

# 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, huqun.hu, ssp439458, dagui.cdg,  
zhuhan.zh, xiyu.xj, bozheng}@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 However, classic mechanisms are no longer applicable in this setting where the auction object is shifted from discrete ad slots to distributions over LLM outputs, and existing methods are impractical in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 advertisers’ 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 LLM-Auction, and prove that a simple first-price payment rule exhibits favorable incentive properties. 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 satisfying key mechanism properties.

## 1 INTRODUCTION

The rise of chatbots and AI search engines powered by large language models (LLMs) (Zhao et al., 2023; Brown et al., 2020) is fundamentally reshaping how users acquire information (Tsai, 2023). 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 slots attached to the LLM responses (Perplexity, 2024; Longo, 2024; Reid, 2023), which fail to exploit LLMs’ capabilities in reasoning and generation.

We focus on *LLM-native advertising*, where promotional content is naturally integrated within generated responses to better utilize LLM capabilities while balancing revenue and user experience. Unlike conventional auction settings with discrete ad slots, the auction object becomes the distribution over LLM outputs. In this new setting, 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 (Edelman et al., 2007) inapplicable, creating urgent demand for tailored auction mechanisms.

Existing auction mechanisms for LLM-native advertising mainly decouple auction from generation (Dütting et al., 2024; Soumalias et al., 2025; Hajiaghayi et al., 2024; Dubey et al., 2024). These approaches can be categorized into pre-generation and post-generation mechanisms, as shown in Figure 1. Pre-generation mechanisms allocate ads before generation, presetting insertion ad counts and failing to model externalities, leading to suboptimal performance. Post-generation mechanisms require multiple LLM forward passes to generate tokens or responses conditioned on ads and then run an auction on these intermediate results to determine the final output, incurring inference costs that scale with the number of advertisers or candidate responses.

---

\*Corresponding author.

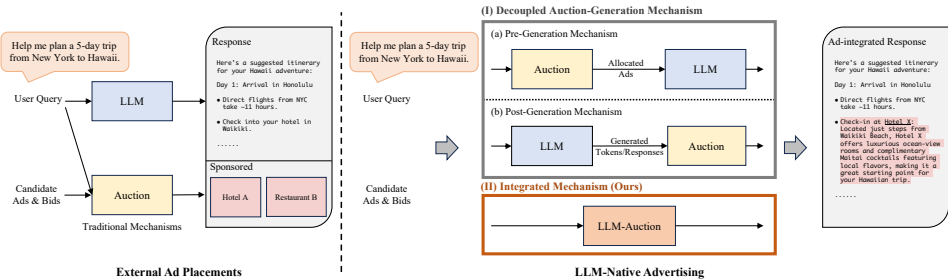


Figure 1: Comparison of ad formats and auction mechanisms in LLM-based AI applications.

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. We formulate the allocation as a preference alignment task, post-training the LLM to align outputs with a mechanism objective balancing revenue and user experience, conditioned on candidate ads and their bids. Specifically, we develop a reward model based on advertising performance metrics (e.g., click-through rate) to provide preference feedback. We then introduce *Iterative Reward-Preference Optimization* (IRPO), a novel algorithm that iteratively optimizes the reward model and the LLM to maximize the mechanism’s performance. After training, the LLM itself implements the allocation rule, implicitly modeling externalities without additional inference costs. Furthermore, we demonstrate that the allocation rule in LLM-Auction exhibits monotonicity and continuity. These properties ensure that a simple first-price payment rule achieves desirable incentive properties under two advertiser behavior models.

Additionally, inspired by the LLM-as-a-judge paradigm (Li et al., 2025), we design a simulation environment that spans the generation of user query, ad-integrated response, and user feedback, facilitating large-scale training and quantitative evaluation of auction mechanisms in the LLM-native advertising setting. Extensive experiments demonstrate that LLM-Auction significantly outperforms existing mechanisms in allocation efficiency while satisfying key mechanism properties.

Our main contributions are summarized as follows:

- We lay down general definitions and desired incentive properties for LLM-native advertising auction mechanisms, establishing foundations for future work.
- We propose LLM-Auction, the *first* learning-based generative auction mechanism for the LLM-native advertising setting that integrates allocation and generation through theoretically-grounded preference alignment.
- We identify the allocation monotonicity and continuity of LLM-Auction, and prove that a simple first-price payment rule exhibits favorable incentive properties.
- 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 (2007) established the game-theoretic foundation of position auctions in sponsored search. Edelman et al. (2007) formalized the GSP auction, analyzing its equilibrium properties and efficiency, with Caragiannis et al. (2015) proving GSP achieves constant-factor approximation to optimal social welfare even in worst-case equilibria. Recent work extends GSP to learning-based mechanisms (Zhang et al., 2021; Liu et al., 2021), and Zhu et al. (2025) propose a multi-slot generative auction mechanism that accounts for externalities.

**Mechanism design for LLMs.** An emerging research frontier that designs auction mechanisms for LLM-generated content with desirable incentive properties. Most existing work decouple auction and generation, falling into two categories: (i) *Pre-generation mechanisms* allocate ads via auction, then generate responses based on auction results. Dubey et al. (2024) introduce a prominence-based auction for LLM summarization, Hajiaghayi et al. (2024) design mechanisms using retrieval-augmented generation (RAG), and Balseiro et al. (2025) propose an extended position auction for

AI-generated content. However, these mechanisms require specifying ad numbers and positions upfront while failing to model allocation externalities (context and co-exposure effects), yielding suboptimal performance. (ii) *Post-generation mechanisms* generate multiple ad-conditioned candidates and then select or aggregate intermediate results via auction. Feizi et al. (2023) and Mordo et al. (2024) incorporate ad-rewriting stages into position auctions. Dütting et al. (2024) pioneer a token-level auction aggregating advertisers’ preferences during decoding, and Soumalias et al. (2025) propose a response-level auction selecting among multiple candidates. While more context-aware, these mechanisms incur additional LLM inference costs scaling with the number of advertisers or candidate responses, creating bottlenecks for high-concurrency industrial deployment. Besides, Bergemann et al. (2025) adapt VCG mechanism to LLM settings theoretically.

**LLM alignment methods.** LLM alignment aims to steer model outputs toward external objectives or human preferences. Reinforcement Learning from Human Feedback (RLHF) (Christiano et al., 2017; Ouyang et al., 2022) optimizes LLMs using reward models representing human preferences. Direct Preference Optimization (DPO) (Rafailov et al., 2023) approximates RLHF objective in closed form, enabling direct optimization from preference pairs. Several DPO variants including IPO (Gheshlaghi Azar et al., 2024), KTO (Ethayarajh et al., 2024), and SimPO (Meng et al., 2024) enhance DPO in different aspects. More recently, tuning-free methods such as LA (Gao et al., 2024) and Amulet (Zhang et al., 2025b) perform inference-time preference adaptation via latent space.

