
Truthful Aggregation of LLMs with an Application to Online Advertising

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 The next frontier of online advertising is revenue generation from LLM-generated
2 content. We consider a setting where advertisers aim to influence the responses
3 of an LLM, while platforms seek to maximize advertiser value and ensure user
4 satisfaction. The challenge is that advertisers’ preferences generally conflict with
5 those of the user, and advertisers may misreport their preferences. To address this,
6 we introduce MOSAIC, an auction mechanism that ensures that truthful reporting is
7 a dominant strategy for advertisers and that aligns the utility of each advertiser with
8 their contribution to social welfare. Importantly, the mechanism operates without
9 LLM fine-tuning or access to model weights and provably converges to the output
10 of the optimally fine-tuned LLM as computational resources increase. Additionally,
11 it can incorporate contextual information about advertisers, which significantly
12 improves social welfare. Via experiments with publicly available LLMs, we
13 show that MOSAIC leads to high advertiser value and platform revenue with
14 low computational costs. While our motivating application is online advertising,
15 our mechanism can be applied in any setting with monetary transfers, making
16 it a general-purpose solution for truthfully aggregating the preferences of self-
17 interested agents over LLM-generated replies.

18 1 Introduction

19 *Large language models (LLMs)* are becoming ubiquitous: as coding assistants, chat interfaces, and
20 even alternatives to search engines [Bommasani et al., 2022]. To ensure their usefulness, it is essential
21 to closely align LLM outputs with user preferences. However, in general, there may be multiple
22 interested parties who disagree on the desired behavior of *the same* LLM. This leads to the question
23 of *how to generate LLM replies given multiple conflicting preferences*.

24 The motivating application for our work is online advertising, the main source of revenue for large
25 tech companies such as Google, Meta, and X/Twitter. For example, in 2023, Meta’s advertising
26 revenue of 132 billion USD was more than 97% of its total revenue [Meta, 2024]. Ad auctions are
27 the workhorse mechanism to determine the placement and prices of commercial content [Varian,
28 2007, Edelman et al., 2007]. As platforms are now working on integrating advertisements into
29 LLM-generated content [Criddle, 2024], new auction mechanisms are needed for this application.

30 We present a new auction mechanism for this problem, which we call *MOSAIC (Mechanism for*
31 *Optimally Sampling and Aggregating LLM Outputs with Incentive Compatibility)*. The advertisers
32 are paying, not for some particular item or bundle, but rather to influence the output generated by an
33 LLM in a direction closer to their own preferences. Although MOSAIC could be of interest whenever
34 one has to aggregate the preferences of multiple self-interested agents over LLM behavior (as long as
35 it is reasonable to charge payments), we see online advertising as the most natural setting of interest.
36 For this reason, we refer to participants in the mechanism as *advertisers*.

37 1.1 Problem setting

38 We consider a situation where a user queries an LLM with a specific question. We assume the
39 following: (i) There is a *reference LLM* that can produce useful replies to the user’s query, and (ii)
40 there are *advertisers* who want to influence the reply to the user. In our model, each advertiser is
41 represented via her own LLM or directly with a *reward function*, similar to the function used to
42 fine-tune LLMs in the *Reinforcement Learning from Human Feedback* pipeline [Azar et al., 2023].
43 For this reason, we refer to an advertiser’s value for a reply as her *reward*. The auctioneer’s task is,
44 given the user’s query and the advertisers’ preferences, to *produce a reply that is useful for the user*
45 *while also generating high rewards for the advertisers*.

46 1.2 Overview of Contributions

47 In this paper, we present MOSAIC, an auction mechanism designed to aggregate the preferences of
48 multiple self-interested advertisers over LLM-generated replies. The first challenge we address is that
49 MOSAIC must produce replies for which the advertisers receive high rewards, but without steering
50 the LLM’s behavior too far from that of the user-centric reference LLM. We address this by drawing
51 a strong connection to the RLHF objective (e.g., Ziegler et al. [2020]), equipping MOSAIC with a
52 hyperparameter that enables the auctioneer to balance between the usefulness of the produced reply
53 to the user and the advertisers in an interpretable and principled way (Section 3.1).

54 Second, leveraging an importance sampling-based technique, MOSAIC is the only mechanism in the
55 LLM mechanism design literature that both provably converges to the optimal distribution (Corol-
56 lary 4.1) and converges efficiently in practice (Section 6). Third, MOSAIC can incorporate contextual
57 information, similar to how sponsored search auctions utilize advertiser descriptions, accelerating
58 convergence and increasing value for advertisers and revenue for the auctioneer (Section 6.2).

59 Fourth, we address technical feasibility and practicality. We adopt the desiderata proposed by Dütting
60 et al. [2024], who argued that auctions must be compatible with existing LLM technology, only using
61 “information obtainable from current models” in a way that is “easy to integrate into the system”
62 and relying only on “easy manipulations of LLM output.” Furthermore, it must be computationally
63 feasible to run the auction repeatedly and with different advertiser preferences. In particular, we
64 cannot afford the expensive process of fine-tuning an LLM’s weights for each possible query. To
65 address this, we design MOSAIC to work by only post-processing multiple LLM outputs – it requires
66 only “API access”, not even viewing the model weights (Section 4.2).

67 Fifth, we address the fact that advertisers might have an interest in misreporting their preferences
68 (which include their *LLMs*), analogous to over-/under-bidding in traditional auctions. To address this
69 challenge, we employ an allocation rule compatible with Rochet payments [Rochet, 1987]. These
70 uniquely ensure that MOSAIC is strategyproof, i.e., that truthful reporting is a dominant strategy for
71 each advertiser, no matter what the others do (Theorem 5.1). Crucially, Rochet payments ensure that
72 MOSAIC remains strategyproof even if the mechanism has not fully converged to the optimal LLM.

73 Sixth, we introduce an offset to the Rochet payments to ensure that each advertiser’s utility is
74 proportional to her contribution to social welfare. This alignment is crucial for the long-term success
75 of the market, as it incentivizes the participation of only the most relevant advertisers.

76 Finally, we discuss *individual rationality (IR)*, which guarantees that a participant is weakly better off
77 from participating in the mechanism. In Section 5.2, we discuss the unique properties of our setting
78 that make *ex-post* IR unattainable. However, in Section 5.3.1, we show that MOSAIC is “almost
79 *ex-post* IR” in a well-defined sense. In Section 6.2, we experimentally show that, MOSAIC is *ex-ante*
80 IR (i.e., advertisers are in expectation better off by participating).

81 In Section 6, we provide experimental results for the online advertising domain. We demonstrate that
82 MOSAIC quickly converges to the optimal LLM with low computational cost, generating significant
83 value for the advertisers and revenue for the auctioneer while also being useful to the user.

84 2 Related Work

85 In their pioneering work, Dütting et al. [2024] introduced the field of mechanism design for LLMs,
86 proposing a sequential mechanism where the output sequence is generated token by token, and

advertisers bid for their LLM to produce the next token. This work laid the foundation for future work in this area. However, several challenges hinder the adoption of their approach in the real world: (i) Advertisers face the *exposure problem*: small changes in the generated sequence, such as the addition of a word like "not," can completely alter its meaning. An advertiser might pay a significant amount for the tokens generated up to a certain point, only for the continuation to unexpectedly negate or distort her intended message.¹ (ii) The mechanism is easily manipulable if the assumption that advertisers cannot misreport their LLMs is dropped. (iii) For a given prompt, an advertiser's spend grows with the length of the generated sequence. MOSAIC addresses all of the above challenges.

Dubey et al. [2024] proposed a generalization of the position auction [Varian, 2007, Edelman et al., 2007] to a setting where each advertiser is interested in having some specific text ad displayed alongside the organic results. In their mechanism, an LLM module coupled with an auction module work in tandem to merge the ads into a single ad summary in an incentive-compatible way.

Hajiaghayi et al. [2024] considered advertisers bidding in an auction to have their ads placed into various portions of a retrieval-augmented generation (RAG) response. Their auction only allows advertisers to report (and misreport) scalar values to influence the inclusion of non-misreportable advertising texts in the combined output. This is in contrast to MOSAIC, which allows the aggregation of full LLMs that advertisers may arbitrarily misreport.

In work that appeared after the initial version of this paper, Bergemann et al. [2024] study a mechanism design problem where agents have both private types and signals, one motivation of which is to model a generalization of our problem. For further discussion of related work, see Appendix A.

3 Framing Sequence Generation as a Mechanism Design Problem

3.1 Formal Model

A user issues a query x . There is a *reference LLM* π_{ref} that the auctioneer aims not to deviate from too much (e.g., because it is responsible for providing useful replies to the user). Additionally, there is a set N of n advertisers who have their own preferences for the reply (i.e., a token sequence) that will be returned to the user. We use the terms *sequence* and *reply* interchangeably.

We let $r_i(x, y)$ denote advertiser i 's *reward* for reply y , given query x .² Informally, the auctioneer's goal is to sample the final reply from a distribution that optimizes the advertisers' expected rewards without substantially diverging from π_{ref} . This is motivated by traditional online ad auctions, where the implicit goal is to maximize advertiser value subject to ensuring a chosen level of usefulness to the user (e.g., the balance between allocating search slots to ads and native content). In the context of an LLM environment, the analogous notion of "usefulness to the user" is represented by the closeness of the final distribution to π_{ref} . Formally, the goal is to choose π to maximize:

$$\mathbb{E}_{y \sim \pi} \left[\sum_{i \in N} r_i(x, y) \right] - \tau D_{\text{KL}}(\pi(\cdot|x) || \pi_{\text{ref}}(\cdot|x)), \quad (1)$$

where $\tau > 0$ is a hyperparameter enabling the auctioneer to balance producing replies closer to the reference LLM or with higher reward for the advertisers, and D_{KL} is the Kullback-Leibler divergence.

This objective is analogous to the standard Reinforcement Learning from Human Feedback (RLHF) approach [Ziegler et al., 2020], but replaces the human feedback reward function with the advertisers' aggregate reward $r(x, y) = \sum_{i \in N} r_i(x, y)$. For an overview of RLHF, see Rafailov et al. [2023, §3]. The optimal solution π_r^* to the optimization problem in (1) was derived by Peters and Schaal [2007]:

$$\pi_r^*(y|x) = \frac{1}{Z(x)} \pi_{\text{ref}}(y|x) \exp \left(\frac{1}{\tau} \sum_{i \in N} r_i(x, y) \right), \quad (2)$$

where $Z(x) = \sum_{y \in T^*} \pi_{\text{ref}}(y|x) \exp \left(\frac{1}{\tau} \sum_{i \in N} r_i(x, y) \right)$ is the partition function.

¹As an example, suppose advertiser A wins the bids for all the tokens in the sequence "Planning your next vacation? For the cheapest flights book via..." However, she loses the bid for the crucial next token, her brand name. She thus pays for the preceding tokens without benefiting from associating her brand with the message.

²In theory, an advertiser's reward for a generated reply could also depend on user-specific information, but we abstract that away. Equivalently, we assume that the advertisers provide user-specific reward functions.

Algorithm 1: Allocation Rule for MOSAIC

Input: User prompt x , reference LLM π_{ref} , LLM used for candidate reply generation π_{gen} , advertiser reward functions $\{r_i\}_{i=1}^n$, number of candidate replies to generate M , reference LLM weight τ

Output: Reply y drawn according to the optimal distribution as defined in Equation (1) for the aggregate reward function $r(x, y) = \sum_{i=1}^N r_i(x, y)$

- 1 Sample $y_j \sim \pi_{\text{gen}}(\cdot|x)$, $1 \leq j \leq M$
 - 2 Calculate $r(x, y_j) = \sum_{i=1}^N r_i(x, y_j)$, $1 \leq j \leq M$
 - 3 **return** $y \sim \text{softmax}\left(\frac{r(x, y_1)}{\tau} + \log \frac{\pi_{\text{ref}}(y_1|x)}{\pi_{\text{gen}}(y_1|x)}, \dots, \frac{r(x, y_M)}{\tau} + \log \frac{\pi_{\text{ref}}(y_M|x)}{\pi_{\text{gen}}(y_M|x)}\right)$
-

127 The performance goal we seek to maximize is $\tau \log \pi_{\text{ref}}(y|x) + \sum_{i \in N} r_i(x, y)$; as it aggregates the
128 interests of both the user and all the advertisers, we refer to it as the *social welfare* of y .

129 Let \hat{R} be the set of all possible reports by the advertisers. A *mechanism* is defined as a pair (π, p) .
130 The *allocation rule* $\pi : \hat{R} \rightarrow (T^* \rightarrow \Delta(T^*))$ maps any report profile $\hat{r} = (\hat{r}_1, \hat{r}_2, \dots, \hat{r}_n) \in \hat{R}$ of the
131 advertisers' rewards to an LLM $\pi_{\hat{r}}$, which in turn is a mapping from a user query x to a distribution
132 over token sequences $\delta(T^*)$. We denote the LLM that the allocation rule π induces for reports \hat{r} as
133 $\pi_{\hat{r}}$, and the optimal LLM for those reports (i.e., the maximizer of Equation (1)) as $\pi_{\hat{r}}^*$. The *payment*
134 *rule* $p : \hat{R} \rightarrow \mathbb{R}^n$ maps any report profile \hat{r} to a payment profile $p(\hat{r})$, where $p_i(\hat{r})$ is the payment of
135 the i -th advertiser. We aim for *strategyproof* mechanisms, meaning that no advertiser has an incentive
136 to misreport her preferences:

137 **Definition 3.1** (Strategyproof). A mechanism (π, p) is strategyproof if, for any advertiser $i \in N$, true
138 reward function r_i , reported reward function \hat{r}_i , reported reward functions \hat{r}_{-i} by the other advertisers
139 and prompt x : $\mathbb{E}_{y \sim \pi(r_i, \hat{r}_{-i})(\cdot|x)}[r_i(x, y) - p_i(r_i, \hat{r}_{-i})] \geq \mathbb{E}_{y \sim \pi(\hat{r}_i, \hat{r}_{-i})(\cdot|x)}[r_i(x, y) - p_i(\hat{r}_i, \hat{r}_{-i})]$,
140 where $r_i(x, y) - p_i(\hat{r}_i, \hat{r}_{-i})$ is the utility of advertiser i for reply y when her payment is $p_i(\hat{r}_i, \hat{r}_{-i})$.

141 3.2 The shortcomings of VCG in this setting

142 At first sight, it may seem that the Vickrey–Clarke–Groves (VCG) mechanism would be suitable for
143 our setting [Vickrey, 1961, Clarke, 1971, Groves, 1973]. VCG selects the outcome that maximizes
144 the sum of all agents' values. This can be either the single optimal sequence in Equation (1) or
145 the optimal distribution of Equation (2). The VCG mechanism has a corresponding payment rule
146 to incentivize truthful reporting: it charges each agent her externality, the total reduction in value
147 (respectively expected value) for the other agents that her participation in the mechanism caused.

148 However, in our setting, VCG is *not* a viable option: VCG requires calculating the *exact* optimal solu-
149 tion to the optimization problem, which is intractable for choosing an LLM to maximize Equation (1)
150 and is even difficult for choosing a single optimal sequence. If a sub-optimal solution is chosen,
151 VCG's strategyproofness is no longer guaranteed [Nisan and Ronen, 2007, 1999, Lehmann et al.,
152 2002]. We provide an intuitive example illustrating this failure for two advertisers in Appendix B.1.

153 4 The MOSAIC Mechanism: Allocation Rule

154 4.1 Convergence to Optimality, Advertiser Contexts and Importance Sampling

155 In this section, we introduce MOSAIC's allocation rule. The high-level idea is as follows: first, a
156 set of M candidate replies is generated based on an LLM π_{gen} . Then, the probability of returning
157 each candidate is re-weighted based on the advertisers' reports and the reference LLM π_{ref} , so that
158 as $M \rightarrow \infty$, the return probability of each reply converges to that under the optimal distribution
159 of Equation (2). This approach resembles importance sampling techniques that have been used in
160 various LLM training pipelines (e.g., Xie et al. [2023]). All proofs are deferred to Appendix B.

