity, improve sleep, and relieve joint discomfort. Here are some recommendations:

1. Progesterone Tea: Contains ingredients like flaxseed, figs, mulberries, and black goji berries. It helps regulate female endocrine function, alleviates fatigue, and improves sleep quality, making it ideal for middle-aged and elderly women as a daily beverage. @Progesterone Tea@[Ad-2489].
2. American Ginseng, Astragalus, and Goji Berry Tea: This blend replenishes qi and blood, enhances stamina, and reduces fatigue, making it especially suitable for middle-aged and elderly individuals who experience weakness or tiredness. For a milder flavor, steep with a few red dates or longan. @American Ginseng Tea@[Ad-3356].

Ad-Integrated Response (LLM-Auction), Ad-1 (Progesterone Tea) Bid 1, Ad-2 (American Ginseng) Bid 41:
As the weather turns cooler, middle-aged and elderly individuals, who generally have weaker constitutions, need to pay extra attention to staying warm and maintaining their health. Below are some recommended health-boosting teas and recipes suitable for this age group, designed to enhance immunity, regulate qi and blood, and alleviate fatigue.
Recommended Health Teas:
For example, you might try @American Ginseng Tea@[Ad-3356]. This tea is rich in natural ingredients such as astragalus, American ginseng, and goji berries, which help replenish qi and blood while strengthening the body. It is particularly suitable for consumption during autumn and winter.

Ad-Integrated Response (LLM-Auction), Ad-1 (Progesterone Tea) Bid 100, Ad-2 (American Ginseng) Bid 41:
As the weather turns cooler, it's the perfect season for health maintenance. Middle-aged and elderly individuals can benefit from warm, nourishing teas that moisturize dryness and regulate qi and blood. Paired with a balanced diet, these teas may help boost immunity, improve sleep, and ease joint discomfort.
@Progesterone Tea@[Ad-2489] is a wellness tea blend specially designed for middle-aged and senior women. With natural ingredients like flaxseeds, mulberries, and black goji berries, it supports endocrine regulation and enhances blood circulation—ideal for those experiencing cold hands/feet or frequent late nights.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Section 1 state the scope, contributions, and experimental claims of LLM-Auction. Sections 4 and 5, together with the limitations in Section 6 and Appendix I, support these claims and bound their scope.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 6 and Appendix I discuss the limitations in detail.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The assumptions and theoretical statements are given in Sections 3 and 4, including Assumptions 4.1 and 4.2. Full proofs of Theorem 4.3, Theorem 4.4, Theorem 4.5, and Theorem 4.6, as well as the closed form of the DSIC payment, are provided in Appendix C.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Sections 5.1–5.6 describe the simulation environment, dataset construction, baselines, metrics, and evaluation protocol. Appendices E and F provide prompt templates, model configurations, reward design, pCTR architecture, training procedure, and baseline implementation details.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility.

In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The data and code will be open-sourced in a GitHub repository after completing internal approval.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 5 and Appendix presents all the details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Table 1 reports the mean $\pm$ standard deviation of all main metrics over 10 disjoint repetition groups of sampled responses (3 independent full runs for MOSAIC), which captures response-sampling variability. Table 5 reports the mean $\pm$ standard deviation over 3 User-LLM runs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Refer to Appendix F.4.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work uses synthetic profiles, LLM-simulated click feedback, and controlled offline evaluation rather than collecting private human-subject data. The paper also discusses deployment risks and mitigation directions for trust, safety, bias, and online optimization in Section 6 and Appendix I.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The paper does not include a standalone broader-impact section, but it briefly discusses relevant positive and negative impacts in Section 1, Section 6, and Appendix I.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The submission does not release a high-risk pre-trained model, image generator, or scraped dataset. Appendix I discusses future safeguards such as safety filters, audits, monitoring, rollback, and richer reward signals for deployment.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All the references are made.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper constructs a simulation environment for evaluation but does not release a new dataset, code package, or model asset as part of this submission. The environment design, prompts, configurations, and examples are documented in Appendix E.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The experiments do not involve crowdsourcing. User queries and click feedback used for training and evaluation are generated by User-LLM simulators, as described in Section 5.1 and Appendix E. The only human involvement is the small-scale validation in Appendix G.2, where three in-house professional annotators made click decisions on 200 synthetic ad-integrated responses given the same context as the User-LLM.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing or human-subject research. The human validation in Appendix G.2 only asks annotators to label click decisions on synthetic ad-integrated responses and collects no personal data. All other evaluations use synthetic profiles and User-LLM-generated feedback rather than data collected from study participants.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: LLMs are central to both the proposed allocation rule and the simulation environment; specifically, we train an LLM to serve as the allocation rule for the auction mechanism, while simultaneously employing the LLM-as-a-judge approach to construct the simulation evaluation environment. For detailed implementation specifics, please refer to the sections on experimental details within the main text and the appendix.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.