### 3 PRELIMINARIES

We formalize the LLM-native advertising auction as a single-turn interaction between a user and an LLM. Consider a set of  $n$  advertisers  $[n]$ , where each advertiser  $i$  is characterized by a private value  $v_i \in \mathbb{R}_{\geq 0}$  and a fixed ad  $a_i \in \mathcal{A}$ . Let  $\mathbf{v} = (v_1, \dots, v_n)$  denote the vector of advertisers’ private values. Let  $s = (x, h, \mathbf{a}) \in \mathcal{S}$  denote the *context*, which concatenates the user query  $x \in \mathcal{X}$ , the user profile  $h \in \mathcal{H}$ , and the candidate ad profile  $\mathbf{a} = (a_1, \dots, a_n) \in \mathcal{A}^n$ .

Upon observing the context  $s$  and the submitted bid profile  $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{R}_{\geq 0}^n$ , the LLM  $\pi_\theta$  maps the inputs to an allocation distribution  $\pi_\theta(\cdot \mid s, \mathbf{b}) \in \Delta(\mathcal{Y})$ , where  $\Delta(\mathcal{Y})$  denotes probability simplex over the response space  $\mathcal{Y}$ . Then, a response  $y \in \mathcal{Y}$  is sampled from this distribution, and it may contain the content related to *any subset* of candidate ads. An auction mechanism is thus defined as a tuple  $\mathcal{M} = (\pi_\theta, \mathbf{p})$ , consisting of:

- **Allocation Rule:** The LLM, parameterized by  $\theta$ , acts as the allocation rule  $\pi_\theta : \mathcal{S} \times \mathbb{R}_{\geq 0}^n \rightarrow \Delta(\mathcal{Y})$ . The mechanism then samples a realized response  $y \sim \pi_\theta(\cdot \mid s, \mathbf{b})$ .
- **Payment Rule:** The function  $\mathbf{p} : \Delta(\mathcal{Y}) \times \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}_{\geq 0}^n$  determines payment vector  $\mathbf{p}(\pi_\theta(\cdot \mid s, \mathbf{b}), \mathbf{b})$  based on the allocation distribution and bids.

To quantify the advertising effectiveness of  $a_i$  given response  $y$  and context  $s$ , we introduce the performance metric  $f_i(y, s)$ , which can be practical measures such as click-through rate (CTR) or conversion rate (CVR). With slight abuse of notation, we write  $f_i(\pi_\theta(\cdot \mid s, \mathbf{b}), s)$  to denote  $\mathbb{E}_{y \sim \pi_\theta(\cdot \mid s, \mathbf{b})}[f_i(y, s)]$ . In a well-behaved mechanism, an advertiser should not receive a less-preferred expected performance by submitting a higher bid. This leads to the following desideratum:

**Definition 3.1** (Allocation Monotonicity). An allocation rule  $\pi_\theta$  is *monotone* if, for any fixed context  $s$  and any competing bids  $\mathbf{b}_{-i}$ , for any pairs of bids  $b'_i \geq b_i$ , the resulting allocation satisfies:

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

Given the potential stochasticity of generative allocations, advertisers formulate their bidding strategies based on their specific optimization objectives and constraints. We consider two standard advertiser models: Utility Maximizers (UM) and Value Maximizers (VM).

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

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

**Definition 3.3** (Value Maximizer (VM)). Let  $\tau_i > 0$  denote advertiser  $i$ 's target ROI. Value maximizers seek to maximize their expected value subject to an ROI constraint:

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

For notation consistency, we also refer to the VM's objective as their "utility"  $u_i$ .

Based on these behavior models, we can define the ideal incentive properties of the mechanism:

**Definition 3.4** (Incentive Compatibility (IC)). A mechanism  $\mathcal{M} = (\pi_\theta, \mathbf{p})$  is *incentive compatible* if truthful bidding is a dominant strategy:  $\forall i, b_i \geq 0, u_i(v_i; s, v_i, \mathbf{b}_{-i}) \geq u_i(b_i; s, v_i, \mathbf{b}_{-i})$ .

**Definition 3.5** (Individual Rationality (IR)). A mechanism  $\mathcal{M} = (\pi_\theta, \mathbf{p})$  is *individually rational* if an advertiser's expected payment does not exceed their expected value:  $p_i(\pi_\theta(\cdot \mid s, \mathbf{b}), \mathbf{b}) \leq v_i \cdot f_i(\pi_\theta(\cdot \mid s, \mathbf{b}), s)$ .

Ultimately, the mechanism designer's goal is to optimize a policy that balances social welfare and user experience. The global objective function is formulated as follows:

$$J(\pi_\theta, \mathbf{v}, \mathbf{b}, s) = \sum_i v_i \cdot f_i(\pi_\theta(\cdot \mid s, \mathbf{b}), s) + \lambda \cdot \Omega_{\text{resp}}(\pi_\theta(\cdot \mid s, \mathbf{b}), s) - \beta \cdot \mathcal{D}_{\text{KL}}(\pi_\theta(\cdot \mid s, \mathbf{b}) \parallel \pi_{\theta_0}(\cdot \mid s, \mathbf{b})), \quad (3)$$

where  $\Omega_{\text{resp}}(\cdot)$  represent the expected penalty term for ad insertion constraint violations over distribution  $\pi_\theta(\cdot \mid s, \mathbf{b})$ , and the KL-divergence term  $\mathcal{D}_{\text{KL}}$  penalizes deviations from the pre-trained LLM  $\pi_{\theta_0}$ .

**Definition 3.6** (Optimal LLM-Native Advertising Mechanism). In the LLM-native advertising setting, the optimal mechanism  $\mathcal{M}^* = (\pi_{\theta^*}, \mathbf{p}^*)$  in LLM-native advertising consists of: (i) the optimal allocation rule  $\pi_{\theta^*}$  that maximizes the objective function in Eq. (3) for every realization  $(\mathbf{v}, \mathbf{b}, s)$  while satisfying allocation monotonicity and (ii)  $\mathbf{p}^*$  that ensures IC and IR.

Considering practical constraints in industrial deployment scenarios, it is common practice to permit a relaxation of the optimal mechanism defined in Definition 3.6. For instance, strict IC may be relaxed when computationally efficient payment rule is necessary.

## 4 LLM-AUCTION

We now formally introduce LLM-Auction, the first integrated generative auction mechanism for the LLM-native advertising setting. Full proofs of the theoretical results are deferred to Appendix C.

### 4.1 ALLOCATION RULE

As discussed in Section 3, the allocation rule of our auction setting is the LLM  $\pi_\theta$  itself, with the allocation result being the LLM's output distribution  $\pi_\theta(\cdot \mid s, \mathbf{b})$ . Thus, optimizing the allocation is equivalent to adjusting this distribution to maximize the objective defined in Eq. (3). We formulate this problem as a preference alignment task between LLMs and the mechanism's objective, and fine-tune the LLMs with candidate ads and bids as input using a principled reward model.