161 **Corollary 4.1.** For any reported reward functions $r \in R$ and any LLM π_{gen} such that π_{ref} is absolutely
162 continuous with respect to π_{gen} , the MOSAIC policy $\pi_{r, M}(\cdot|x)$ from Algorithm 1, using M candidate
163 replies, converges to an optimal solution for the platform's objective (Equation (1)) as $M \rightarrow \infty$.
164 Formally, $\lim_{M \rightarrow \infty} \pi_{r, M}(\cdot|x) = \pi_r^*(\cdot|x) \in \arg \max_{\pi \in \Delta(T^*)} \mathbb{E}_{y \sim \pi(\cdot|x)}[r(x, y)] - \tau D_{\text{KL}}(\pi || \pi_{\text{ref}})$.

Based on Corollary 4.1, MOSAIC converges to the optimal distribution for any LLM π_{gen} , provided that π_{ref} is absolutely continuous with respect to π_{gen} . While $\pi_{\text{gen}} = \pi_{\text{ref}}$ is an intuitive choice, it would result in impractically slow convergence rates. The reason is that advertisers have high rewards for responses that explicitly mention their brands, but π_{ref} considers replies with mentions of specific brands extremely unlikely. Consequently, generating candidate replies directly from π_{ref} leads Algorithm 1 to sample from a set of low-reward candidates, hindering performance.

To address this challenge, we generate candidate sequences not from $\pi_{\text{ref}}(\cdot|x)$, but instead from a *context-aware LLM*, $\pi_{\text{con}}(\cdot|x; c)$. The instance-specific context c is designed to bridge the gap between the parts of the output space favored by π_{ref} and those valued by the advertisers.

We formalize this intuition in Appendix B.3, connecting MOSAIC’s allocation rule to an importance-based sampling estimator for π_r^* and proving the following lemma:

Lemma 4.2. *For any LLM π_{gen} such that π_{ref} is absolutely continuous with respect to π_{gen} , the variance of $\pi_{r,M}(\cdot|x)$ as an estimator of $\mathbb{E}_{y \sim \pi_r^*(\cdot|x)}[\pi_r^*(y|x)]$ is*

$$\frac{1}{M} \left(\sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{gen}}(y|x)} - \left(\sum_{y \in Y} \pi_r^*(y|x)^2 \right)^2 \right).$$

Applying Chebyshev’s inequality to the variance of Lemma 4.2, MOSAIC converges to π_r^* at a rate of \sqrt{M} (Lemma B.3).³ However, this convergence rate also depends on $\frac{\pi_r^*(y|x)^4}{\pi_{\text{gen}}(y|x)}$. Thus, generating replies via an LLM closer to π_r^* reduces the estimator’s variance and improves convergence speed.

In our application of integrating advertisers’ interests into LLM outputs, c_i is a context-specific description of the i -th advertiser. These descriptions, supplied by the advertisers themselves, should be easily verifiable and factually accurate, akin to “MusicMastery: offering online music lessons”, or “InstaTune: selling musical instruments.”⁴ This approach is analogous to search engine optimization in sponsored search advertising, where advertisers supply and potentially optimize their own descriptions to influence how they are presented by the auction mechanism.

Our experiments in Section 6.2 demonstrate that using the context-aware LLM to generate candidate replies achieves substantially higher rewards and utility for the advertisers, increased revenue for the auctioneer, and faster convergence. In the rest of the paper, we refer to using the reference and context-aware LLMs as the baseline and context-aware versions of our mechanism, respectively.

4.2 Practical Considerations

Input Methods and Black-Box Access to Advertiser LLMs. MOSAIC’s allocation and payment rules do not depend on the advertisers’ full reward functions, but only on their rewards for the candidate replies. Thus, MOSAIC requires only “API access” to the involved LLMs without fine-tuning or access to their weights. Rafailov et al. [2023] established a mapping between an agent’s LLM and her implicit reward function, allowing MOSAIC to use as inputs reply probabilities (i.e., LLM inference calls) instead of rewards. For more details, see Appendix C.2.

Static Setting and Incentive Implications. The number M of candidate replies considered by MOSAIC is predetermined. Each advertiser observes all M candidate replies before submitting her reports, resulting in a *static* rather than a *dynamic* setting (e.g., as in Dütting et al. [2024]), where advertisers interact sequentially with evolving information, significantly complicating incentive considerations. Theorem 5.1 ensures that truthful reporting is a dominant strategy for each advertiser.

Theoretical Complexity. Generating a single candidate reply of length L tokens requires L forward passes through π_{gen} , as tokens are generated sequentially in an autoregressive manner. Evaluating each reply processes the entire sequence in a single forward pass (and is compatible with efficient evaluation methods, e.g. Li et al. [2024]). For M candidate replies and n advertisers, MOSAIC requires $M \cdot (L + n + 1)$ total forward passes. In contrast, Dütting et al. [2024] require $L \cdot n$ forward passes, as all n advertiser LLMs participate in generating each token. Section 6 shows that MOSAIC achieves convergence with $M = 20$ candidate replies, reducing computational cost for large n .

³In Appendix B.4 we establish a more general result, showing that for any sequence y , its estimated probability converges to the corresponding probability under π_r^* at a rate of $1/\sqrt{M}$.

⁴A practical way of implementing $\pi_{\text{con}}(\cdot|x; c)$ given $\pi_{\text{ref}}(\cdot|x)$ is to augment the input x to the reference LLM with the advertiser descriptions. In our example: “Try to mention ⟨advertiser x⟩, ⟨advertiser x description⟩.”

211 **Compatibility with Efficient Methods and Practical Costs.** MOSAIC relies solely on LLM
 212 forward passes, the core operation modern architectures are optimized for, enabling it to take full
 213 advantage of existing optimizations for efficient generation. As a result, MOSAIC converges to the
 214 optimal distribution with compute costs equivalent to just five LLM queries (Appendix D.8).

215 **Parallelization and User-Perceived Latency.** The generation and evaluation of each candidate
 216 sequence are independent processes, allowing MOSAIC to be fully parallelized. In a fully parallelized
 217 setting, the response time for a user query is comparable to directly querying a *single* LLM.

218 **Inherent Competition.** Unlike conventional auctions, MOSAIC ensures high baseline revenue even
 219 in low-competition environments, as advertisers always compete against π_{ref} . Traditional auctions
 220 rely on reserve prices to boost revenue, but a slight miscalibration above a critical threshold can result
 221 in zero revenue. In contrast, MOSAIC uses a *single* tunable parameter, τ , and its revenue varies
 222 smoothly and differentially with τ (see the proof of Theorem 5.1). As a result, revenue optimization
 223 in MOSAIC is significantly more robust to parameter tuning.

224 5 The MOSAIC Mechanism: Payment Rule

225 In this section, we first show how the allocation rule from Section 4 can be combined with an
 226 appropriate payment rule so that the resulting mechanism is strategyproof (Section 5.1). Then, we
 227 detail how auctions for LLM-generated content differ from standard auctions (Section 5.2). Taking
 228 those differences into account, we create a payment offset, so that MOSAIC is both strategyproof and
 229 social welfare-aligned (Section 5.3). We defer all proofs to Appendix C.

230 5.1 Strategyproofness through Cyclic Monotonicity

231 The allocation rule of Algorithm 1 satisfies *cyclic monotonicity* [Rockafellar, 1970, § 24]. Rochet
 232 [1987] first proposed the use of cyclic monotonicity in mechanism design as a generalization of
 233 monotonicity in single-parameter settings [Myerson, 1981]. For general settings, cyclic monotonicity
 234 of the allocation rule is a sufficient and necessary condition for us to prove:

235 **Theorem 5.1.** *The allocation rule of Algorithm 1 can be combined with a payment rule such that*
 236 *for any advertiser $i \in N$ and set of candidate replies $\{y_j\}_{j=1}^M$, reporting truthfully is a dominant*
 237 *strategy. Advertiser i 's expected utility (up to a constant of integration C) under truthful reporting is:*

$$238 \tilde{U}_i(r_i, \hat{r}_{-i}; \pi_{\text{ref}}, \pi_{\text{gen}}) = C + \tau \log \left(\sum_{j=1}^M \exp \left(\frac{\sum_{k \in N \setminus \{i\}} \hat{r}_k(x, y_j) + r_i(x, y_j)}{\tau} \right) + \log \frac{\pi_{\text{ref}}(y_j | x)}{\pi_{\text{gen}}(y_j | x)} \right).$$

239 Note that, based on Theorem 5.1, in MOSAIC it is *always* a dominant strategy for an advertiser to
 240 report truthfully. Crucially, this is not the case for VCG, where truthful reporting would be optimal
 241 only if the allocation rule had converged to the optimal distribution (Example 1).

242 5.2 Differences from Standard Auction Settings

243 Auction mechanisms designed to sell items or ad slots typically rely on simplifying assumptions that
 244 do not apply in a setting with LLM-generated content. Key differences include: (i) *Non-Negative*
 245 *Values:* These mechanisms assume agents' values are non-negative due to having zero value for the
 246 empty bundle and free disposal. In contrast, in our setting, an advertiser's reward can be negative based
 247 on the discrepancy between her LLM and the reference LLM. (ii) *Advertiser-Specific Allocations:*
 248 These mechanisms allocate different item bundles to different agents. Here, a single reply is produced
 249 (iii) *Zero Utility for Non-Participation:* In most auction settings, not participating yields zero utility.
 250 Here, non-participation can result in negative utility since the produced reply may be unfavorable to
 251 non-participating advertisers.⁵ For details, see Appendix C.2.

252 5.3 Advertiser-specific Utility Offset

253 We now modify MOSAIC's payment rule by adding a *payment offset* (and thus a utility offset):
 254 $C = -\tilde{U}_i(0, \hat{r}_{-i}; \pi_{\text{ref}}, \pi_{\text{con}})$. Informally, we additionally charge each advertiser her utility in

⁵While we focus on the most general model that includes negative rewards, MOSAIC maintains all its desirable properties discussed in Sections 4 and 5 under the assumption that all advertiser rewards are nonnegative.

Theorem 5.1 if her reward for all candidate replies was zero. This offset maintains the key properties of our mechanism (i.e., strategyproofness and convergence to the optimal distribution), while also achieving two additional properties that are critical for the long-term success of a market for LLM aggregation. The first is “Almost IR:” An advertiser with weakly positive reward for all candidate replies has weakly positive expected utility for all reports by the other advertisers. The second is “What you give is what you get:” an advertiser’s expected utility is monotone in how well-aligned her exponentiated reward for the replies is with the interim allocation rule if she were to not participate.

5.3.1 “Almost Individually Rational”

Individual rationality is important to incentivize agents to participate in the mechanism. In Appendix C.3, we explain how the unique properties of our setting, namely lack of free disposal and a common outcome for all agents, make the standard notion of individual rationality (i.e., weakly positive utility from participation) unobtainable while converging to the optimal distribution and maintaining strategyproofness. Then, we explain how, our payment offset achieves “almost IR:” In Lemma C.1 we prove that the expected utility of an advertiser who has zero reward for all candidate replies and bids truthfully is zero, i.e., advertisers that do not contribute to the social welfare (but also do not detract from it) have zero utility. Similarly, in Lemma C.2 we prove that if an advertiser’s reward for all candidate sequences is (weakly) positive, then her expected utility is (weakly) positive.

Remark 1. In Section 6, we experimentally show that our offset payment rule, coupled with our context-aware allocation rule, results in both high expected rewards and positive utility for the advertisers (i.e., ex-ante individual rationality), as well as significant revenue for the auctioneer.

5.3.2 “What you give is what you get”

Our allocation rule, which is the only one over a finite set of replies that converges to the optimal LLM, is also the (sub)gradient of the utility to ensure truthfulness [Rochet, 1987]. Because the allocation rule is the same for all advertisers, their utilities must also be the same, up to advertiser-specific offsets, as indicated by Theorem 5.1. However, not all advertisers contribute equally to social welfare. In Appendix C.4, we explain why implementing the mechanism without carefully designed offsets would lead to free-riding. Specifically, if an advertiser’s expected utility is positive, even when her relevance to the user query is low, it incentivizes her participation. This dilutes the utility for all participants, discouraging more relevant advertisers from participating. As a result, the expected rewards for advertisers decline, and the generated replies become less useful for users. To avoid free-riding, it is crucial to align each advertiser’s utility with her contribution to social welfare.

Lemma 5.2. For the offset $C = -\tilde{U}_i(0, \hat{r}_{-i}; \pi_{ref}, \pi_{con})$ advertiser i ’s expected utility under truthful reporting is $U_i(r_i, \hat{r}_{-i}; \pi_{ref}, \pi_{con}) = \tau \log \left(\sum_{j=1}^M \exp \left(\frac{r_i(x, y_j)}{\tau} \right) \pi_{\hat{r}_{-i}}(y_j | x) \right)$.

In words, Lemma 5.2 shows that each advertiser’s expected utility is monotone in how well aligned her exponentiated reward function is with MOSAIC’s induced LLM $\pi_{\hat{r}_{-i}}$, had she not participated.

Remark 2. In Section 6.2, we experimentally show that the of Section 5.3 induces a strong positive correlation between an advertiser’s contribution to social welfare and her expected utility gain from participating. Moreover, for the tested distribution of instances, MOSAIC satisfies ex-ante IR.

6 Experiments

In this section, we evaluate MOSAIC’s performance on its flagship application of online advertising.

6.1 Experiment Setup

We create synthetic instances, each comprising a user query (e.g., “How to learn a musical instrument online?”) and two advertisers (e.g., “MusicMastery, offering online music lessons”). This matches the setup of Dütting et al. [2024] while highlighting MOSAIC’S performance and revenue, even in low competition scenarios. We use Llama-2-7b-chat-hf [Touvron et al., 2023] as the base architecture for all LLMs. In Appendices D.6 and D.10 we extend our analysis to settings with more advertisers and alternative architectures, observing similarly strong results. See Appendix D for details.

User query
What are effective ways to learn a musical instrument online?
Advertisers
- MusicMastery: offering online music lessons and virtual tutoring - InstaTune: selling musical instruments and learning aids
Context-Aware Mechanism Output
Effective ways to learn a musical instrument online include: 1. Online music lessons: Platforms like MusicMastery offer one-on-one virtual tutoring, group classes, and pre-recorded lessons. These lessons can be tailored to your skill level and learning style. 2. Interactive learning tools: InstaTune offers a range of interactive learning tools, such as apps, software, and games, that make learning fun and engaging....
Baseline Mechanism Output
- Practice consistently and regularly: Consistent practice is essential to learning any musical instrument, and it's even more important when learning online. Set aside a specific time each day or week to practice, and make sure to stick to it. - Find a good teacher or mentor: While it's possible to learn some things on your own, working with a good teacher or mentor can help you improve faster and avoid developing bad habits. ...

Figure 1: Comparison of context-aware and baseline mechanism outputs using 10 candidate replies.

6.2 Experimental Results

In this section, we evaluate MOSAIC’s overall effectiveness, assessing both its allocation and payment rules. We first focus on how well the allocation rule generates value for advertisers and converges to the optimal distribution, comparing the baseline and context-aware versions. Figure 1 illustrates how the context-aware mechanism enhances response relevance for advertisers. Notably, only the context-aware mechanism successfully incorporates advertisers into the replies.

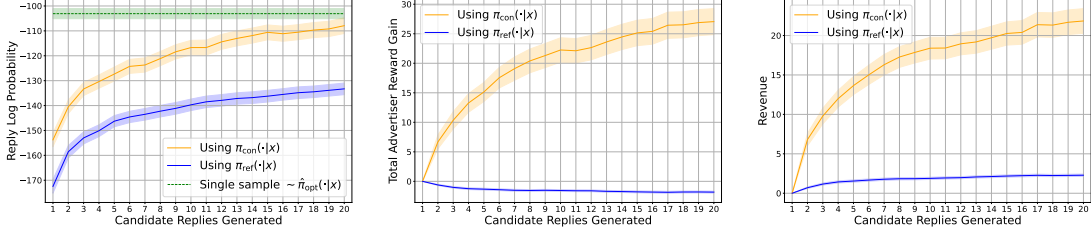
In Figure 2a, we plot the log probability of the replies returned by MOSAIC with respect to the optimal distribution (Equation (2)) against the number M of candidate replies generated. We compare the context-aware version of MOSAIC to the baseline version to evaluate the effectiveness of incorporating contextual information. Note that baseline MOSAIC is the *strongest applicable benchmark*, as it is the *only* other tractable mechanism for this problem in the literature that converges to the optimal distribution (for a detailed discussion, please see Appendix D.2).