#### 4.1.1 MECHANISM REWARD MODEL

Aligning LLMs with mechanism objectives through post-training requires access to pairwise preference samples over different ad-integrated responses given identical inputs  $(s, \mathbf{b})$ . However, we cannot obtain such counterfactual feedback from users in the LLM-native advertising setting. To address this, we construct a reward model based on a performance metric prediction model. This reward model provides feedback signals for the post-training of the LLMs by estimating the advertising performance of an allocation result.

**Performance Metric Prediction Model in LLM-Auction.** We employ a parameterized prediction model  $\hat{f}^\psi$  to estimate performance metrics based on the response. Distinct from conventional prediction models,  $\hat{f}^\psi$  is used exclusively for offline LLM training, enabling it to use the response

context  $y$  for accurate estimation. Let  $\hat{f}_i^\psi(y, s)$  denote the predicted performance metric for ad  $i$  given inputs  $(y, s)$ . The model is trained on feedback logs using binary cross-entropy (BCE) loss:

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

where  $c_i \in \{0, 1\}$  is the feedback label for ad  $i$  (e.g., click or conversion) and  $\hat{f}_i^\psi(y, s)$  denotes the related predicted performance metric. We assume the prediction model is unbiased, which ensures that optimizing the induced reward aligns with the mechanism objective in expectation:

**Assumption 4.1** (Unbiased performance prediction). The prediction model is unbiased under allocation rule  $\pi_\theta$  if for any advertiser  $i$ , contexts  $s$ , and bid vectors  $\mathbf{b}$ , we have

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

**Reward model.** We fine-tune the LLM to maximize the objective in Eq. (3) using Direct Preference Optimization (DPO) (Rafailov et al., 2023) for its training stability and theoretical foundations. Drawing on DPO’s reward maximization derivation with KL constraints, and assuming truthful bidding ( $b_i = v_i$ ) and an unbiased prediction model (Assumption 4.1), we construct a response-level reward model  $\mathcal{R}$  by combining Eq. (3) with the predicted performance metrics:

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

where  $\Omega_{\text{resp}}(y)$  serves as a penalty term for ad insertion constraint violations associated with the generated response  $y$ . This reward model serves as the feedback signal for LLM alignment training, steering the output distribution toward the optimal allocation rule.

#### 4.1.2 ITERATIVE REWARD-PREFERENCE OPTIMIZATION ALGORITHM

Based on the reward model in Eq. (5), we can fine-tune the LLM using DPO to maximize the mechanism’s objective. However, this relies on the unbiased prediction model as stated in Assumption 4.1. In practice, distributional shift exists between the pre-trained LLM’s output and the optimal allocation rule. This shift biases the training distribution of the prediction model, leading to inaccurate reward estimation and suboptimal allocation. To mitigate this, we propose Iterative Reward-Preference Optimization (IRPO), which alternately refines the reward model and the LLM policy through interaction with real user feedback.

Let  $\pi_{\theta_t}$  and  $\mathcal{R}_{\psi_t}$  denote the LLM and the reward model with the performance metric prediction model at epoch  $t$ , respectively. IRPO consists of the following two phases:

**Reward model update.** We freeze the LLM  $\pi_{\theta_{t-1}}$  and deploy it online to collect user feedback, constructing an online dataset  $\mathcal{D}_{\text{on}}^t = \{s^k, y^k, \mathbf{c}^k\}_{k=1}^{N_{\text{on}}}$ , where  $N_{\text{on}}$  denotes the number of samples. Then we optimize the prediction model using Eq. (4) to obtain the updated reward model  $\mathcal{R}_{\psi_t}$ .

**LLM update.** We fix  $\mathcal{R}_{\psi_t}$  and construct an offline dataset  $\mathcal{D}_{\text{off}}^t$  from  $\mathcal{D}_{\text{on}}^t$  or synthetic data. For each sample  $(s, \mathbf{b}) \in \mathcal{D}_{\text{off}}^t$ , we sample  $M$  responses:  $y^m \sim \pi_{\theta_{t-1}}(\cdot | s, \mathbf{b})$ ,  $\forall m \in [M]$ , and compute the reward for each response:  $r^m = \mathcal{R}_{\psi_t}(s, \mathbf{b}, y^m)$ . Next, we select the highest-reward response  $y^w$  as the winner, and form the set of losing responses based on the threshold  $\delta_{\text{th}}$ :  $\mathcal{K} = \{y^l \mid r^w - r^l > \delta_{\text{th}}\}$ . The DPO loss is then computed as:

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

where  $\beta$  controls the KL penalty strength. We update the LLM parameters by minimizing Eq. (6) and deploy the updated LLM  $\pi_{\theta_t}$  for the next epoch.

Our proposed IRPO progressively reduces the distributional shift from the optimal allocation rule, improving the allocation performance of LLM-Auction. The pseudocode is provided in Algorithm 1.

#### 4.1.3 PROPERTIES OF THE ALLOCATION RULE

The LLM-Auction optimizes the allocation rule  $\pi_\theta$  to maximize the mechanism objective  $J(\pi_\theta, \mathbf{v}, \mathbf{b}, s)$  defined in Eq. (3). We now analyze the theoretical properties of the resulting optimal allocation rule  $\pi_{\theta^*}$ .

**Theorem 4.2** (Monotonicity of the Optimal Allocation). *The optimal mechanism  $\mathcal{M}^* = (\pi_{\theta^*}, \mathbf{p}^*)$  that maximizes the mechanism’s objective function satisfies allocation monotonicity. Formally, for any advertiser  $i$ , fixed context  $s$ , and competing bids  $\mathbf{b}_{-i}$ , if  $b'_i \geq b_i$ , then:*

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

Theorem 4.2 ensures that an advertiser will not receive a lower expected performance for increasing their bid. This property provides the necessary foundation for incentive compatibility.

Beyond monotonicity, we identify a distinct property in the LLM-native setting: *allocation continuity*. In traditional position auctions, advertisers compete for discrete slots, often resulting in binary (win/loss) outcomes and step-function expected values. In contrast, LLM-native allocations operate over a continuous distribution space of generated responses. The LLM’s output distribution  $\pi_{\theta}(\cdot | s, \mathbf{b})$  shifts smoothly, leading to continuous expected performance:

**Theorem 4.3** (Continuity of the Optimal Allocation). *In the LLM-native advertising setting, if the pre-trained policy  $\pi_{\theta_0}(\cdot | s, \mathbf{b})$  is continuous in  $\mathbf{b}$  for all  $s$ , then the expected performance of the optimal allocation rule  $\pi_{\theta^*}$  satisfies continuity:  $f_i(\pi_{\theta^*}(\cdot | s, \mathbf{b}), s)$  is continuous in  $\mathbf{b}$  for all  $i, s$ .*

We will discuss the implications of this continuity property in Theorem 4.3 for designing the payment rule in the next section and provide empirical validation of this property in Section 5.4.2.

## 4.2 FIRST-PRICE PAYMENT RULE

While incentive-compatible mechanisms like VCG (Vickrey, 1961; Clarke, 1971) are theoretically appealing, they are computationally infeasible at scale. Discrete auctions typically use generalized second-price (GSP) rules (Edelman et al., 2007), but in our continuous allocation settings, infinitesimal bid changes yield only infinitesimal allocation changes. Consequently, the “second price”—the minimum bid to preserve allocation—converges to the bid itself, collapsing the distinction between first-price and second-price payments. We adopt a first-price-style payment rule, where the expected payment of advertiser  $i$  equals its bid  $b_i$  multiplied by its expected performance:

$$p_i(\pi_{\theta}(\cdot | s, \mathbf{b}), \mathbf{b}) = b_i \cdot f_i(\pi_{\theta}(\cdot | s, \mathbf{b}), \mathbf{b}), s. \quad (7)$$

The simplicity of this first-price rule, coupled with the continuity of  $\pi_{\theta^*}$ , ensures both computational efficiency and favorable incentive properties. We now analyze the incentive properties of the first-price payment rule under the two standard advertiser behavior models defined in Section 3. We first consider the case where all advertisers are UMs. Based on Theorem 4.3 and Glicksberg’s theorem (Glicksberg, 1952), we have the following corollary:

**Corollary 4.4** (Equilibrium under UM). *In the LLM-native advertising setting with the optimal allocation rule, the first-price payment rule admits a mixed-strategy Nash equilibrium under the UM behavior model.*

Next, we consider value maximizers (VMs). We prove that truthful reporting of both private value and return-on-Investment (ROI) is an optimal strategy under the first-price payment rule:

**Theorem 4.5** (Truthfulness under VM with ROI constraint). *In the LLM-native advertising setting with the optimal allocation rule, the first-price payment rule is IC for VM with ROI constraint  $\tau_i$ , where the optimal bidding strategy is  $b_i = v_i/\tau_i$ .*

Furthermore, individual rationality (IR) is naturally satisfied for UM under the first-price payment rule. For VM, IR holds when  $\tau_i \geq 1$ , which is a standard assumption in practice.

In summary, the first-price payment rule is computationally efficient, satisfies IR, admits equilibria under UM behavior model, and achieves IC under ROI-constrained VM behavior model.

## 5 EXPERIMENTS

### 5.1 LLM-NATIVE ADVERTISING SIMULATION ENVIRONMENT

A key bottleneck in mechanism design for LLMs is the lack of standardized, quantitative evaluation at scale. Previous work predominantly relies on qualitative or small-scale experiments, which fail to capture the scale and complexity of ad auctions in real-world industrial scenarios.

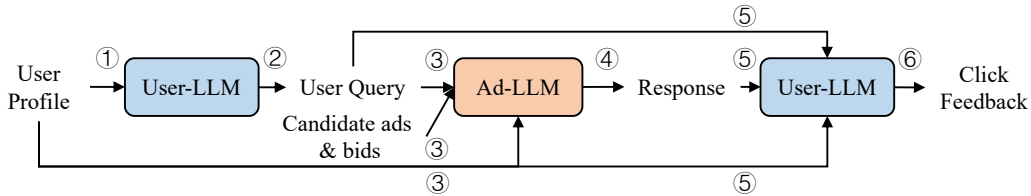


Figure 2: Illustration of our proposed simulation environment. The Ad-LLM generates ad-integrated responses based on the evaluated mechanisms. The numbers indicate the order of data flow.

To bridge this gap, we design a simulation environment inspired by the LLM-as-a-Judge paradigm (Li et al., 2025), simulating the entire ad serving and user feedback pipeline in LLM-based applications (e.g., chatbots and AI search). This enables large-scale dataset construction for both LLM alignment training and mechanism evaluation at an industrial scale. As shown in Figure 2, our pipeline consists of two modules: the Ad-LLM generates responses incorporating native advertisements based on the auction mechanism evaluated. Specifically, in our LLM-Auction framework, it is the trained LLM itself. The User-LLM in our framework simulates personalized queries and expresses user feedback through click events.

The overall process involves three stages: (i) **User query generation:** The User-LLM generates user queries based on user profiles that contain demographic information and personal interests. (ii) **Ad-Integrated response generation:** Conditioned on the user query, profile, and *all* candidate ads with bids, the Ad-LLM then generates the response that integrates the clickable ads based on the mechanism.<sup>1</sup> (iii) **User click feedback:** Based on the user profile, query, and the ad-integrated response, the User-LLM simulates click behavior and returns structured feedback.

Table 1: 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."}]

<sup>1</sup>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.

Table 2: Comparison of allocation efficiency in our proposed LLM-native advertising simulation environment.

Method	Revenue	Reward
Pre-trained LLM	70.37	-217.97
RAG-Auction	82.15	-20.87
MOSAIC	107.40	-15.27
LLM-Auction	170.86	81.86
LLM-Auction (Oracle)	223.30	125.24

Table 1 presents a concrete example of our proposed simulation environment. While the 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 inserting multiple ads can be found in Appendix F. Please refer to Appendix D for detailed implementation and prompt templates.

## 5.2 EXPERIMENTAL SETUP

### 5.2.1 DATASET

We construct 3k synthetic user profiles with demographic information and interests, and use the User-LLM to generate five different queries per profile, yielding 15k queries in total. We keep the query set fixed across experiments and split it into 12k training queries and 3k test queries. The candidate ad set comprises 100 real-world advertisements from a leading e-commerce platform *Taobao*, spanning 50 distinct categories to ensure diversity and reflect industrial scenarios.

### 5.2.2 BASELINES

We compare with the pre-trained LLM, two advanced LLM auction mechanisms, RAG-Auction (Hajiaghayi et al., 2024) and MOSAIC (Soumalias et al., 2025), and LLM-Auction (Oracle), a variant using direct User-LLM feedback. See Appendix E.2 for implementation details.

### 5.2.3 IMPLEMENTATION DETAILS OF LLM-AUCTION

**Reward design.** Following Eq. (5), we instantiate the reward model. Specifically, the performance metric prediction model  $\hat{f}_i^{\psi}$  is implemented as a predicted click-through rate (pCTR) model, and  $\Omega_{\text{resp}}(y)$  penalizes excessive ad insertions and formatting errors. Refer to Appendix E for the details of the pCTR model and the reward model.

**Training protocol.** The training process follows the IRPO algorithm with Qwen3-4B-Instruct (Yang et al., 2025) as our base model. In the reward model update stage, we use the Ad-LLM from the previous epoch to sample 10 responses per query, and collect the feedback from the User-LLM to form the dataset  $\mathcal{D}_{\text{on}}$  for the pCTR model updates. The LLM update stage samples 5 responses per training query, scores them via the reward model with updated pCTR model, and optimizes the Ad-LLM. We provide the complete candidate ad set in the Ad-LLM prompt, and bids for each training query during each epoch are sampled uniformly over positive integers between 1 and 100 per query to prevent overfitting to static bid patterns. Additional implementation details and hyperparameters are provided in Appendix E.1.3.