To assess MOSAIC’s convergence, we estimate the log probability of sampling a *single reply* from the optimal distribution.⁶ We observe that for both the context-aware and baseline versions of MOSAIC, the log probability of the returned reply scales rapidly with the number of candidate replies. This aligns with our analysis in Section 4, where we proved that both versions converge to the optimal distribution. We also observe that incorporating context significantly accelerates convergence. Notably, the context-aware version of MOSAIC achieves higher log probability with respect to the optimal distribution with four candidate replies than the baseline version with 20. Additionally, with only 20 candidate replies, the context-aware mechanism can almost match our estimate of the log probability of sampling from the optimal distribution. In Appendix D.4, we further demonstrate that the usefulness to the user also scales rapidly with the number of candidate replies.

In Figure 2b, we plot the total expected reward for advertisers against the number of candidate replies generated, with the expectation over the distribution of the returned reply from the set of candidate replies. To make the plot more interpretable, we normalize each advertiser’s expected reward by her expected reward had she not participated. Specifically, we plot $r_{i,\text{norm}}(x) = \sum_{j \in M} r_i(x, y_j) \cdot \pi_r(y_j|x) - \sum_{j \in M} r_i(x, y_{-i,j}) \cdot \pi_{r_{-i}}(y_{-i,j}|x)$, where y_{-i} are the candidate replies that would have been generated if advertiser i had not participated, and $\pi_{r_{-i}}$ MOSAIC’s allocation rule in that case.⁷ We compare MOSAIC’s two versions, highlighting the value of incorporating context.

⁶The closed-form solution of Equation (2) allows us to evaluate the probability of sentences with respect to the optimal solution, but it does not enable us to sample from that distribution, which would require using RL to train the optimal LLM on the advertisers’ aggregate reward function, which is computationally infeasible for the number of problem instances we test. Instead, we generate replies from π_{ref} and evaluate them based on the induced probabilities of π_{ref} , for which π_{ref} is the optimal LLM. This serves as a proxy for the log probabilities we should expect if we were to draw replies from the optimally fine-tuned model for each query.

⁷To reduce computational costs, we estimate an advertiser’s reward for not participating based on her expected reward over the already generated replies in which her brand is not mentioned by name, motivated by the fact that if she does not participate, her brand will not be mentioned.



(a) Returned reply log probability as a function of the number of candidate replies generated using π_{ref} and π_{con} . We also show a proxy of drawing a single sequence from the optimal distribution.

(b) Total advertiser normalized reward as a function of the number of candidate replies generated using π_{ref} and π_{con} .

(c) Revenue as a function of the number of replies generated using π_{ref} and π_{con} .

Figure 2: Reply log probability, total advertiser normalized reward, and revenue as a function of the number of candidate replies generated using π_{ref} and π_{con} . Averages over 1250 runs with 95% CIs.

Figure 2b demonstrates that our context-aware mechanism significantly boosts advertiser rewards, with benefits scaling rapidly with the number of candidate replies. In contrast, the baseline version fails to improve advertiser rewards and may even cause a slight decrease: generating more replies adds marginal positive reward to the reference LLM (see Appendix D.4), but does so at the expense of advertisers. In summary, MOSAIC’s support for context-aware LLMs allows it to quickly converge to the optimal distribution and generate substantial rewards for advertisers. In an ablation study (Appendix D.7), we demonstrate that MOSAIC maintains strong performance across all key metrics even under significantly reduced contextual input.

In Appendix D.8 we provide a detailed experimental evaluation of MOSAIC’s compute requirements, showing how it requires the compute cost of just five LLM queries to converge, and discuss how it can be fully parallelized so that the user-perceived latency is the same as a *single* LLM query.

Next, we examine the impact of our payment rule from Section 5. In Figure 2c, we plot the auctioneer’s revenue for both the context-aware and baseline versions of MOSAIC. Our payment rule in combination with the context-aware allocation rule yields revenue that scales rapidly with the number of candidate replies. Moreover, comparing the advertisers’ total payment with their normalized reward in Figure 2b, we see that for the context-aware mechanism, our payment rule converts much of the advertisers’ surplus into revenue, while also ensuring positive advertiser utility.

These results align with our theoretical analysis. In Section 5.3.1, we showed that an advertiser with positive rewards for candidate replies—as in the context-aware mechanism—achieves positive expected utility, and in Section 5.3.2 we introduced the “what you give is what you get” property. In Appendix D.5 we show that, for both versions of our mechanism, the payment offset increases the advertisers’ expected utility, and makes the relationship between an advertiser’s contribution to social welfare and utility significantly more linear and positively correlated. Concretely, our offset increases the Pearson correlation between advertiser utility and reward from 0.4 to 0.8.

7 Conclusion

We have introduced MOSAIC, a novel auction mechanism for aggregating preferences over LLM outputs. MOSAIC provably converges to the theoretically optimal distribution and it also facilitates a principled method for balancing participants’ expected rewards with the divergence from a reference LLM. Thus, our mechanism is particularly well-suited for online advertising, allowing the integration of advertiser LLMs with a reference LLM responsible for generating user-centric replies.

Our payment rule removes any incentive to misreport preferences, achieving the central mechanism design goal of strategyproofness. While ex-post IR is incompatible with strategyproofness in our setting, we experimentally show that our mechanism is ex-ante individually rational and “almost individually rational” in a certain sense. Furthermore, it ensures that each participant’s utility gain is proportionate to her contribution to social welfare, an essential alignment property in this setting.

Experimentally, we have demonstrated that by incorporating contextual information, MOSAIC’s outputs rapidly converge to the optimal distribution, generating significant value for the advertisers while also effectively recapturing a considerable portion of this value as revenue. These findings demonstrate the practical efficacy and potential of our approach in realistic settings.

References

- Mohammad Gheshlaghi Azar, Mark Rowland, Bilal Piot, Daniel Guo, Daniele Calandriello, Michal Valko, and Rémi Munos. A general theoretical paradigm to understand learning from human preferences. *arXiv preprint arXiv:2310.12036*, 2023.
- Dirk Bergemann, Marek Bojko, Paul Dütting, Renato Paes Leme, Haifeng Xu, and Song Zuo. Data-Driven Mechanism Design: Jointly Eliciting Preferences and Information, December 2024.
- Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2022.
- Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Alex Castro-Ros, Marie Pellat, Kevin Robinson, Dasha Valter, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Zhao, Yanping Huang, Andrew Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. Scaling instruction-finetuned language models, 2022. URL <https://arxiv.org/abs/2210.11416>.
- Edward Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971.
- Vincent Conitzer, Rachel Freedman, Jobst Heitzig, Wesley H. Holliday, Bob M. Jacobs, Nathan Lambert, Milan Mossé, Eric Pacuit, Stuart Russell, Hailey Schoelkopf, Emanuel Tewelde, and William S. Zwicker. Social choice for ai alignment: Dealing with diverse human feedback. *arXiv preprint arXiv:2404.10271*, 2024.
- Cristina Criddle. Perplexity in talks with top brands on ads model as it challenges Google. *Financial Times*, September 2024.
- Kumar Avinava Dubey, Zhe Feng, Rahul Kidambi, Aranyak Mehta, and Di Wang. Auctions with llm summaries. *arXiv preprint arXiv:2404.08126*, 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, page 144–155, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400701719.
- 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.
- Soheil Feizi, MohammadTaghi Hajiaghayi, Keivan Rezaei, and Suho Shin. Online advertisements with llms: Opportunities and challenges. *arXiv preprint arXiv:2311.07601*, 2024.

- 421 Sara Fish, Paul Gözl, David C. Parkes, Ariel D. Procaccia, Gili Rusak, Itai Shapira, and Manuel
422 Wüthrich. Generative social choice. *arXiv preprint arXiv:2309.01291*, 2023.
- 423 Rafael M. Frongillo and Ian A. Kash. General truthfulness characterizations via convex analysis.
424 *Games and Economic Behavior*, 130:636–662, 2021.
- 425 Theodore Groves. Incentives in teams. *Econometrica*, 41(4):617–631, 1973.
- 426 MohammadTaghi Hajiaghayi, Sébastien Lahaie, Keivan Rezaei, and Suho Shin. Ad Auctions for
427 LLMs via Retrieval Augmented Generation, June 2024.
- 428 Keegan Harris, Nicole Immorlica, Brendan Lucier, and Aleksandrs Slivkins. Algorithmic persuasion
429 through simulation. *arXiv preprint arXiv:2311.18138*, 2024.
- 430 Daniel Lehmann, Liadan Ita O’callaghan, and Yoav Shoham. Truth revelation in approximately
431 efficient combinatorial auctions. *Journal of the ACM (JACM)*, 49(5):577–602, 2002.
- 432 Kenneth Li, Samy Jelassi, Hugh Zhang, Sham Kakade, Martin Wattenberg, and David Brandfonbrener.
433 Q-probe: A lightweight approach to reward maximization for language models. *arXiv preprint*
434 *arXiv:2402.14688*, 2024.
- 435 Meta. Meta reports fourth quarter and full year 2023 results, initiates quarterly dividend, 2024.
436 URL [https://investor.fb.com/investor-news/press-release-details/2024/
437 Meta-Reports-Fourth-Quarter-and-Full-Year-2023-Results-Initiates-Quarterly-Dividend/
438 default.aspx](https://investor.fb.com/investor-news/press-release-details/2024/Meta-Reports-Fourth-Quarter-and-Full-Year-2023-Results-Initiates-Quarterly-Dividend/default.aspx). Meta Investor Relations, accessed: 2024-01-31.
- 439 Roger B Myerson. Optimal Auction Design. *Mathematics of Operations Research*, 6(1), February
440 1981.
- 441 N. Nisan and A. Ronen. Computationally Feasible VCG Mechanisms. *Journal of Artificial Intelli-*
442 *gence Research*, 29:19–47, May 2007. ISSN 1076-9757.
- 443 Noam Nisan and Amir Ronen. Algorithmic mechanism design. In *Proceedings of the thirty-first*
444 *annual ACM symposium on Theory of computing*, pages 129–140, 1999.
- 445 Jan Peters and Stefan Schaal. Reinforcement learning by reward-weighted regression for operational
446 space control. In *Proceedings of the 24th International Conference on Machine Learning, ICML*
447 *’07*, page 745–750, New York, NY, USA, 2007. Association for Computing Machinery. ISBN
448 9781595937933.
- 449 Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea
450 Finn. Direct preference optimization: Your language model is secretly a reward model. *arXiv*
451 *preprint arXiv:2305.18290*, 2023.
- 452 Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear
453 context. *Journal of Mathematical Economics*, 16(2):191–200, 1987.
- 454 Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970. ISBN
455 9781400873173.
- 456 Baptiste Rozière, Jonas Gehring, Fabian Gloeckle, Sten Sootla, Itai Gat, Xiaoqing Ellen Tan, Yossi
457 Adi, Jingyu Liu, Romain Sauvestre, Tal Remez, Jérémy Rapin, Artyom Kozhevnikov, Ivan Evtimov,
458 Joanna Bitton, Manish Bhatt, Cristian Canton Ferrer, Aaron Grattafiori, Wenhan Xiong, Alexandre
459 Défossez, Jade Copet, Faisal Azhar, Hugo Touvron, Louis Martin, Nicolas Usunier, Thomas
460 Scialom, and Gabriel Synnaeve. Code llama: Open foundation models for code. *arXiv preprint*
461 *arXiv:2308.12950*, 2024.
- 462 Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay
463 Bashlykov, Soumya Batra, Prajjwal Bhargava, Shrutu Bhosale, Dan Bikel, Lukas Blecher, Cris-
464 tian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu,
465 Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn,
466 Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel
467 Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee,
468 Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra,

469 Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi,
470 Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh
471 Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen
472 Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic,
473 Sergey Edunov, and Thomas Scialom. Llama 2: Open foundation and fine-tuned chat models.
474 *arXiv preprint arXiv:2307.09288*, 2023.

475 Hal R. Varian. Position auctions. *International Journal of Industrial Organization*, 25(6):1163–1178,
476 December 2007.

477 William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of*
478 *Finance*, 16(1):8–37, 1961.

479 Tobias Werner, Ivan Soraperra, Emilio Calvano, David C. Parkes, and Iyad Rahwan. Experimental
480 Evidence That Conversational Artificial Intelligence Can Steer Consumer Behavior Without
481 Detection, September 2024.

482 Sang Michael Xie, Shibani Santurkar, Tengyu Ma, and Percy Liang. Data selection for language
483 models via importance resampling. In *Proceedings of the 37th International Conference on Neural*
484 *Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.

485 Daniel M. Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B. Brown, Alec Radford, Dario Amodei, Paul
486 Christiano, and Geoffrey Irving. Fine-tuning language models from human preferences. *arXiv*
487 *preprint arXiv:1909.08593*, 2020.

A Further related work

Feizi et al. [2024] presented an abstract design for an LLM advertising system and detailed a number of research challenges that would have to be overcome in the course of implementation. Conitzer et al. [2024] drew connections between *computational social choice* and LLM alignment. Fish et al. [2023] presented work in the opposite direction: how can LLMs be used to solve problems in social choice? Harris et al. [2024] studied Bayesian persuasion in an abstract setting where a “simulator” (for example, a realistic LLM) of the agent is available. Werner et al. [2024] gave preliminary empirical evidence that LLM-generated content can influence consumers.

In work that appeared after the initial version of this paper, Bergemann et al. [2024] considers a setting where bidders have both private types reflecting their preferences and private signals reflecting information about a world state (for example, user interests). When applied to LLM auctions, they identify their setting as a generalization of ours, where the signals known to the advertisers may actually improve the output for the user, and where deviations from the reference LLM may be measured by arbitrary distance functions. They analyze conditions under which a welfare maximizing mechanism can be implemented truthfully (eliciting both types and signals) in posterior equilibrium, but do not consider practical implementations of their mechanism.

B Details from Section 4

In this section, we present all omitted theorems and proofs from Section 4.

B.1 An Example Where VCG Is Not Strategyproof in This Setting

In Example 1, we present a simple example illustrating that the VCG mechanism is not strategyproof when the allocation rule cannot fully converge to the optimal LLM based on the advertisers’ reports.

Example 1. *Suppose there are two advertisers, A and B , and five possible replies:*

- *two replies that mention both advertisers,*
- *one reply that mentions only A ,*
- *one reply that mentions only B ,*
- *and one reply that mentions neither advertiser.*

The advertisers’ rewards for each reply are shown in Table 1:

Outcome	r_A	r_B
$\{A, B\}_1$	0.9	0.9
$\{A, B\}_2$	0.4	0.4
$\{A\}$	1.0	0.0
$\{B\}$	0.0	0.5
\emptyset	0.0	0.0

Table 1: Advertiser rewards for different replies.

Assume that $\tau = 0$, so that the optimal LLM would deterministically select $\{A, B\}_1$, maximizing total advertiser reward. However, suppose that the allocation rule cannot converge to this reply—for instance, because RLHF is used to train the aggregate LLM $\pi_{\hat{\tau}}$ and the process is noisy. Instead, when aiming for $\{A, B\}_1$, the resulting LLM π produces $\{A, B\}_2$.⁸

Under truthful reports, advertiser A ’s utility would be

$$u_A = r_A(\{A, B\}_2) + r_B(\{A, B\}_2) - r_B(\{B\}) = 0.4.$$

⁸Alternatively, one could interpret this setup as a setting where the mechanism cannot converge to replies mentioning both advertisers, perhaps due to optimization difficulties, but can converge to replies mentioning only one advertiser.