### 5.2.4 EVALUATION METRICS

We evaluate mechanisms on a held-out test set  $\mathcal{D}_{\text{test}}$  with unseen queries and independently sampled bid profiles. Consistent with our simulation environment’s setup where click events represent user feedback, the realized payment triggered by click events. Let  $p_i^k$ ,  $c_i^k$ , and  $b_i^k$  denote the payment, click event, and bid of advertiser  $i$  for the  $k$ -th sample, respectively. We have  $p_i^k = b_i^k \cdot \mathbb{1}(c_i^k = 1)$ . Based on the User-LLM’s click feedback, we compute the following two 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:  $\frac{1}{|\mathcal{D}_{\text{test}}|} \sum_{k \in \mathcal{D}_{\text{test}}} \mathcal{R}(s^k, \mathbf{b}^k, y^k)$ .

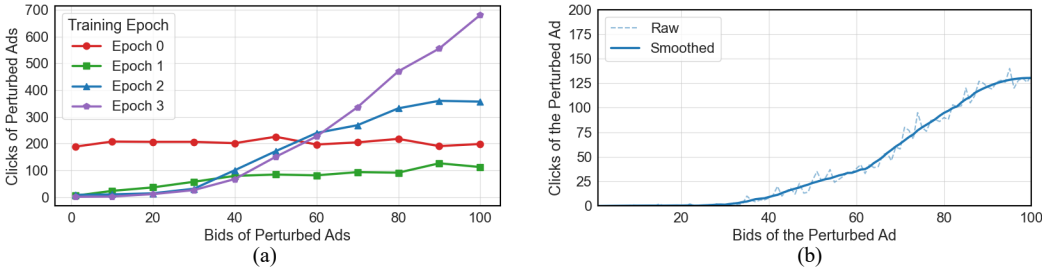


Figure 3: Experiment results of mechanism properties. **(a) Allocation monotonicity:** The allocation monotonicity of LLM-Auction gradually increases during training. **(b) Allocation continuity:** With a fixed query and perturbed ad, the allocation results change continuously as the bid increases.

### 5.3 COMPARISON OF ALLOCATION EFFICIENCY

Table 2 compares LLM-Auction with baselines, demonstrating significant and consistent improvements in both revenue and reward (which also accounts for user experience). Compared to the pre-generation baseline RAG-Auction, LLM-Auction achieves substantial gains in revenue and reward per query, highlighting the advantage of leveraging post-generation information for allocation. Notably, LLM-Auction achieves substantial improvements over MOSAIC, the current state-of-the-art post-generation mechanism, increasing revenue by 59.1% and reward from -15.27 to 81.86, with no additional inference costs. This indicates that LLM-Auction learns superior allocation strategies compared to inference-time sampling and ensemble methods. Furthermore, the oracle variant of LLM-Auction trained with direct feedback from the User-LLM achieves higher performance, showing substantial room for improvement as the accuracy of the performance metric prediction model increases. This validates the effectiveness of our learning-based auction framework and suggests promising directions for future work.

### 5.4 VERIFICATION OF MECHANISM PROPERTIES

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

#### 5.4.1 ALLOCATION MONOTONICITY

We perform bid perturbation experiments on  $\mathcal{D}_{\text{test}}$ . For each sample, we randomly select a candidate ad and vary its bid over  $\{1, 10, 20, \dots, 100\}$  (with other bids fixed). We measure allocation results of LLM-Auction from different epochs via User-LLM clicks on the perturbed ad. Figure 3(a) shows that the pre-trained LLM exhibits no correlation between bids and clicks. As training progresses, this correlation gradually strengthens. After 3 epochs, the mechanism achieves allocation monotonicity, where higher bids consistently yield higher click rates. This indicates that even without achieving the optimal allocation, the monotonicity of LLM-Auction improves progressively throughout training, empirically supporting and extending the theoretical result established in Theorem 4.2 under the optimality assumption.

#### 5.4.2 ALLOCATION CONTINUITY

Allocation continuity requires that allocation results vary smoothly (non-stepwise) with bids under fixed conditions. We fix an unseen test query and a candidate ad, then sweep the ad’s bid from 1 to 100. For each bid value, we sample 200 responses and count User-LLM clicks on the fixed perturbed ad. As shown in Figure 3(b), the allocation changes smoothly with the bid rather than in the step-function manner typical of discrete slot auctions. This empirically validates Theorem 4.3 and supports the first-price payment rule in our setting.

## 5.5 ABLATION STUDY

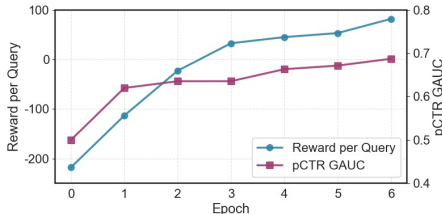
We conduct ablation studies to analyze the key components of LLM-Auction. We measure the accuracy of the pCTR model using Group AUC (GAUC), which computes the weighted average of per-user AUC scores. Additional experimental details are provided in Appendix E.3.

Table 3: 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

**Importance of response context.** As shown in Table 3, removing the response context from the performance prediction model dramatically degrades LLM-Auction, with a loss of revenue of 55%. This shows that leveraging full response context in LLM-Auction is critical for accurate performance metric estimation and effective allocation optimization.

**Correlation between prediction accuracy and mechanism performance.** As shown in Table 3, the performance of LLM-Auction variants aligns with GAUC variations. The oracle variant in Table 2, which approximates an ideal pCTR model, achieves the best performance. These results empirically validate the positive relationship between the accuracy of the pCTR model and the performance of LLM-Auction.



**Effectiveness of IRPO algorithm.** As shown in Figure 4, during the iterative training of LLM-Auction with IRPO, both the mechanism performance and the GAUC of the pCTR model improve steadily across 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 mechanism performance.

## 5.6 QUALITATIVE RESULTS

We also qualitatively compare the ad-integrated responses generated before and after LLM-Auction training. Please refer to Appendix F for details.

## 6 LIMITATIONS AND FUTURE WORK

This work is an initial step toward learning-based LLM-native advertising mechanisms, revealing many exciting future directions. (i) Our simulation environment is limited in scale and diversity; developing standardized industrial-scale benchmarks is crucial to advance the field. (ii) Current implementation employs offline multi-stage optimization with simple rewards; integrating advanced RL algorithms (Schulman et al., 2017; Shao et al., 2024; Yu et al., 2025; Zheng et al., 2025) and refined reward functions could enable single-stage optimization methods that learn directly from real-time user feedback. (iii) While our first-price payment rule achieves favorable incentive properties, designing computationally feasible IC payment rules for broader advertiser behaviors remains critical to realizing optimal mechanisms. (iv) While clickable ads simplify attribution, establishing content-based measurement methodologies better aligned with LLM applications will be foundational for emerging LLM ecosystems.

## REFERENCES

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.

- 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, pp. 507, 2025.
- 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, pp. 1877–1901, 2020.
- 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.
- 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.
- Edward H. Clarke. Multipart pricing of public goods. *Public Choice*, 11:17–33, 1971.
- 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, pp. 713–722, 2024.
- 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, pp. 144–155, 2024.
- 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.
- Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. Kto: Model alignment as prospect theoretic optimization. *arXiv preprint arXiv:2402.01306*, 2024.
- Soheil Feizi, MohammadTaghi Hajiaghayi, Keivan Rezaei, and Suho Shin. Online advertisements with llms: Opportunities and challenges. *arXiv preprint arXiv:2311.07601*, 2023.
- 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, pp. 14702–14722, 2024.
- 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, pp. 4447–4455, 2024.
- 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.
- 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, pp. 18445–18480, 2024.
- 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.

- 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.
- 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*, pp. 3354–3364, 2021.
- 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.
- 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, pp. 124198–124235, 2024.
- 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*, pp. 167–173, 2024.
- 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, pp. 27730–27744, 2022.
- 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.
- 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, pp. 53728–53741, 2023.
- Elizabeth Reid. Supercharging search with generative ai, 2023. URL <https://blog.google/products/search/generative-ai-search/>. Last accessed 10 November 2025.
- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
- 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.
- 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.
- 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.
- Hal R. Varian. Position auctions. *International Journal of Industrial Organization*, 25(6):1163–1178, 2007.
- William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.
- 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*, pp. 25–32, 2017. ISBN 9781450349130. doi: 10.1145/3038912.3052687.

- 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.
- Qiyang Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, 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, Jiase 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.
- 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*, 2025a.
- 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)*, 2025b.
- 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*, pp. 993–1001, 2021.
- 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.
- 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.
- 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*, pp. 2171–2181, 2025.

## A OPTIMIZATION AND TRAINING PROCEDURES

---

### Algorithm 1 Iterative Reward-Preference Optimization (IRPO)

---

**Input:** Initial LLM parameters  $\theta_0$ , initial prediction model parameters  $\psi_0$ , number of sampled responses  $M$ , DPO pair threshold  $\delta_{th}$ , learning rates  $\gamma, \eta$ .

**Output:** Optimized LLM parameters  $\theta_T$  and prediction model parameters  $\psi_T$ .

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

## B DEPLOYMENT FRAMEWORK

Figure 5 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 via the IRPO algorithm and deploys the updated LLM online.

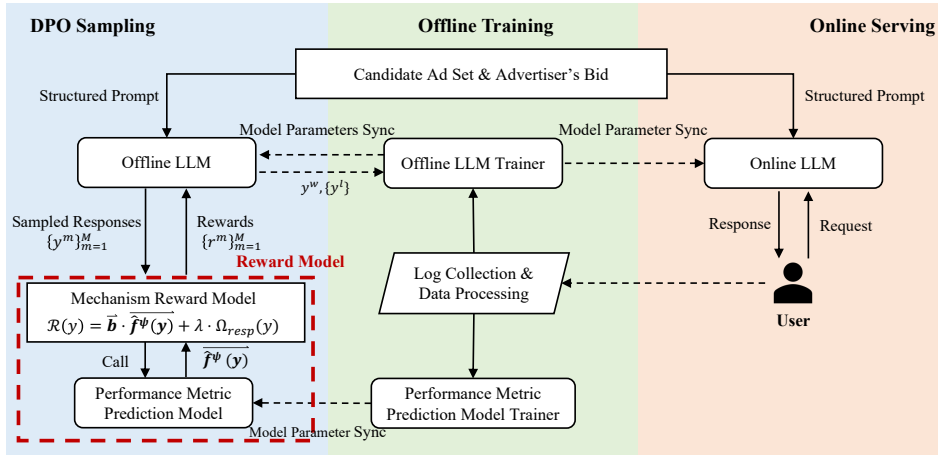


Figure 5: Deployment framework of LLM-Auction.

## C PROOFS OF THEORETICAL RESULTS

### C.1 PROOFS OF THEOREM 4.2

**Theorem C.1** (Restatement of Theorem 4.2). *The optimal mechanism  $\mathcal{M}^* = (\pi_{\theta^*}, \mathbf{p}^*)$  that maximizes the mechanism’s objective function satisfies allocation monotonicity. Formally, for any advertiser  $i$ , fixed context  $s$ , and competing bids  $\mathbf{b}_{-i}$ , if  $b'_i \geq b_i$ , then:*

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

*Proof.* To streamline notation, we henceforth denote the conditional policy  $\pi_{\theta}(\cdot | s, b_i, \mathbf{b}_{-i})$  simply as  $\pi_{\theta}(b_i)$ , treating  $(s, \mathbf{b}_{-i})$  as fixed context. Under an optimal mechanism that satisfies *Incentive Compatibility* (IC), truthful bidding is incentivized, implying  $\mathbf{v} = \mathbf{b}$ . Consequently, we simplify the objective function notation from  $J(\pi_{\theta}, \mathbf{v}, \mathbf{b}, s)$  to  $J(\pi_{\theta^*}, \mathbf{b})$ , where  $\theta^*$  denotes the optimal parameters. Accordingly, the user experience term, denoted by  $e(\pi_{\theta^*}(b_i))$ , is defined as:  $e(\pi_{\theta}(b_i)) = \lambda \mathbb{E}_{y \sim \pi_{\theta}(\cdot | s, \mathbf{b})} [\Omega_{\text{resp}}(y)] - \beta \mathcal{D}_{\text{KL}}(\pi_{\theta}(\cdot | s, \mathbf{b}) \| \pi_{\theta_0}(\cdot | s, \mathbf{b}))$ . First, according to Definition 3.6, the optimality of  $\pi_{\theta^*}$  implies that for any bid  $b_i$  of advertiser  $i$ , we have:

$$\pi_{\theta^*}(b_i) = \arg \max_{\pi_{\theta}} J(\pi_{\theta}(b_i), (b_i, \mathbf{b}_{-i})). \quad (8)$$

Consequently, it follows from Eq. (8) that for any two bids  $b_i$  and  $b'_i$ , the following inequalities hold:

$$\begin{aligned} J(\pi_{\theta^*}(b_i), (b_i, \mathbf{b}_{-i})) &\geq J(\pi_{\theta^*}(b'_i), (b_i, \mathbf{b}_{-i})), \\ J(\pi_{\theta^*}(b'_i), (b'_i, \mathbf{b}_{-i})) &\geq J(\pi_{\theta^*}(b_i), (b'_i, \mathbf{b}_{-i})). \end{aligned}$$