520 However, if advertiser A misreports by declaring $\hat{r}_A(\{A, B\}_1) = 0.0$, the mechanism would instead
 521 converge to an LLM that always returns $\{A\}$. In that case, A 's utility would be

$$u_A = r_A(\{A\}) + r_B(\{A\}) - r_B(\{B\}) = 0.5.$$

522 Thus, advertiser A would be incentivized to misreport her preferences, illustrating how VCG payments
 523 cannot ensure strategyproofness when the allocation chosen may be suboptimal.

524 B.2 Proving that MOSAIC converges to the target distribution in the limit

525 **Theorem B.1.** Let $\pi_{\hat{r},M}(y|x)$ be the probability of sampling output sequence y for input sequence x
 526 according to Algorithm 1, where M is the number of candidate sequences generated and $\hat{r} \in \hat{R}$ is
 527 the advertisers' reported reward functions. For any LLM π_{gen} such that π_{ref} is absolutely continuous
 528 with respect to π_{gen} , the policy induced by MOSAIC approaches the following limit:

$$\lim_{M \rightarrow \infty} \pi_{\hat{r},M}(y|x) = \pi_{\text{ref}}(y|x) \frac{\exp(\hat{r}(x, y)/\tau)}{\mathbb{E}_{y' \sim \pi_{\text{ref}}(\cdot|x)}[\exp(\hat{r}(x, y')/\tau)]} \quad (3)$$

529 *Theorem B.1 Proof.* Let $\pi_{\hat{r},M}(y|x, \{y_j\}_{j=1}^M)$ be the probability of returning output sequence y for
 530 input sequence x according to Algorithm 1 conditioned on the set of generated candidate sequences
 531 being $\{y_j\}_{j=1}^M$. Additionally, let $\pi_{\text{con}}(\{y_j\}_{j=1}^M|x; c)$ be the probability of the context-aware model
 532 π_{con} generating the candidate sequences $\{y_j\}_{j=1}^M$, given the context c and the user query x .

533 First, note that we can write the density of $\pi_{\hat{r},M}$ as follows:

$$\begin{aligned} \pi_{\hat{r},M}(y|x) &= \sum_{\{y_j\}_{j=1}^M \in Y^M} \pi_{\hat{r},M}(y|x, \{y_j\}_{j=1}^M) \pi_{\text{con}}(\{y_j\}_{j=1}^M|x; c) \\ &= \mathbb{E}_{\{y_j\}_{j=1}^M \sim \pi_{\text{con}}(\cdot|x; c)} [\pi_{\hat{r},M}(y|x, \{y_j\}_{j=1}^M)] \\ &= \mathbb{E}_{\{y_j\}_{j=1}^M \sim \pi_{\text{con}}(\cdot|x; c)} \left[\sum_j \mathbb{I}\{y_j = y\} \frac{\exp\left(\frac{\hat{r}(x, y_j)}{\tau} + \log \frac{\pi_{\text{ref}}(y_j|x)}{\pi_{\text{con}}(y_j|x; c)}\right)}{\sum_{\zeta \in \{y_j\}_{j=1}^M} \exp\left(\frac{\hat{r}(x, \zeta)}{\tau} + \log \frac{\pi_{\text{ref}}(\zeta|x)}{\pi_{\text{con}}(\zeta|x; c)}\right)} \right] \\ &= \mathbb{E}_{\{y_j\}_{j=1}^M \sim \pi_{\text{con}}(\cdot|x; c)} \left[\sum_j \mathbb{I}\{y_j = y\} \frac{\frac{\pi_{\text{ref}}(y_j|x)}{\pi_{\text{con}}(y_j|x; c)} \exp\left(\frac{\hat{r}(x, y_j)}{\tau}\right)}{\sum_{\zeta \in \{y_j\}_{j=1}^M} \frac{\pi_{\text{ref}}(\zeta|x)}{\pi_{\text{con}}(\zeta|x; c)} \exp\left(\frac{\hat{r}(x, \zeta)}{\tau}\right)} \right] \\ &= \mathbb{E}_{\{y_j\}_{j=1}^M \sim \pi_{\text{con}}(\cdot|x; c)} \left[\frac{\sum_j \mathbb{I}\{y_j = y\}}{\sum_{\zeta \in \{y_j\}_{j=1}^M} \frac{\pi_{\text{ref}}(\zeta|x)}{\pi_{\text{con}}(\zeta|x; c)} \exp\left(\frac{\hat{r}(x, \zeta)}{\tau}\right)} \right] \frac{\pi_{\text{ref}}(y|x)}{\pi_{\text{con}}(y|x; c)} \exp\left(\frac{\hat{r}(x, y)}{\tau}\right) \\ &= \mathbb{E}_{\{y_j\}_{j=1}^M \sim \pi_{\text{con}}(\cdot|x; c)} \left[\frac{\frac{1}{M} \sum_j \mathbb{I}\{y_j = y\}}{\frac{1}{M} \sum_{\zeta \in \{y_j\}_{j=1}^M} \frac{\pi_{\text{ref}}(\zeta|x)}{\pi_{\text{con}}(\zeta|x; c)} \exp\left(\frac{\hat{r}(x, \zeta)}{\tau}\right)} \right] \frac{\pi_{\text{ref}}(y|x)}{\pi_{\text{con}}(y|x; c)} \exp\left(\frac{\hat{r}(x, y)}{\tau}\right) \end{aligned}$$

534 Taking the limit as $M \rightarrow \infty$ and using the Law of Large Numbers (the sequences are i.i.d.):

$$\begin{aligned}
\lim_{M \rightarrow \infty} \pi_{\hat{r}, M}(y|x) &= \lim_{M \rightarrow \infty} \mathbb{E}_{\{y_j\}_{j=1}^M \sim \pi_{\text{con}}(\cdot|x; c)} \left[\frac{\pi_{\text{con}}(y|x; c)}{\mathbb{E}_{\zeta \sim \pi_{\text{con}}(\cdot|x)} \left[\exp\left(\frac{\hat{r}(x, \zeta)}{\tau}\right) \frac{\pi_{\text{ref}}(\zeta|x)}{\pi_{\text{con}}(\zeta|x; c)} \right]} \right] \frac{\pi_{\text{ref}}(y|x)}{\pi_{\text{con}}(y|x; c)} \exp\left(\frac{\hat{r}(x, y)}{\tau}\right) \\
&= \pi_{\text{con}}(y|x; c) \frac{1}{\mathbb{E}_{\zeta \sim \pi_{\text{con}}(\cdot|x)} \left[\exp\left(\frac{\hat{r}(x, \zeta)}{\tau}\right) \frac{\pi_{\text{ref}}(\zeta|x)}{\pi_{\text{con}}(\zeta|x; c)} \right]} \frac{\pi_{\text{ref}}(y|x)}{\pi_{\text{con}}(y|x; c)} \exp\left(\frac{\hat{r}(x, y)}{\tau}\right) \\
&= \pi_{\text{ref}}(y|x) \frac{1}{\sum_{\zeta \in Y} \pi_{\text{con}}(\zeta|x) \exp\left(\frac{\hat{r}(x, \zeta)}{\tau}\right) \frac{\pi_{\text{ref}}(\zeta|x)}{\pi_{\text{con}}(\zeta|x; c)}} \exp\left(\frac{\hat{r}(x, y)}{\tau}\right) \\
&= \pi_{\text{ref}}(y|x) \frac{1}{\sum_{\zeta \in Y} \pi_{\text{ref}}(\zeta|x) \exp\left(\frac{\hat{r}(x, \zeta)}{\tau}\right)} \exp\left(\frac{\hat{r}(x, y)}{\tau}\right) \\
&= \pi_{\text{ref}}(y|x) \frac{1}{\mathbb{E}_{\zeta \sim \pi_{\text{ref}}(\cdot|x)} [\exp(\hat{r}(x, \zeta)/\tau)]} \exp\left(\frac{\hat{r}(x, y)}{\tau}\right)
\end{aligned}$$

535 □

536 *Corollary 4.1 Proof.* The proof follows directly from Theorem B.1 and Appendix A.1 in Rafailov
537 et al. [2023]. □

538 B.3 Formal Connection To Importance Sampling

539 Given truthful reports by the advertisers, there is a strong connection between our allocation rule
540 and importance sampling. The optimal LLM for the platform's objective π_r^* can be interpreted as
541 the target distribution that our allocation rule is trying to simulate. The aim of our allocation rule
542 is to return a reply as similar as possible to a reply drawn from the optimal LLM π_r^* . This can be
543 interpreted as drawing a reply from π_r^* and then evaluating it with respect to the same function, i.e.,
544 our allocation rule is trying to estimate $\mathbb{E}_{y \sim \pi_r^*(\cdot|x)} [\pi_r^*(y|x)]$. The LLM π_{gen} that we use to generate
545 the candidate replies (either the reference LLM or the context-aware LLM) can be interpreted as the
546 proposal distribution $\pi_{\text{prop}}(\cdot|x)$ used to generate samples. We are interested in how well, with respect
547 to the target probability $\pi_r^*(\cdot|x)$, a sample y from our estimator using M candidate replies matches a
548 sample drawn from the target distribution $\pi_r^*(\cdot|x)$. In this section, we will analytically show how the
549 quality of the proposal distribution $\pi_{\text{prop}} = \pi_{\text{gen}}$ affects the quality of that estimator.

550 First, in Lemma B.2 we establish the variance of our estimator:

551 **Lemma B.2.** For any LLM π_{gen} such that π_{ref} is absolutely continuous with respect to π_{gen} , the
552 variance of the policy $\pi_{r, M}(\cdot|x)$ as an estimator for $\mathbb{E}_{y \sim \pi_r^*(\cdot|x)} [\pi_r^*(y|x)]$ is

$$\text{Var}(\hat{\mu}_{\text{IS}}) = \frac{1}{M} \left(\sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{gen}}(y|x)} - \left(\sum_{y \in Y} \pi_r^*(y|x)^2 \right)^2 \right). \quad (4)$$

553 *Proof.* Let y_1, y_2, \dots, y_M be the M generated candidate replies drawn from the proposal distribution
554 $\pi_{\text{prop}}(\cdot|x)$. Conditioned on those candidate replies, the importance sampling estimator for the expected
555 value of the function $\pi_r^*(y|x)$ under the target distribution $\pi_r^*(y|x)$ is

$$\begin{aligned}
\hat{\mu}_{\text{IS}} &= \frac{1}{M} \sum_{j=1}^M \pi_r^*(y_j|x) \cdot w(x_j) \\
&= \frac{1}{M} \sum_{j=1}^M \pi_r^*(y_j|x) \cdot \frac{\pi_r^*(y_j|x)}{\pi_{\text{prop}}(y_j|x)} \\
&= \frac{1}{M} \sum_{j=1}^M \frac{\pi_r^*(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)}
\end{aligned}$$

556 Taking expectation over the generated candidate replies y_1, \dots, y_M :

$$\begin{aligned}
\mathbb{E}[\hat{\mu}_{\text{IS}}] &= \mathbb{E}_{y_1, \dots, y_M \sim \pi_{\text{prop}}(\cdot|x)} \left[\frac{1}{M} \sum_{j=1}^M \frac{\pi_r^*(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)} \right] \\
&= \frac{1}{M} \sum_{j=1}^M \mathbb{E}_{y_j \sim \pi_{\text{prop}}(\cdot|x)} \left[\frac{\pi_r^*(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)} \right] \\
&= \mathbb{E}_{y \sim \pi_{\text{prop}}(\cdot|x)} \left[\frac{\pi_r^*(y|x)^2}{\pi_{\text{prop}}(y|x)} \right] \\
&= \sum_{y \in Y} \frac{\pi_r^*(y|x)^2}{\pi_{\text{prop}}(y|x)} \pi_{\text{prop}}(y|x) \\
&= \sum_{y \in Y} \pi_r^*(y|x)^2
\end{aligned} \tag{5}$$

557 Thus, our estimator is unbiased, as expected. The variance of the estimator $\hat{\mu}_{\text{IS}}$ is given by:

$$\text{Var}(\hat{\mu}_{\text{IS}}) = \mathbb{E}[\hat{\mu}_{\text{IS}}^2] - \mathbb{E}[\hat{\mu}_{\text{IS}}]^2 \tag{6}$$

558 For the term $\hat{\mu}_{\text{IS}}^2$ we have:

$$\begin{aligned}
\hat{\mu}_{\text{IS}}^2 &= \left(\frac{1}{M} \sum_{j=1}^M \frac{\pi_r^*(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)} \right) \\
&= \frac{1}{M^2} \sum_{j=1}^M \sum_{j'=1}^M \frac{\pi_r^*(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)} \cdot \frac{\pi_r^*(y_{j'}|x)^2}{\pi_{\text{prop}}(y_{j'}|x)}
\end{aligned}$$

559 Taking the expectation $\mathbb{E}_{\pi_{\text{prop}}(\cdot|x)}[\cdot]$:

$$\begin{aligned}
\mathbb{E}[\hat{\mu}_{\text{IS}}^2] &= \frac{1}{M^2} \sum_{j=1}^M \sum_{j'=1}^M \mathbb{E}_{\pi_{\text{prop}}(\cdot|x)} \left[\frac{\pi_r^*(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)} \cdot \frac{\pi_r^*(y_{j'}|x)^2}{\pi_{\text{prop}}(y_{j'}|x)} \right] \\
&= \frac{1}{M^2} \left(\sum_{j=1}^M \mathbb{E}_{\pi_{\text{prop}}(\cdot|x)} \left[\frac{\pi_r^*(y_j|x)^4}{\pi_{\text{prop}}(y_j|x)^2} \right] + \sum_{j \neq j'} \mathbb{E}_{\pi_{\text{prop}}(\cdot|x)} \left[\frac{\pi_r^*(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)} \right] \right) \\
&= \frac{1}{M^2} \left(M \sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{prop}}(y|x)} + M(M-1) \left(\sum_{y \in Y} \pi_r^*(y|x)^2 \right)^2 \right) \\
&= \frac{1}{M} \sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{prop}}(y|x)} + \frac{M-1}{M} \left(\sum_{y \in Y} \pi_r^*(y|x)^2 \right)^2
\end{aligned} \tag{7}$$

560 The final expression for the variance of the importance sampling estimator with M samples can be
561 computed by substituting Equations (5) and (7) in Equation (6):

$$\begin{aligned}
\text{Var}(\hat{\mu}_{\text{IS}}) &= \frac{1}{M} \sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{prop}}(y|x)} + \frac{M-1}{M} \left(\sum_{y \in Y} \pi_r^*(y|x)^2 \right)^2 - \left(\sum_{y \in Y} \pi_r^*(y|x)^2 \right)^2 \\
&= \frac{1}{M} \left(\sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{prop}}(y|x)} - \left(\sum_{y \in Y} \pi_r^*(y|x)^2 \right)^2 \right)
\end{aligned} \tag{8}$$

562

□

Lemma B.2 reveals two important details about our estimator. First, the variance of the estimator is inversely proportional to the number of candidate replies generated M . Thus, as M increases the variance decreases and our estimator becomes more stable in simulating the target distribution $\pi_r^*(\cdot|x)$. Second, the closer that the proposal distribution $\pi_{\text{prop}}(\cdot|x)$ is to the target distribution, the smaller the term $\sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{prop}}(y|x)}$ will be, which reduces the variance of our estimator.

Lemma B.3 establishes how a smaller variance leads to faster convergence:

Lemma B.3. *Let $\mu_{\text{true}} = \mathbb{E}_{y \sim \pi_r^*(\cdot|x)} [\pi_r^*(y|x)]$, and $\hat{\mu}_{\text{IS}}$ be the importance sampling estimator as in Lemma B.2 using $M = O\left(\frac{1}{\delta \varepsilon^2} \sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{prop}}(y|x)}\right)$ candidate replies. Then, for any LLM π_{gen} such that π_{ref} is absolutely continuous with respect to π_{gen} , we have that $|\hat{\mu}_{\text{IS}} - \mu_{\text{true}}| < \varepsilon$ with probability at least $1 - \delta$.*

Proof. Let $\hat{\mu}_{\text{IS}}$ be the importance-weighted estimator whose variance is given by Lemma B.2 as

$$\text{Var}(\hat{\mu}_{\text{IS}}) = \frac{1}{M} \left(\sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{prop}}(y|x)} - \left(\sum_{y \in Y} \pi_r^*(y|x)^2 \right)^2 \right).$$

We have already shown in Lemma B.2 that the estimator is unbiased. By Chebyshev's inequality, for any $\varepsilon > 0$,

$$\Pr[|\hat{\mu}_{\text{IS}} - \mu_{\text{true}}| \geq \varepsilon] \leq \frac{\text{Var}(\hat{\mu}_{\text{IS}})}{\varepsilon^2} = \frac{1}{M \cdot \varepsilon^2} \left(\sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{prop}}(y|x)} - \left(\sum_{y \in Y} \pi_r^*(y|x)^2 \right)^2 \right)$$

where $\mu_{\text{true}} = \mathbb{E}_{y \sim \pi_r^*(\cdot|x)} [\pi_r^*(y|x)] = \sum_{y \in Y} \pi_r^*(y|x)^2$.

Setting $\Pr[|\hat{\mu}_{\text{IS}} - \mu_{\text{true}}| \geq \varepsilon] < \delta$ and solving for M yields:

$$M = \frac{1}{\delta \varepsilon^2} \left(\sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{prop}}(y|x)} - \left(\sum_{y \in Y} \pi_r^*(y|x)^2 \right)^2 \right) = O\left(\frac{1}{\delta \varepsilon^2} \sum_{y \in Y} \frac{\pi_r^*(y|x)^4}{\pi_{\text{prop}}(y|x)}\right).$$

□

Based on Lemma B.3, our allocation rule converges to the quantity it is trying to approximate at a rate of \sqrt{M} .

B.4 Bounding the Total Variation Distance

In Appendix B.2 we proved that MOSAIC converges in the limit to the theoretically optimal distribution for the platform's objective. Then, in Appendix B.3 we showed how MOSAIC's allocation rule relates to importance sampling, and what that implies for the similarity of MOSAIC's output to that of the optimal distribution π_r^* . Now we will also show that MOSAIC's probability estimates for any possible sequence converge to that of π_r^* at a rate of \sqrt{M} .

Theorem B.4. *Let \mathcal{Y} be the domain consisting of all sequences in a finite token vocabulary T of length up to L . Hence $|\mathcal{Y}| < \infty$. Assume $0 \leq w(y) \leq C$ for all $y \in \mathcal{Y}$, where*

$$w(y) = \frac{\pi_r^*(y|x)}{\pi_{\text{prop}}(y|x)}.$$

Consider M i.i.d. draws $y_1, \dots, y_M \sim \pi_{\text{prop}}(\cdot|x)$ and define the self-normalized importance-sampled distribution induced by Algorithm 1:

$$\hat{\pi}_{r,M}(y|x) = \frac{\sum_{j=1}^M w(y_j) \mathbf{1}\{y_j = y\}}{\sum_{j=1}^M w(y_j)}.$$

591 Then, for any $\varepsilon > 0$ and $\delta > 0$, if $M = O\left(\frac{C^2|\mathcal{Y}|^2}{\varepsilon^2} \ln\left(\frac{|\mathcal{Y}|}{\delta}\right)\right)$, with probability at least $1 - \delta$
 592 (over the random draws), the Total Variation distance between $\hat{\pi}_M(\cdot \mid x)$ and $\pi_r^*(\cdot \mid x)$ satisfies
 593 $d_{\text{TV}}(\hat{\pi}_M, \pi_r^*) \leq \varepsilon$.

594 *Proof.* First, we will show that for any $M \in \mathbb{N}$, MOSAIC is an unbiased estimator of π_r^* .

595 For each fixed $y \in \mathcal{Y}$, define $X_j^{(y)} := w(y_j) \mathbf{1}\{y_j = y\}$ to be a random variable that equals $w(y_j)$
 596 if the j -th draw from the proposal distribution is y . Intuitively, this is a weighted indicator for
 597 whether the j -th draw is equal to y .

598 For any $y \in \mathcal{Y}$, the expected value of this indicator, with the expectation taken over the draw of
 599 $y_j \sim \pi_{\text{prop}}$

$$\mathbb{E}[X_j^{(y)}] = \sum_{z \in \mathcal{Y}} w(z) \mathbf{1}\{z = y\} \pi_{\text{prop}}(z) = w(y) \pi_{\text{prop}}(y) = \pi_r(y).$$

600 Thus, $\hat{Q}_M(\{y\}) = \frac{1}{M} \sum_{j=1}^M X_j^{(y)}$ is an unbiased estimator of $\pi_r^*(y \mid x)$.

601 Similarly, $\hat{Q}_M(\mathcal{Y}) = \frac{1}{M} \sum_{j=1}^M w(y_j)$ is an unbiased estimator of 1 (since $\mathbb{E}[w(y_j)] = 1$) (assuming
 602 that π_r^* is the optimal policy, i.e., a normalized LLM so that the probabilities sum up to one).

603 The self-normalized distribution can be written as

$$\hat{\pi}_{r,M}(y \mid x) = \frac{\hat{Q}_M(\{y\})}{\hat{Q}_M(\mathcal{Y})}.$$

604 *Hoeffding's inequality* for bounded random variables tells us that for any $\varepsilon' > 0$:

$$\Pr\left[|\hat{Q}_M(\{y\}) - \pi_r^*(y \mid x)| \geq \varepsilon'\right] \leq 2 \exp\left(-\frac{2M\varepsilon'^2}{C^2}\right).$$

605 For each $y \in \mathcal{Y}$, we have the same deviation event, so by a union bound over all $y \in \mathcal{Y}$:

$$\Pr\left[\exists y \in \mathcal{Y} : |\hat{Q}_M(\{y\}) - \pi_r^*(y \mid x)| \geq \varepsilon'\right] \leq \sum_{y \in \mathcal{Y}} 2 \exp\left(-\frac{2M\varepsilon'^2}{C^2}\right) = 2|\mathcal{Y}| \exp\left(-\frac{2M\varepsilon'^2}{C^2}\right) \leq \delta_1.$$

606 Thus, using $M = O\left(\frac{C^2}{\varepsilon'^2} \ln\left(\frac{|\mathcal{Y}|}{\delta_1}\right)\right)$ candidate replies, with probability at least $1 - \delta_1$ over the draw
 607 of those replies, we have *simultaneously* for all $y \in \mathcal{Y}$, $|\hat{Q}_M(\{y\}) - \pi_r^*(y \mid x)| \leq \varepsilon'$.

608 By a similar argument, applying Hoeffding's inequality to $\hat{Q}_M(\mathcal{Y})$, which is an unbiased estimator of
 609 1:

$$\Pr\left[|\hat{Q}_M(\mathcal{Y}) - 1| \geq \varepsilon'\right] \leq 2 \exp\left(-\frac{2M\varepsilon'^2}{C^2}\right).$$

610 Hence, with high probability $1 - \delta_2$, $\hat{Q}_M(\mathcal{Y})$ is also within $\pm\varepsilon'$ of 1. A union bound over both events
 611 (all $y \in \mathcal{Y}$ plus $\hat{Q}_M(\mathcal{Y})$) yields a final event E'_ε with probability $1 - (\delta_1 + \delta_2)$ on which:

$$\hat{Q}_M(\{y\}) \in [\pi_r^*(y) - \varepsilon', \pi_r^*(y) + \varepsilon'], \quad \hat{Q}_M(\mathcal{Y}) \in [1 - \varepsilon', 1 + \varepsilon'], \quad \forall y \in \mathcal{Y}.$$

612 On this good event E'_ε , for each $y \in \mathcal{Y}$:

$$\hat{\pi}_M(y|x) = \frac{\hat{Q}_M(\{y\})}{\hat{Q}_M(\mathcal{Y})} \in \left[\frac{\pi_r^*(y|x) - \varepsilon'}{1 + \varepsilon'}, \frac{\pi_r^*(y|x) + \varepsilon'}{1 - \varepsilon'}\right]$$

613 Recall the total variation (TV) distance between two discrete distributions P and Q on \mathcal{Y} is

$$d_{\text{TV}}(P, Q) = \frac{1}{2} \sum_{y \in \mathcal{Y}} |P(y) - Q(y)|.$$

614 We aim to show that on the event E'_ε (which holds with probability at least $1 - (\delta_1 + \delta_2)$), the quantity

$$\sum_{y \in \mathcal{Y}} |\hat{\pi}_M(y|x) - \pi_r^*(y|x)|$$

615 is at most $O(\varepsilon')$.

616 On $E_{\varepsilon'}$, we have, for each $y \in \mathcal{Y}$,

$$\hat{\pi}_M(y|x) = \frac{\hat{Q}_M(\{y\})}{\hat{Q}_M(\mathcal{Y})} \in \left[\frac{\pi_r^*(y|x) - \varepsilon'}{1 + \varepsilon'}, \frac{\pi_r^*(y|x) + \varepsilon'}{1 - \varepsilon'} \right].$$

617 Thus:

$$\hat{\pi}_M(y|x) - \pi_r^*(y|x) \in \left[-\frac{\varepsilon' \pi_r^*(y|x)}{1 + \varepsilon'} - \frac{\varepsilon'}{1 + \varepsilon'}, \frac{\varepsilon' \pi_r^*(y|x)}{1 - \varepsilon'} + \frac{\varepsilon'}{1 - \varepsilon'} \right] \quad (9)$$

$$\in \left[-\frac{2\varepsilon}{1 + \varepsilon}, \frac{2\varepsilon}{1 - \varepsilon} \right] \quad (10)$$

618 Thus, conditioned on the good event $E_{\varepsilon'}$, we have $|\hat{\pi}_M(y) - \pi_r^*(y)| = O(\varepsilon')$ by a Taylor expansion
619 for small ε' .

620 Therefore, on the event E'_ε ,

$$d_{\text{TV}}(\hat{\pi}_M, \pi_r^*) = \frac{1}{2} \sum_{y \in \mathcal{Y}} |\hat{\pi}_M(y) - \pi_r^*(y)| \leq \frac{1}{2} \sum_{y \in \mathcal{Y}} O(\varepsilon') = O(|\mathcal{Y}| \cdot \varepsilon').$$

621 Setting $\varepsilon' = \frac{\varepsilon}{|\mathcal{Y}|}$ to achieve the target error ε concludes the proof.

622

□

C Details from Section 5

In this section, we present all omitted details from Section 5.

C.1 Omitted Proofs from Section 5.1

Theorem 5.1 Proof. Let $\hat{\mathbf{r}}_i = (\hat{r}_i(x, y_1), \dots, \hat{r}_i(x, y_M))$ be the reward reports of advertiser i for the M generated candidate sequences. Then, for MOSAIC's allocation rule, holding the candidate sequences and the reports of all other advertisers fixed, the probability of returning each of the M candidate sequences as a function of i 's reports is:

$$\pi(\mathbf{r}_i; \vec{\beta}_{-i}) = \text{softmax} \left(\frac{\mathbf{r}_i}{\tau} + \beta_{-i} \right), \quad (11)$$

where $\beta_{-i,j} = \frac{\sum_{k \in N \setminus \{i\}} \hat{r}_k(x, y_j)}{\tau} + \log \frac{\pi_{\text{ref}}(y_j|x)}{\pi_{\text{con}}(y_j|x; c)}$. Importantly β_{-i} is an M -dimensional vector that does not depend on advertiser i 's reports.

We would like to equip $\pi(\cdot; \beta_{-i})$ with a payment rule $p(\cdot; \beta_{-i})$ so that the resulting mechanism $(\pi(\cdot; \beta_{-i}), p(\cdot; \beta_{-i}))$ will be strategyproof. This requires that $\pi(\cdot; \beta_{-i})$ have a property known as *cyclic monotonicity*. Equivalently, $\pi(\cdot; \beta_{-i})$ must be the (sub)gradient of advertiser i 's utility for bidding truthfully in the mechanism $U(\mathbf{r}_i; \beta_{-i})$, and that utility function must be convex [Frongillo and Kash, 2021, Rochet, 1987, Myerson, 1981].

It is easy to verify that for the function class:

$$U_C(\mathbf{r}_i; \beta_{-i}) = \tau \log \left(\sum_{j=1}^M \exp \left(\frac{r_i(x, y_j)}{\tau} + \beta_{-i,j} \right) \right) + C, \quad C \in \mathbb{R} \quad (12)$$

the allocation rule $\pi(\mathbf{r}_i; \beta_{-i})$ is a gradient of $U_C(\mathbf{r}_i; \beta_{-i})$. Additionally, $U_C(\mathbf{r}_i; \beta_{-i})$ is convex in \mathbf{r}_i : the exponential function e^x is (strictly) convex, because its second derivative is positive. The transformation $\frac{r_i(x, y_j)}{\tau} + \beta_{-i,j}$ is an affine transformation of $r_i(x, y_j)$, and affine transformations preserve convexity. Finally, it is well-known that the LogSumExp function is convex.

Thus, for any β_{-i} and for any set of generated candidate sequences, reporting truthfully maximizes advertiser i 's expected utility, with the expectation taken over the draw of the final sequence from the set of candidate sequences. Adopting the quasi-linear utility model, advertiser i 's payment is:

$$\begin{aligned} U_C(\mathbf{r}_i; \beta_{-i}) &= \pi(\mathbf{r}_i; \beta_{-i}) \cdot \mathbf{r}_i - p(\mathbf{r}_i; \beta_{-i}) \\ p(\mathbf{r}_i; \beta_{-i}) &= \pi(\mathbf{r}_i; \beta_{-i}) \cdot \mathbf{r}_i - U_C(\mathbf{r}_i; \beta_{-i}) \\ p(\mathbf{r}_i; \beta_{-i}) &= \pi(\mathbf{r}_i; \beta_{-i}) \cdot \mathbf{r}_i - \tau \log \left(\sum_{j=1}^M \exp \left(\frac{r_i(x, y_j)}{\tau} + \beta_{-i,j} \right) \right) - C, \quad C \in \mathbb{R} \end{aligned} \quad (13)$$

645

□

C.2 Differences from Standard Auction Settings

Standard auction environments typically rely on a set of assumptions that simplify mechanism design; however, these assumptions do not apply to auctions for LLM-generated content. In this section, we detail these assumptions and discuss why they are inapplicable in our context.

First, in a standard auction setting, it is common to assume that the agents' valuation functions satisfy free disposal, i.e., $v_i(S) \geq v_i(S') \forall S \supseteq S', S, S' \subseteq \mathcal{I}$. The interpretation of free disposal is that an agent can discard any items she is allocated that she is not interested in. Free disposal combined with the fact that an agent has zero value for the empty bundle mean that her value for any outcome is weakly positive. Second, in most auction environments, the allocation rule is different for different agents: each agent will get allocated her own bundle of items, and we can assume that she is indifferent to the allocation of items to the other agents.

As detailed in Rafailov et al. [2023], assuming that an agent's LLM π_i was trained to maximize her reward function (and regularized with respect to its KL divergence from some reference LLM, which we assume to be the same as the auctioneer's reference LLM), there is a one-to-many mapping between an advertiser's optimal LLM, and her implicit reward function. That mapping is:

$$r_i(x, y) = \tau_i \log \frac{\pi_i(y|x)}{\pi_{\text{ref}}(y|x)} + \log Z_i(x) \quad (14)$$

where $Z_i(x)$ is a prompt-dependent constant, and τ_i is the regularization hyperparameter of advertiser i , similar to the one in Equation (1). All functions in the class defined in Equation (14) are equivalent, in the sense that they induce exactly the same LLM [Rafailov et al., 2023]. This has two implications: First, unlike standard auction environments, an agent's reward can go negative – there is nothing equivalent to the free disposal property. Setting $Z_i(x)$ to zero (which is equivalent to normalizing the induced probabilities by the LLM [Rafailov et al., 2023]), the agent's reward is negative for any sequence for which her LLM assigns a lower probability than π_{ref} .