Then, by invoking the definition of the objective function  $J$  from Eq. (3), we can expand the first inequality as:

$$\begin{aligned} &b_i f_i(\pi_{\theta^*}(b_i), s) + \sum_{j \neq i}^n b_j f_j(\pi_{\theta^*}(b_i), s) + e(\pi_{\theta^*}(b_i)) \\ &\geq b_i f_i(\pi_{\theta^*}(b'_i), s) + \sum_{j \neq i}^n b_j f_j(\pi_{\theta^*}(b'_i), s) + e(\pi_{\theta^*}(b'_i)). \end{aligned} \quad (9)$$

$$\begin{aligned} &b'_i \cdot f_i(\pi_{\theta^*}(b'_i), s) + \sum_{j \neq i}^n b_j \cdot f_j(\pi_{\theta^*}(b'_i), s) + e(\pi_{\theta^*}(b'_i)) \\ &\geq b'_i \cdot f_i(\pi_{\theta^*}(b_i), s) + \sum_{j \neq i}^n b_j \cdot f_j(\pi_{\theta^*}(b_i), s) + e(\pi_{\theta^*}(b_i)). \end{aligned} \quad (10)$$

Combining inequalities (9) and (10), we have

$$(b_i - b'_i)(f_i(\pi_{\theta^*}(b_i), s) - f_i(\pi_{\theta^*}(b'_i), s)) \geq 0,$$

which completes the proof.  $\square$

### C.2 PROOFS OF THEOREM 4.3

**Theorem C.2** (Restatement of Theorem 4.3). *In the LLM-native advertising setting, if for any context  $s$ ,  $\pi_{\theta_0}(\cdot | s, \mathbf{b})$  is continuous in the bid profile  $\mathbf{b}$ , then for the optimal allocation rule  $\pi_{\theta^*}$ , the expected performance  $f_i(\pi_{\theta^*}(\cdot | s, \mathbf{b}), s)$  is a continuous function of  $\mathbf{b}$  for any advertiser  $i$  and context  $s$ .*

*Proof.* Given bid profile  $\mathbf{b}$ , we rewrite our objective function Eq. (3):

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

By Assumption 4.1, the performance prediction model is unbiased. We can therefore substitute the reward model  $R(s, \mathbf{b}, y)$  defined in Eq. (5) into the objective:

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

According to prior work Rafailov et al. (2023), based on the reward model defined in Eq. (5), the optimal allocation rule  $\pi_{\theta^*}$  obtained by the DPO algorithm is as follows:

$$\pi_{\theta^*}(y | s, \mathbf{b}) = \frac{1}{Z(s, \mathbf{b})} \pi_{\theta_0}(y | s, \mathbf{b}) \exp\left(\frac{1}{\beta} R(s, \mathbf{b}, y)\right), \quad (11)$$

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

$$\begin{aligned} f_i(\pi_{\theta^*}(\cdot | s, \mathbf{b}), s) &= \sum_y \pi_{\theta^*}(y | s, \mathbf{b}) f_i(y, s) \\ &= \frac{\sum_y \pi_{\theta_0}(y | s, \mathbf{b}) \exp\left(\frac{1}{\beta} R(s, \mathbf{b}, y)\right) f_i(y, s)}{\sum_y \pi_{\theta_0}(y | s, \mathbf{b}) \exp\left(\frac{1}{\beta} R(s, \mathbf{b}, y)\right)}. \end{aligned} \quad (12)$$

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

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

is linear—and hence continuous—in the bid profile  $\mathbf{b}$ . Under the assumption that  $\pi_{\theta_0}(y | s, \mathbf{b})$  is continuous in the bid profile  $\mathbf{b} \in [0, \bar{b}]^n$ , for each fixed  $y$ , the map

$$\mathbf{b} \mapsto \pi_{\theta_0}(y | s, \mathbf{b}) \cdot \exp\left(\frac{1}{\beta} R(s, \mathbf{b}, y)\right)$$

is continuous on  $[0, \bar{b}]^n$ , as it is the product of two continuous functions. Since  $f_i(y, s)$  does not depend on  $\mathbf{b}$ , we conclude:

- The numerator of Eq. (12) is a finite sum of continuous functions of  $\mathbf{b}$ , hence continuous in  $\mathbf{b}$ .
- The denominator of Eq. (12) is also a finite sum of continuous functions of  $\mathbf{b}$ , and is strictly positive for all  $\mathbf{b} \in [0, \bar{b}]^n$ , because  $\exp(\cdot) > 0$  and  $\pi_{\theta_0}(\cdot)$  is a probability distribution (thus nonnegative and not identically zero).

Therefore, the ratio in Eq. (12) defines a continuous function of  $\mathbf{b}$  on the compact set  $[0, \bar{b}]^n$ .  $\square$

### C.3 PROOFS OF COROLLARY 4.4

**Corollary C.3** (Restatement of Corollary 4.4). *In the LLM-native advertising setting with the optimal allocation rule, the first-price payment rule admits a mixed-strategy Nash equilibrium under the UM behavior model.*

*Proof.* We now establish the existence of a mixed-strategy Nash equilibrium in the bidding game among advertisers.

For each advertiser  $i$ , let the pure strategy space be the bid interval

$$B_i := [0, \bar{b}] \subset \mathbb{R}_{\geq 0},$$

for some finite upper bound  $\bar{b} > 0$ . Clearly, each  $B_i$  is a nonempty, compact, and convex subset of  $\mathbb{R}$ , and hence a compact metric space under the standard Euclidean metric.

According to Definition 3.2, the utility of advertiser  $i$  under the optimal allocation rule  $\pi_{\theta^*}$  is given by

$$u_i(\pi_{\theta^*}(\cdot | s, \mathbf{b}), v_i) = (v_i - b_i) \cdot f_i(\pi_{\theta^*}(\cdot | s, \mathbf{b}), s).$$

By the continuity of the optimal allocation rule with respect to the bid profile  $\mathbf{b}$  in Theorem 4.3, we can obtain that the performance for the ad of the advertiser  $i$ ,  $f_i(\pi_{\theta^*}(\cdot \mid s, \mathbf{b}), s)$  is continuous on the joint pure strategy space  $B = \prod_{j=1}^n B_j$ . Since  $(v_i - b_i)$  is also continuous in  $b_i$ , it follows that  $u_i : B \rightarrow \mathbb{R}$  is continuous on  $B$ .

Thus, the game satisfies the conditions of **Glicksberg’s theorem** Glicksberg (1952): (i) each pure strategy space  $B_i$  is a nonempty compact metric space, and (ii) each utility function  $u_i$  is continuous on the product space  $B$ .

Consequently, there exists at least one mixed-strategy Nash equilibrium and completes the proof.  $\square$

#### C.4 PROOFS OF THEOREM 4.5

**Theorem C.4** (Restatement of Theorem 4.5). *In the LLM-native advertising setting with the optimal allocation rule, the first-price payment rule is IC for VM with ROI constraint  $\tau_i$ , where the optimal bidding strategy is  $b_i = v_i/\tau_i$ .*

*Proof.* We first establish that when all advertisers have ROI constraint parameter  $\tau_i = 1$ , the mechanism is incentive compatible for value maximizers. We then extend this to general  $\tau_i \geq 1$  via a reduction. For simplicity, we hereafter denote  $\pi_{\theta}(\cdot \mid s, (b_i, \mathbf{b}_{-i}))$  as  $\pi_{\theta}(b_i)$ , assuming  $s$  and  $\mathbf{b}_{-i}$  are fixed.