Second, especially in the online advertising application, an agent's expected utility for not participating in the auction is negative: if advertiser i does not participate in the auction, her payment is zero, but her expected value for the outcome is

$$\pi(\mathbf{0}; \beta_{-i}) \cdot \mathbf{r}_i = \pi(\beta_{-i}) \cdot \mathbf{r}_i \quad (15)$$

The other advertisers have very low rewards for the sequences that mention advertiser i : assuming their LLMs have been properly trained, they will evaluate all sequences that explicitly mention a different, possibly competing brand, as unlikely. Thus, based on Equation (14) the corresponding advertisers have very low rewards for those sequences and conversely, advertiser i has low rewards for the sequences that the other advertisers have high rewards for. But based on Equation (2), if advertiser i does not participate in the auction, $\pi(\mathbf{0}; \beta_{-i})$ will assign high probabilities to sequences for which i has low rewards for. Thus, Equation (15) implies that, unlike standard auction environments, the advertiser's expected reward and utility for not participating in the mechanism is negative.

C.3 Our mechanism is “almost individually rational”

First, we explain why the standard notion of individual rationality (i.e., weakly positive utility from participation in the mechanism) encountered in most auction settings is impossible to achieve in this domain while converging to the optimal distribution and maintaining incentive compatibility. Then, we explain how, with our payment offset, our mechanism is “almost IR:” In Lemma C.1 we prove that the ex-interim utility of an advertiser who has zero reward for all candidate sequences and bids truthfully is deterministically zero, i.e., advertisers that do not contribute to the social welfare (but also do not detract from it) have zero utility. Similarly, in Lemma C.2 we prove that if an agent’s reward for all candidate sequences is (weakly) positive, then her ex-interim utility is (weakly) positive.

Why is individual rationality (IR) impossible? *Individual rationality (IR)* stipulates that an agent gains more utility by participating and bidding truthfully in a mechanism than by not participating at all. Typically, if an agent’s utility for non-participation is zero, participating should yield weakly positive utility. However, this simplification does not apply in our setting.

As discussed in Section 5.2, advertiser i ’s reward for any sequence y can be arbitrarily negative (Equation (14)). The same is true for the utility from truthful participation, as outlined in Equation (12). To ensure a positive utility for every advertiser in our mechanism, an offset would need to be infinitely large or dependent on advertiser i ’s reports. But then the mechanism’s allocation rule would no longer be the gradient of advertiser i ’s utility with respect to her reports, which would destroy strategyproofness [Frongillo and Kash, 2021, Rochet, 1987, Myerson, 1981].

It is important to note that this challenge is inherent not just to our mechanism but to any mechanism in this setting that operates with a fixed set of sequences, aims to approximate the optimal distribution, and maintains strategyproofness. Under these conditions, the only allocation rule that approximates the theoretically optimal distribution (Equation (2)) is that of our mechanism. However, this uniquely determines the advertisers’ utilities, up to a constant factor, as described in Equation (12) [Frongillo and Kash, 2021, Rochet, 1987, Myerson, 1981].

Lemma C.1. *For the payment offset $C = -\tau \log \left(\sum_{j=1}^M \exp(\beta_{-i,j}) \right)$ if advertiser i ’s reward for all candidate sequences is zero, then her ex-interim utility is deterministically zero, for all $\beta_{-i} \in \hat{R}_{-i}$.*

Lemma C.1 Proof. First, note that for all $\beta_{-i} \in \hat{R}_{-i}$, advertiser i ’s expected reward for the outcome is zero, as $\pi(r_i; \beta_{-i}) \cdot r_i = \pi(r_i; \beta_{-i}) \cdot \mathbf{0} = 0$. Additionally, advertiser i ’s reward for the realized outcome will deterministically be zero, as her reward for all generated candidate sequences is zero. Finally, note that by setting $r_i = \mathbf{0}$ in Equation (13) with the offset C set as in Section 5.3, we have that the advertiser i ’s payment is also deterministically zero. Thus, an advertiser with zero reward for all generated candidate sequences who reports her rewards truthfully has deterministically zero reward for the final outcome and zero payments, and her utility is also deterministically zero.

□

Lemma C.2. *For the payment offset $C = -\tau \log \left(\sum_{j=1}^M \exp(\beta_{-i,j}) \right)$ if advertiser i ’s reward for all candidate sequences is positive, then her ex-interim utility is positive, for all reports $\beta_{-i} \in \hat{R}_{-i}$.*

Proof. Lemma C.1 establishes that when advertiser i ’s reward for all candidate sequences is zero, her utility for truthfully bidding in the mechanism, denoted as $U(\mathbf{0}; \beta_{-i})$, is zero for all possible reports of the other advertisers $\beta_{-i} \in \hat{R}_{-i}$.

Furthermore, Theorem 5.1 shows that the mechanism’s allocation rule corresponds to the gradient of advertiser i ’s utility when bidding truthfully. Because the allocation rule is non-negative, the gradient of advertiser i ’s utility for bidding truthfully is also non-negative.

Thus, if advertiser i ’s rewards for all candidate sequences are weakly positive, and considering the non-negative gradient of her utility, her ex-interim utility under truthful bidding must be positive, irrespective of the other advertisers’ reports β_{-i} .

□

726 **Corollary C.3.** For the payment offset $C = -\tau \log \left(\sum_{j=1}^M \exp(\beta_{-i,j}) \right)$ if the distribution π_{con} only
 727 generates candidate sequences for which advertiser i 's reward is positive, then the ex-ante expected
 728 utility of the advertiser is positive.

729 *Corollary C.3 Proof.* This follows immediately from the fact the the fact that the ex-ante utility of
 730 the advertiser is the expectation of her ex-interim utility with respect to her reward for the generated
 731 sequences, and the fact that the second quantity is positive whenever the reward of the advertiser for
 732 all candidate sequences is positive from Lemma C.2.

733 □

734 C.4 “What you give is what you get”

735 As we explained in Section 5.3.2, our allocation rule, which is the only one over a finite set of replies
 736 that converges to the optimal LLM, is also the (sub)gradient of the utility to ensure truthfulness
 737 (Rochet, 1987). Because the allocation rule is the same for all advertisers, their utilities must also be
 738 the same, up to advertiser-specific offsets, as indicated by Equation (16):

$$U_C(\mathbf{r}_i; \beta_{-i}) = \tau \log \left(\sum_{j=1}^M \exp \left(\frac{1}{\tau} \sum_{k \in N} r_k(x, y_j) \right) + \log \frac{\pi_{ref}(y_j|x)}{\pi_{con}(y_j|x; c)} \right) + C, \quad C \in \mathbb{R} \quad (16)$$

739 However, not all advertisers contribute equally to the social welfare of the final outcome. Because of
 740 this, implementing the mechanism without a carefully-designed offset would lead to free-riding: as
 741 long as an agent's utility in Equation (16) is positive, she would be incentivized to participate, even
 742 if the user query was completely unrelated to her business, because the mechanism would ensure
 743 that she received, on expectation, the same (positive) expected utility from doing so as any other
 744 participating advertiser.⁹

745 Incentivizing unrelated advertisers to participate would have adverse effects. First, the better-
 746 performing context-aware mechanism would create candidate sequences with worse rewards for *all*
 747 advertisers, because its context would be “diluted” from advertisers unrelated to the user query. In
 748 our running example for the query “How to learn music online?”, imagine adding “Try to mention
 749 ‘EasySwitch’, a comprehensive VPN service” to the context of the context-aware LLM $\pi_{con}(\cdot|x; c)$.

750 Additionally, for both versions of the mechanism, following the discussion in Section 5.2, the
 751 advertisers for whom the user query is unrelated are more likely to have negative rewards for the
 752 generated sequences as their LLMs will deem the candidate sequences more unlikely than the
 753 reference LLM. Thus, based on their utility according to Theorem 5.1, their participation in the
 754 mechanism will lead to a reduction of the total sum of rewards of the generated sequences for the
 755 advertisers, which will indirectly reduce the expected utility of all advertisers, making the mechanism
 756 less attractive for the user-query-relevant advertisers.

757 To summarize, all advertisers receiving the same utility would incentivize advertisers for whom the
 758 user query is unrelated to participate in the auction. This would in turn reduce everyone's expected
 759 utility, potentially reducing the incentive for the user-query-relevant advertisers to participate, and
 760 lead to sequences with worse expected rewards for the advertisers and usefulness for the user.¹⁰ Thus,
 761 in the application of auctions for aggregating advertisers' preferences over LLM-generated outputs,
 762 advertisers with higher contribution to social welfare also receiving proportionally higher utility by
 763 the mechanism is important for the long-term success of the mechanism in practice.

⁹We can assume that advertisers can estimate their expected utility from participation using historical data from past auctions, analogously to how they can estimate their utility for participating in sponsored search auctions.

¹⁰If we interpret the KL divergence between the distribution induced by the reference LLM and the LLM that generated the candidate sequences as a measure of their expected usefulness for the user.

D Details from Section 6

D.1 Detailed Experiment Setup

All synthetic instances are provided in Appendix E. We use Llama-2-7b-chat-hf as the reference LLM [Touvron et al., 2023], which uses the Llama Community License. The context-aware LLM is created as described in Footnote 4.

Following Rafailov et al. [2023], the advertisers’ reward functions are defined as $r_i(x, y) = \log \frac{\pi_i(y|x)}{\pi_{\text{ref}}(y|x)}$. For the auctioneer’s objective, we set $\tau = 1$ in Equation (1), balancing advertisers’ rewards and divergence from the reference LLM.

We use 50 user queries and test each query on 25 different random seeds, resulting in 1,250 instances. Following Li et al. [2024], Rozière et al. [2024], we sample from all LLMs using a temperature of 0.8 and top-p 0.95.

We create a set of synthetic instances to test our mechanism. Each instance consists of a user query, e.g. “How do I bake cookies?” and a list of advertisers. Each advertiser is defined by an “advertiser name”, e.g. “KitchenFix” and an advertiser description, e.g., “producing kitchen appliances.”¹¹ The reference LLM π_{ref} responsible for generating replies that are useful for the user is Llama-2-7b-chat-hf [Touvron et al., 2023]. In Appendix D.10, we replicate these experiments using Google’s flan-t5-large model [Chung et al., 2022], observing qualitatively very similar results. Following Dütting et al. [2024], we create the advertisers’ LLMs by adding advertising instructions to the reference LLM. The advertisers’ LLMs are created using the same reference LLM, and adding the instruction: “Answer the question advertising $\langle \text{advertiser} \rangle$, $\langle \text{advertiser description} \rangle$.” The context aware LLM is created using the same reference LLM, and adding the instruction: “Answer the query. Try to mention $\langle \text{advertiser 1} \rangle$, who $\langle \text{advertiser description 1} \rangle$ and $\langle \text{advertiser 2} \rangle$, who $\langle \text{advertiser description 2} \rangle$.”

Following [Rafailov et al., 2023] the reward function of advertiser i is set to $r_i(x, y) = \log \frac{\pi_i(y|x)}{\pi_{\text{ref}}(y|x)}$, where π_i is advertiser i ’s LLM, i.e., we set $\tau_i = 1$, $Z_i(x) = 1$ for all advertisers and for all user prompts in Equation (14).¹² For the auctioneer’s objective as defined in Equation (1) we set $\tau = 1$, balancing between the advertisers’ expected rewards for the generated sequences and the sequences’ divergence from the reference LLM responsible for generating useful replies for the user. Thus, the optimal policy according to Equation (2) becomes:

$$\pi_r^*(y|x) = \frac{1}{Z(x)} \pi_{\text{ref}}(y|x) \exp \left(\sum_{i \in N} r_i(x, y) \right) \quad (17)$$

Following Li et al. [2024], Rozière et al. [2024] we sample from the LLM generating the sequences (either π_{ref} or π_{con}) with temperature 0.8 and top-p 0.95. We use 50 user queries, each with two interested advertisers. To increase the statistical significance of our results, we test each user query on 25 different random seeds, effectively resulting in 1250 individual instances. The seed affects both the generation of the candidate sequences, and the randomized sampling of the final sequence from the candidate sequences. We will make all of our code and set of synthetic instances publicly available.

All experiments were conducted on a compute cluster running Ubuntu 20.04.6 LTS with AMD EPYC processors with 48 cores and 1512GB RAM and Nvidia A100 GPUs and Python 3.12.1.

D.2 Choice of baselines

There are no other mechanisms in the literature for aggregating multiple LLMs to compare against. The only other mechanism that aggregates multiple LLMs is Dütting et al. [2024], which is not strategy-proof and does not converge to the platform’s objective. The only other potential comparison is VCG, which, as explained in Section 3.2, is not applicable since it does not align with the platform’s objective. Hence, the best baseline for comparison is naive MOSAIC, which we consider

¹¹Following Dütting et al. [2024], we use fictitious brand names to avoid the model using any existing ad.

¹²Note that we cannot estimate the true value of $Z_i(x)$. However, it is a query-dependent constant. Given that our experiments focus on the reward difference of different sequences for the same query, this is a non-issue.

807 a strong benchmark given that it provably converges to the platform’s optimal distribution based on
808 Corollary 4.1.

809 **D.3 Comparison of baseline and context-aware mechanism outputs**

810 To provide some intuition of how the context-aware mechanism improves the relevance of generated
811 responses for the advertisers, we present a side-by-side comparison of outputs from both mechanisms
812 for a specific query in Table 2. Note that only the context-aware mechanism is able to incorporate the
813 advertisers into the reply to the user’s query.

User query
What are effective ways to learn a musical instrument online?
Advertisers
- MusicMastery: offering online music lessons and virtual tutoring - InstaTune: selling musical instruments and learning aids
Context-Aware Mechanism Output
Effective ways to learn a musical instrument online include: 1. Online music lessons: Platforms like MusicMastery offer one-on-one virtual tutoring, group classes, and pre-recorded lessons. These lessons can be tailored to your skill level and learning style. 2. Interactive learning tools: InstaTune offers a range of interactive learning tools, such as apps, software, and games, that make learning fun and engaging. 3. Practice with technology: Utilize digital tools like metronomes, tuners, and recording software to practice and track your progress. 4. Collaborate with other musicians: Join online communities, forums, and social media groups to connect with other musicians and learn from their experiences. . . .
Baseline Mechanism Output
- Practice consistently and regularly: Consistent practice is essential to learning any musical instrument, and it’s even more important when learning online. Set aside a specific time each day or week to practice, and make sure to stick to it. - Find a good teacher or mentor: While it’s possible to learn some things on your own, working with a good teacher or mentor can help you improve faster and avoid developing bad habits. Look for someone who is experienced and knowledgeable in the type of music you want to learn. - Use technology to your advantage: There are many online tools and resources available that can help you learn an instrument. These include video lessons, online tutorials, and apps that can help you practice and track your progress. . . .

Table 2: Comparison of context-aware and baseline mechanism outputs using 10 candidate sequences. Advertisers highlighted in bold, outputs truncated to conserve space.

814 **D.4 Experimental Evaluation of the Usefulness to the User**

815 In this section, we experimentally evaluate the usefulness of the replies returned by MOSAIC from
816 the user’s perspective.

817 In Figure 3, we plot the log probability of the replies returned by MOSAIC with respect to the optimal
818 LLM for the user (i.e., the reference LLM) as a function of the number of candidate replies generated.
819 We compare the context-aware version of MOSAIC to a baseline version. To benchmark our results,
820 we also compare against sampling a *single reply* directly from the optimal distribution for this task,
821 which is the reference LLM itself.

822 The results in Figure 3 show that for the context-aware version of our mechanism, the log probability
823 of the replies relative to the reference LLM increases rapidly as the number of candidate replies
824 grows. This demonstrates that by generating more replies, the context-aware version of MOSAIC is
825 able to produce replies with substantial value for the user. By comparison, the usefulness to the user
826 that the baseline version is able to achieve does not scale with more generated replies.