The mechanism selects the optimal allocation rule  $\pi_{\theta^*}(b_i)$  that maximizes the objective function  $J(\pi_{\theta}, \mathbf{b})$  defined in Eq. (3), and charges each advertiser under the first-price payment rule described in Section 4.2:

$$p_i(b_i) = b_i \cdot f_i(\pi_{\theta^*}(b_i), s).$$

Recall from Definition 3.3: the utility of a value maximizer  $i$  is  $u_i(b_i) = v_i \cdot f_i(\pi_{\theta^*}(b_i), s)$ , subject to the ROI constraint  $\tau_i \cdot p_i(b_i) \leq u_i(b_i)$ .

**Case 1:**  $\tau_i = 1$  Consider an advertiser with true value  $v_i$  and  $\tau_i = 1$ . We show that truthful bidding,  $b_i = v_i$ , is optimal by analyzing potential deviations:

- **Underbidding** ( $b_i < v_i$ ): By Theorem 4.2, the optimal allocation rule  $\pi_{\theta^*}$  is monotone, so  $f_i(\pi_{\theta^*}(b_i), s)$  is weakly increasing in  $b_i$ . Since the ROI constraint  $b_i f_i \leq v_i f_i$  is always satisfied for any  $b_i \leq v_i$ , the advertiser’s utility  $v_i f_i$  is maximized by bidding the highest feasible value within this range, which is  $b_i = v_i$ . Any bid  $b_i < v_i$  results in a weakly lower allocation and, consequently, weakly lower utility.
- **Overbidding** ( $b_i > v_i$ ): The advertiser’s expected value remains  $v_i \cdot f_i(\pi_{\theta^*}(b_i), s)$ , but her payment becomes  $b_i \cdot f_i(\pi_{\theta^*}(b_i), s)$ . With  $\tau_i = 1$ , the ROI constraint requires

$$b_i \cdot f_i(\pi_{\theta^*}(b_i), s) \leq v_i \cdot f_i(\pi_{\theta^*}(b_i), s),$$

which is violated whenever  $f_i(\pi_{\theta^*}(b_i), s) > 0$ . Hence, any positive allocation under overbidding is infeasible; if  $f_i(\pi_{\theta^*}(b_i), s) = 0$ , the advertiser receives zero utility. In either case, overbidding does not improve utility.

**Case 2: General**  $\tau_i \geq 1$  We now extend the analysis to the general case. Following the analysis in Wilkens et al. (2017), an advertiser with true type  $(v_i, \tau_i)$  is equivalent to a value maximizer with an “effective value”  $\tilde{v}_i = v_i/\tau_i$  and a normalized ROI constraint  $\tilde{\tau}_i = 1$ . Specifically, the ROI constraint  $\tau_i \cdot p_i \leq v_i \cdot f_i$  can be rewritten as  $p_i \leq (v_i/\tau_i) \cdot f_i$ .

Under this equivalence, the advertiser’s problem of maximizing  $v_i f_i$  subject to  $p_i \leq (v_i/\tau_i) f_i$  is strategically identical to maximizing  $(v_i/\tau_i) f_i$  under the same constraint. Applying the result from Case 1, the optimal strategy is to report a bid equal to the effective value:  $b_i = \frac{v_i}{\tau_i}$ . Consequently, reporting  $b_i = v_i/\tau_i$  is a dominant strategy for an advertiser with type  $(v_i, \tau_i)$ , ensuring that the mechanism is incentive compatible for value maximizers with ROI constraint.  $\square$

## D FURTHER DETAILS OF LLM-NATIVE ADVERTISING SIMULATION ENVIRONMENT

### D.1 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 (Yang et al., 2025). 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 (Kwon et al., 2023).

**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.

### D.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:
  - adid: 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 ADDITIONAL EXPERIMENTAL DETAILS

### E.1 IMPLEMENTATION DETAILS OF LLM-AUCTION

#### E.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  $a_i$  (rep-

resented 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:

$$\hat{f}_i^\psi(y, s) = \sigma(\text{MLP}([\phi_{\text{text}}(x); \phi_{\text{text}}(y); \mathbf{e}_h; \mathbf{e}_{a_i}])), \quad (13)$$

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 (Zhang et al., 2025a) 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  $1 \times 10^{-3}$ .

### E.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:

$$\mathcal{R}(s, \mathbf{b}, y) = \sum_{i=1}^n b_i \cdot \hat{f}_i^\psi(y, s) + \lambda (-10 \cdot N_{\text{ad}}(y)^2 - 500 \cdot \mathbb{1}\{\text{format error}\}), \quad (14)$$

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]`.

### E.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 E.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  $\beta$  is 0.1, the learning rate is  $1 \times 10^{-6}$  with a cosine scheduler and a warmup ratio of 0.1, and the batch size is 64. And we set the threshold  $\delta_{th}$  for constructing preference pairs to 10. For the reward model in Eq. (5), the balancing hyperparameter  $\lambda$  is set to 1.

## E.2 IMPLEMENTATION DETAILS OF BASELINES

**RAG-Auction (Hajiaghayi et al., 2024):** Pre-generation mechanism that allocates ads based on retrieval relevance before the LLM generates the response. Following the setting in LLM-Auction, we reproduce the RAG-Auction using multi-allocation segment auction configuration. Based on the average number of ads inserted by LLM-Auction on the test set, we set the hyperparameter  $K = 3$  for RAG-Auction. Our implementation strictly follows the pseudocode in the original paper, where the retrieval relevance score  $q$  is computed using a pretrained text embedding model. Specifically, we employ the same Qwen3-Embedding-0.6B model used for the pCTR model in LLM-Auction, and define  $q_i$  as the cosine similarity between the embeddings of the ad profile  $a_i$  and the user query  $x$ . While the original RAG-Auction uses a second-price payment rule, we adopt the first-price payment rule to ensure fair comparison of allocation efficiency with LLM-Auction, applying it to compute both revenue and reward per query metrics.

**MOSAIC (Soumalias et al., 2025):** A state-of-the-art post-generation mechanism that samples multiple LLM responses during inference and selects via auction. We reproduce MOSAIC within our proposed LLM-native advertising simulation environment. For the allocation rule, we implement it based on the pseudocode provided in MOSAIC, replacing its reward function with the same reward model used in LLM-Auction as defined in Eq. 14. For the payment rule, since the original payment rule is relatively complex in MOSAIC, we adopt the first-price payment rule as an approximate upper bound to enable fair allocation efficiency comparison with LLM-Auction.

**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.

### E.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 3 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 E.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 3, 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 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 4 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 5 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 6 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.

Table 4: 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 5: 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 6: 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.

---