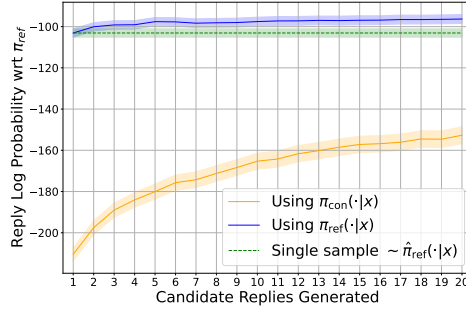
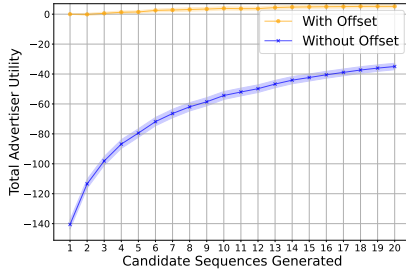


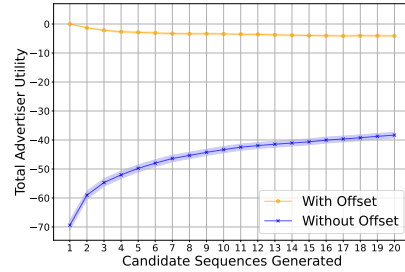
Figure 3: Reply log probability with respect to the reference LLM as a function of the number of replies generated using π_{ref} and π_{con} .

However, there remains a gap between the log probability of those replies and the benchmark set by the reference LLM. This discrepancy arises because the platform optimizes a different objective: the expected advertiser rewards combined with the KL divergence from the reference LLM. By increasing the weight τ that the platform places on the reference LLM, this gap between the context-aware version of MOSAIC and the user’s optimal benchmark will close.

D.5 Comprehensive Experimental Evaluation of the Offset from Section 5.3



(a) Sequences generated using π_{con}



(b) Sequences generated using π_{ref}

Figure 4: Comparison of total advertiser utility gain from participation with, and without the payment offset, as a function of the number of candidate sequences generated using π_{ref} and π_{con} . Averaged over 1250 runs including 95% CIs.

In this section, we provide a comprehensive experimental evaluation of the payment offset introduced in Section 5.3. Figure 4 explores the effectiveness of the payment offset introduced in Section 5.3 in ensuring that our mechanism is ex-ante IR, i.e., advertisers have positive expected utility gains from participating. To that end, we compare the advertiser utility gain from participation with and without the payment offset, as a function of the number of candidate sequences generated by the context-aware (Figure 4a) and baseline (Figure 4b) versions of our mechanism.

In Figure 4a we observe that for the context-aware version of our mechanism, adding the payment offset introduced of Section 5.3 to the payment rule is enough to ensure positive expected utility for the advertisers (conditioned on the fact that they are related to the user’s query), i.e., make the mechanism ex-ante IR. Notably, without the payment offset, advertisers have very negative expected utility from participating in the mechanism.

In Figure 4b we observe that for the baseline version of our mechanism, adding the payment offset introduced of Section 5.3 to the payment rule causes a very large increase in the advertisers’ expected utility, but it is still not enough to ensure positive expected utility for the advertisers. To conclude, Figure 4 demonstrates that the offset introduced in Section 5.3 causes a large increase in

the advertisers' expected utility, which in case of the context-aware version of the mechanism, is also enough to make the mechanism ex-ante IR.

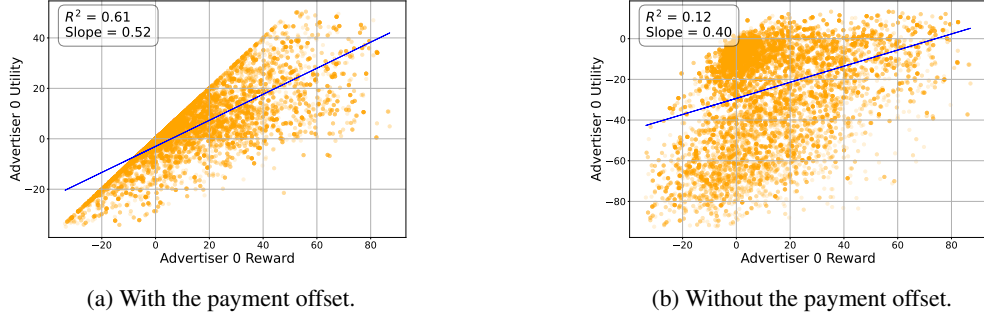


Figure 5: Comparative scatter plots of advertiser reward and utility gain from participation, with and without the payment offset of Section 5.3 for candidate sequences generated by the context-aware LLM π_{gem} . We additionally show a linear regressor fit to that data, its slope and its R^2 .

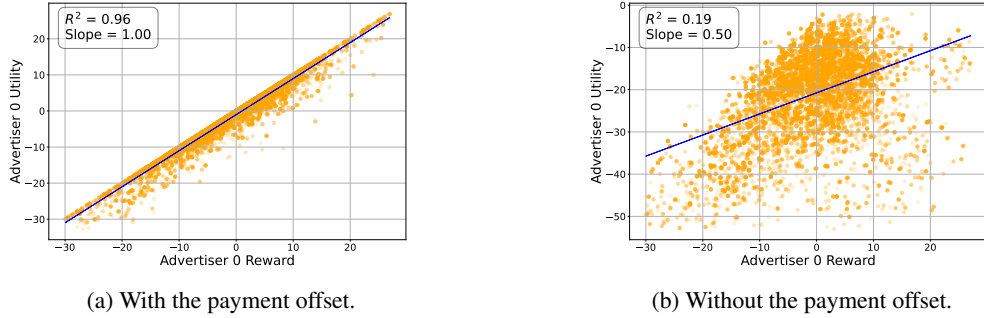


Figure 6: Comparative scatter plots of advertiser reward and utility gain from participation, with and without the payment offset of Section 5.3 for candidate sequences generated by the reference π_{ref} . We additionally show a linear regressor fit to that data, its slope and its R^2 .

Figures 5 and 6 explore the effectiveness of the payment offset introduced in Section 5.3 in aligning an advertiser's utility with her contribution to the social welfare. In Figure 5 we compare the scatter plots of the advertiser reward and utility gain from participation in the mechanism, with and without the payment offset introduced in Section 5.3 for candidate sequences generated using the context-aware LLM π_{ref} . Additionally, for both subfigures, we show a linear regressor fitted to the data, as well as its slope and coefficient of determination. Comparing the two subfigures, it is immediately obvious that adding the offset to the payments makes the relationship between advertiser reward and utility gain far more linear. This is confirmed by the coefficient of determination of the linear regressors fit to each dataset. The coefficient of determination of the linear regressor is far larger when we use the offset. Without the payment offset, the coefficient of determination is almost 0, indicating that, without our payment offset, reward gain is not a predictive measure of an agent's utility. Additionally, the slope of the linear regressor is also higher for the scatter plot with the payment offset.

In Figure 6 we make the same comparison, but for candidate sequences generated using the reference LLM π_{ref} . The results are now even more pronounced. In Figure 6a we observe the relationship between advertiser utility and reward gain with our payment offset is almost perfectly linear, as suggested by the linear regressor fitted to the data having a slope of 1.00 and an extremely high coefficient of determination of 0.96, indicating that it can almost perfectly fit the data. Without our payment offset however, in Figure 6b we can see that the relationship between the two metrics is again both less linear, and less positively correlated, as the slope of the linear regressor is 0.5 and its coefficient of determination is only 0.19.

To conclude, in all cases tested, the use of the advertiser-specific offset introduced in Section 5.3 increases an advertiser's expected utility, makes the relationship between an advertiser's contribution

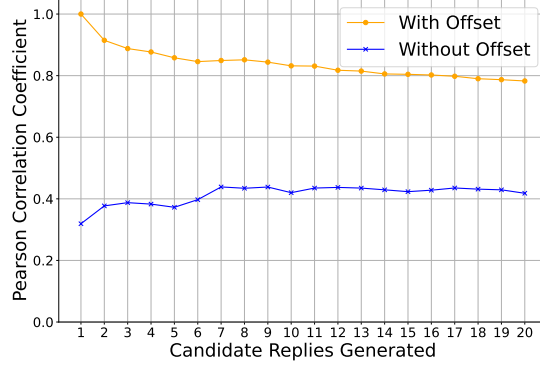


Figure 7: Pearson correlation between advertiser normalized utility and reward.

to social welfare and her utility both more linear and more positively correlated, and, for the context-aware version of our mechanism, can ensure positive expected utility for the advertisers.

D.6 Experiments in Settings with Many Advertisers

In this section, we test MOSAIC’s performance in settings with $n = 5$ and $n = 10$ advertisers, compared to the results of Section 6 where the number of advertisers was set to 2. To maintain the same balance between advertisers and the usefulness to the user as in the experiments of the main paper, we set $\tau = n/2$.¹³ We use the same 50 user queries, but now each query has five and ten advertisers, and test each query on 20 different random seeds, resulting in 1000 instances.

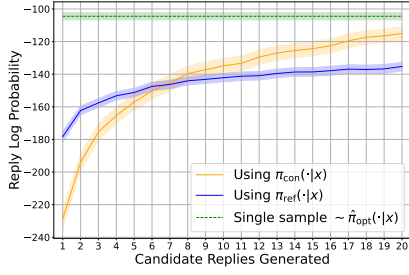
In Figures 8a to 8j, we present the log probability of the returned reply with respect to the optimal LLM for the platform’s objective π_r^* and the reference LLM π_{ref} , the total advertiser value and reward gain from participating in MOSAIC and the platform’s revenue. All plots are with respect to the number of candidate replies (i.e., LLM queries) that MOSAIC used. We compare MOSAIC’s baseline and context-aware versions.

First, in Figures 8a and 8b we observe that the context-aware version of MOSAIC quickly converges to the platform’s objective. By comparison, MOSAIC’s naive version fails to do so. It is noteworthy that for very low numbers of candidate replies, the naive version of the mechanism performs better than the context-aware version. That is because in these experiments, more weight is placed towards the reference LLM, as we have set a higher τ value. However, the superior convergence rate of the context-aware version of the mechanism recovers that difference very quickly. At the same time, Figures 8c and 8d show that the context-aware version of the mechanism produces sequences that closely match the distribution of the reference LLM, thus maintaining the usefulness to the user.

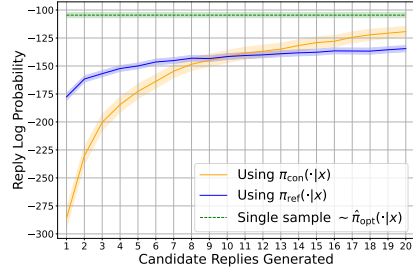
Finally, the context-aware version of MOSAIC is able to generate significant value (Figures 8e and 8f) and utility (Figures 8g and 8h) for the advertisers, while also recapturing a significant portion as revenue for the platform (Figures 8i and 8j).

Taken together, similar to our results in the main paper, we have shown that even with a large number of advertisers, MOSAIC quickly converges to the theoretically optimal distribution, generating significant value and utility for the advertisers and revenue for the platform, while also maintaining its usefulness to the user.

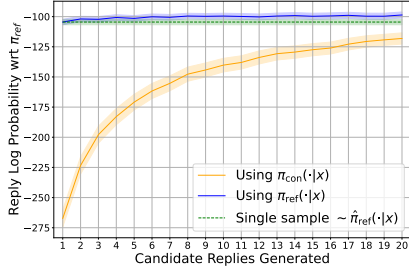
¹³Note that in the main experiments of Section 6 we had $n = 2$ and $\tau = 1$



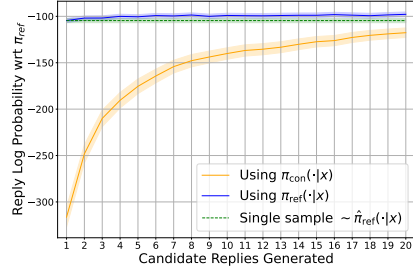
(a) Log probability w.r.t. π_r^* for 5 advertisers.



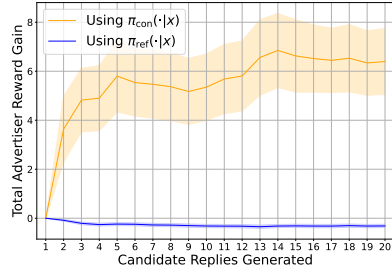
(b) Log probability w.r.t. π_r^* for 10 advertisers.



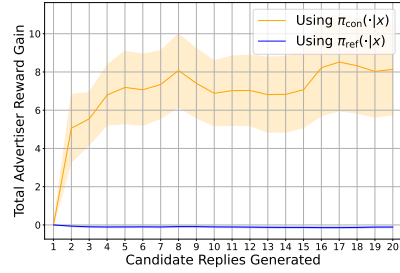
(c) Log probability w.r.t. π_{ref} for 5 advertisers.



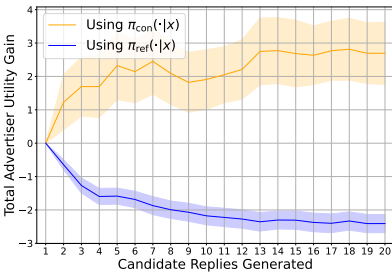
(d) Log probability w.r.t. π_{ref} for 10 advertisers.



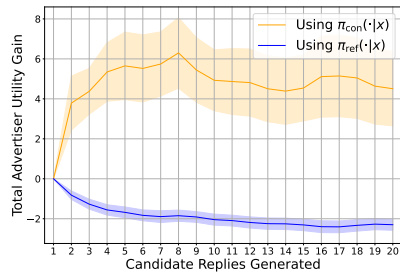
(e) Advertiser value gain for 5 advertisers.



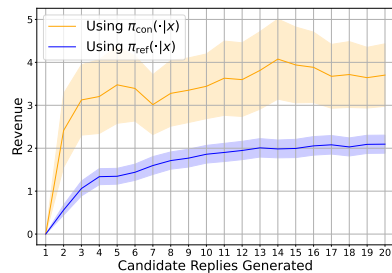
(f) Advertiser value gain for 10 advertisers.



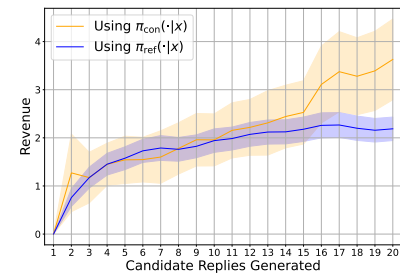
(g) Advertiser utility gain for 5 advertisers.



(h) Advertiser utility gain for 10 advertisers.



(i) Revenue for 5 advertisers.



(j) Revenue for 10 advertisers.

Figure 8: Comparison of results for 5 advertisers (left) and 10 advertisers (right). Each row corresponds to a specific metric: log probability with respect to the optimal LLM π_r^* , the reference LLM π_{ref} , advertiser value and utility gain, and revenue. Shown are averages over 1000 instances including 95% CIs.

900 D.7 Comprehensive Experimental Evaluation of Context

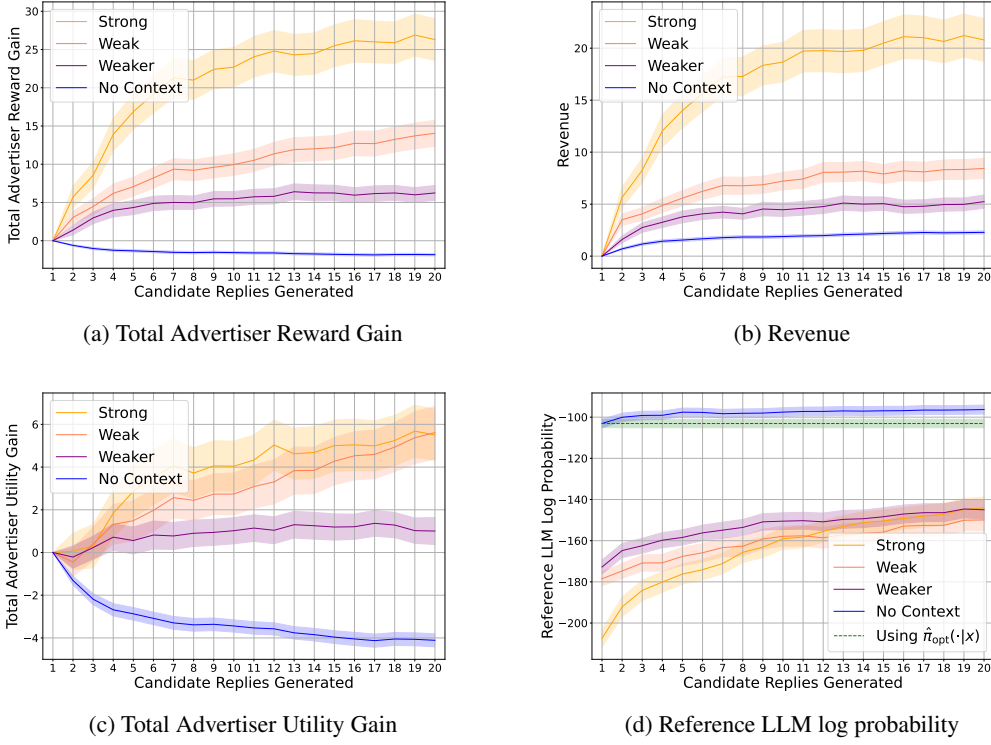


Figure 9: Total advertiser normalized reward, revenue, total advertiser utility gain and reference LLM log probability as functions of the number of candidate replies generated for various levels of context. Averages over 700 runs with 95% CIs.

901 In this section, we experimentally evaluate MOSAIC’s performance under varying levels of *context c*.
 902 Specifically, we test four levels of context:

- 903 • *Strong*: the context-aware LLM $\pi_{\text{con}}(\cdot | x; c)$ has access to both the advertisers’ names and
 904 their descriptions;
- 905 • *Weak*: $\pi_{\text{con}}(\cdot | x; c)$ has access only to the advertisers’ names;
- 906 • *Weaker*: $\pi_{\text{con}}(\cdot | x; c)$ is given neither names nor descriptions, and is simply instructed to
 907 “try to promote related brands”;
- 908 • *No context*: corresponds to omitting context entirely.

909 Using “strong” context corresponds to the context-aware version of MOSAIC evaluated in Section 6,
 910 while “no context” corresponds to the baseline version.

911 For each level of context, we replicate the experiments from Section 6 using the same set of test
 912 instances. To reduce computational costs, we evaluate only 700 instances for the *strong*, *weak*, and
 913 *weaker* context levels. For the *no context* setting, we reuse the results from the 1250 instances
 914 reported in Section 6.

915 In Figure 9, we plot, for each context level, the total advertiser normalized reward and utility gain from
 916 participation, log probability of the reply under the optimal LLM (Equation (2)), and the platform’s
 917 revenue, all as functions of the number of candidate replies generated.

918 In Figure 9a, we observe that for all context levels except “no context,” the total advertiser reward
 919 increases rapidly with the number of generated replies. Comparing these results with the total
 920 advertiser utility gain in Figure 9c reveals an interesting insight: even under the weak context setting,
 921 where the context-aware LLM is provided only with advertiser brand names, MOSAIC generates
 922 substantial utility for advertisers, which also scales quickly with the number of candidate replies.
 923 Notably, the utility gains under strong and weak contexts using twenty candidate replies are nearly

identical. As the level of context increases beyond the weak setting, the additional social welfare generated by MOSAIC is largely recaptured as revenue for the platform, as shown in Figure 9b.

Finally, in Figure 9d, we observe that as the level of context increases, the reference LLM log probability is initially lower for small numbers of candidate replies but quickly matches across all context levels (except no context) as the number of candidates increases. This shows that MOSAIC’s allocation rule (Algorithm 1) is powerful enough that the gains in advertiser value and platform revenue enabled by richer context do not come at the expense of user utility.

To conclude, MOSAIC delivers substantial advertiser utility and platform revenue even with minimal contextual information. As the level of context increases, both advertiser value and platform revenue grow significantly, without compromising the usefulness of responses to the user.

D.8 Compute Experiments

In this section, we experimentally evaluate MOSAIC’s computational requirements. Using the setup described in Section 6.1, we measure the total wall time required by MOSAIC on a *single* A100 GPU, focusing on the number of candidate replies generated for evaluation. This total time includes:

1. Generating candidate replies using the context-aware LLM.
2. Evaluating these replies with the advertisers’ LLMs implemented as discussed in Section 6.1.
3. Calculating the Rochet payments, as detailed in Section 5. In practice the time to calculate payments is negligible (under 10 milliseconds), and need not increase user-perceived latency because they can be calculated after showing outputs to the user.

Figure 10 presents the total time taken by MOSAIC, which includes both the generation and evaluation of candidate replies. These components are further broken down in Figures 11 and 12, respectively.

As previously demonstrated in Section 6 and Appendix D.6, MOSAIC converges to the optimal distribution using only 20 candidate replies, regardless of the number of advertisers. In Figure 11, we observe that generating 10 and 20 candidate replies on a single A100 GPU takes 30.5 and 60.5 seconds, respectively. This represents a 2.5- and 5-fold increase over the 12.0 seconds required to generate a single reply from the same LLM, i.e., the latency perceived by the user when directly querying the LLM.

To reduce user-perceived latency, MOSAIC can parallelize candidate generation across multiple GPUs. For example, distributing the generation of 20 replies across two GPUs (each generating 10 replies) reduces the total time to 30.5 seconds, assuming efficient evaluation techniques as outlined in Section 4.2. This approach results in MOSAIC requiring approximately 5 times the compute time of generating a single reply to converge, but with a perceived latency to the user of only 2.5 times that of directly querying an LLM (of the same architecture and similar size). Although our experiments were run on a single GPU, note that MOSAIC can always be parallelized further so that the user-perceived latency is the same as directly querying a *single* LLM, as explained in Section 4.2.

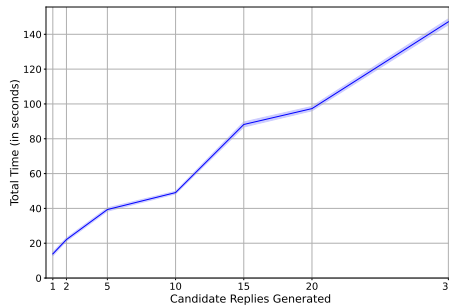


Figure 10: MOSAIC total wall time (seconds) as a function of the number of candidate replies generated using π_{con} . Shown are averages over 50 instances including 95% CIs.

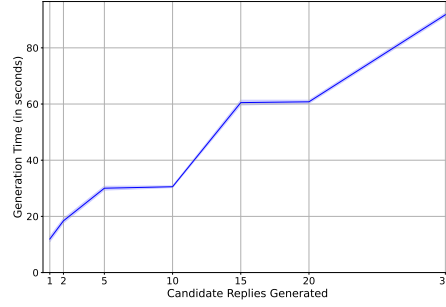


Figure 11: MOSAIC generation time as a function of the number of candidate replies generated using π_{con} . Shown are averages over 50 instances including 95% CIs.

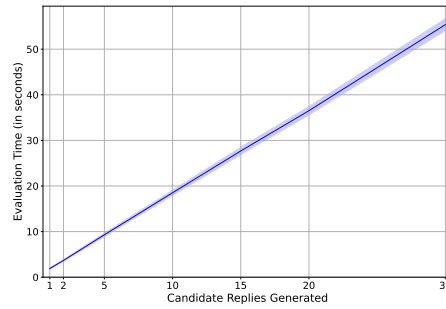


Figure 12: MOSAIC evaluation time as a function of the number of candidate replies. Shown are averages over 50 instances including 95 % CIs.

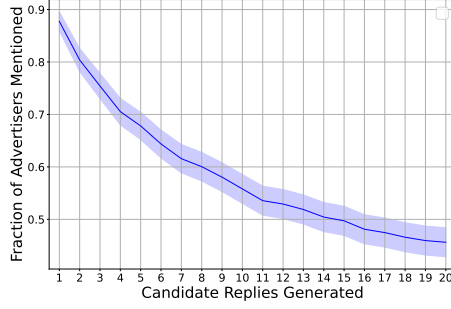


Figure 13: Fraction of advertisers mentioned in the output reply as a function of the number of candidate replies. Shown are averages over 1250 runs including 95% CIs.

959 D.9 Conflicts between advertisers

960 The output of MOSAIC is a single text response to the user. In trying to reflect the interests of
 961 multiple advertisers, there is a risk of a single response being incoherent, just mentioning as many
 962 advertisers as possible without being a good promotion for any of them. However, the platform’s
 963 objective, which takes into account both advertiser utility and usefulness for the user, should account
 964 for this: if advertisers are happy to be cross-promoted, those outcomes should be more likely, but
 965 incoherent advertisements or those that simultaneously promote competitors should be less likely.

966 In Figure 13 we show the fraction of advertisers mentioned by the context-aware version of MOSAIC
 967 as a function of the candidate number of replies used. We observe that initially, the fraction of
 968 advertisers mentioned is very high, but drops significantly once the mechanism has converged.
 969 The reason is precisely the fact that mentioning more advertisers in this case would cause either a
 970 significant drop in advertiser utility, or a drop in the usefulness of the reply to the user. By placing
 971 more weight towards the advertisers (i.e., decreasing the value of τ in Equation (1)), the platform can
 972 increase the number of advertisers shown when the mechanism has converged.

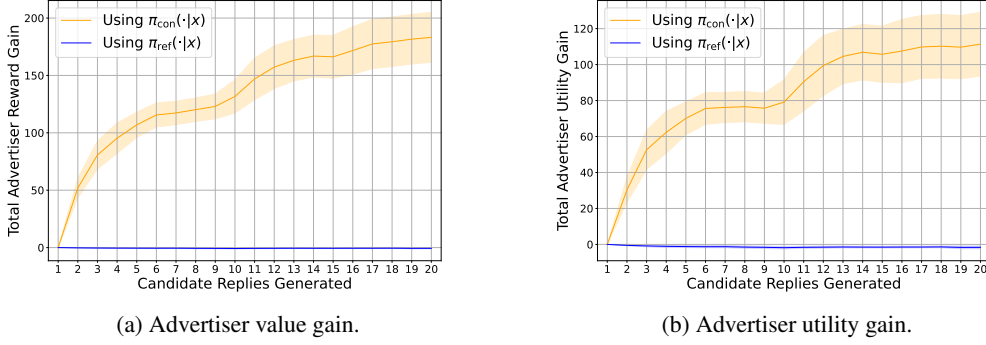


Figure 14: Advertiser gains as a function of the number of candidate replies generated using π_{ref} and π_{con} for the flan-t5-large model. Shown are averages over 1250 instances including 95% CIs.

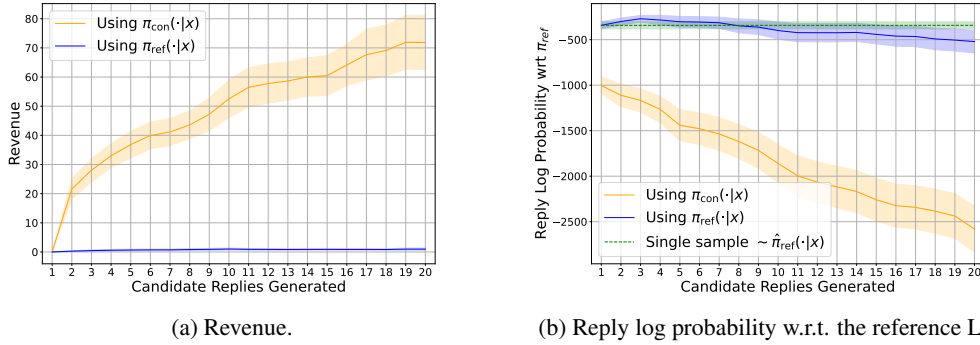


Figure 15: Revenue and reply log probability as a function of the number of candidate replies generated using π_{ref} and π_{con} for the flan-t5-large model. Shown are averages over 1250 instances including 95% CIs.

973 D.10 Testing with additional language models

974 In this section, we test MOSAIC’s performance using Google’s flan-t5-large model [Chung et al.,
975 2022], which uses the Apache 2.0 License.

976 We repeat exactly the same procedure as detailed in Section 6.1, but using flan-t5-large as the basis
977 for both π_{gen} , the advertisers’ LLMs, and the reference LLM. We use the same 50 user queries, each
978 with two advertisers, and test each query on 10 different random seeds, resulting in 500 instances.

979 In Figure 14a to Figure 15b, we present the total advertiser value gain, reward gain from participating
980 in MOSAIC, platform revenue, and the usefulness of the returned reply to users as functions of the
981 number of candidate replies generated.

982 Consistent with the main paper’s results using Llama-2-7b-chat-hf [Touvron et al., 2023], the context-
983 aware version of MOSAIC demonstrates significant value and utility generation for advertisers and
984 successfully captures a substantial portion as platform revenue. In contrast, though, the naive version
985 of MOSAIC fails to generate value or utility for advertisers, resulting in negligible revenue for the
986 platform. Similar to the results of the main paper, despite both mechanisms converging in the limit to
987 the platform’s objective, as shown in Corollary 4.1, their performance for a computationally tractable
988 number of generated replies diverges.

989 Similar to the findings in the main paper, the naive version retains the reference LLM’s usefulness to
990 the user (which is optimal for the task), while the context-aware version prioritizes the platform’s
991 objective, leading to lower user usefulness. Overall, our results using flan-t5-large agree closely
992 with those obtained using Llama-2-7b-chat-hf, demonstrating consistency across models. The main
993 difference is that the context-aware mechanism is even more important when using flan-t5-large.

E Expanded text of prompts

Table 3: Table of user prompts with corresponding advertisers and their descriptions. Only the first two advertisers are shown for each query; the full prompts are available in the code supplement.

<p>User Query 1 Learning to play the guitar.</p> <p>Advertisers</p> <ul style="list-style-type: none"> - StrumMaster: selling guitars, tuners, and learning materials for beginners - Chordify: offering online guitar lessons and interactive practice tools
<p>User Query 2 Best skin care routine for dry skin.</p> <p>Advertisers</p> <ul style="list-style-type: none"> - HydraSkin: providing moisturizers and serums for dry skin - GlowEssentials: selling natural and organic skin care products
<p>User Query 3 Improving productivity while working from home.</p> <p>Advertisers</p> <ul style="list-style-type: none"> - SpaceMakers: offering ergonomic furniture and home office setups - FocusApp: providing productivity apps and time management tools
<p>User Query 4 Starting a small online business.</p> <p>Advertisers</p> <ul style="list-style-type: none"> - EcomLaunch: offering e-commerce platform solutions and web design services - MarketMover: providing digital marketing services and SEO optimization
<p>User Query 5 Healthy meal planning on a budget.</p> <p>Advertisers</p> <ul style="list-style-type: none"> - BudgetBites: selling affordable meal kits and recipe books - NutriSaver: offering discounts on healthy groceries and food delivery services
<p>User Query 6 Mastering digital photography.</p> <p>Advertisers</p> <ul style="list-style-type: none"> - PixelPro: selling cameras, lenses, and photography accessories - EditCraft: offering photo editing software and online tutorials
<p>User Query 7 Effective ways to reduce household energy use.</p> <p>Advertisers</p> <ul style="list-style-type: none"> - EcoSave: offering energy-efficient home appliances and lighting solutions - InsulaTech: providing home insulation and energy audit services
<p>User Query 8 Finding the perfect hiking trails.</p> <p>Advertisers</p> <ul style="list-style-type: none"> - TrailFinder: offering a mobile app with detailed maps and trail reviews - GearUp: selling outdoor gear and apparel for hiking enthusiasts
<p>User Query 9 Building a personal brand on social media.</p> <p>Advertisers</p> <ul style="list-style-type: none"> - BrandBuilder: offering personal branding courses and social media strategy consultations - VisualizeMe: providing graphic design services for social media content
<p>User Query 10 Learning a new language effectively.</p> <p>Advertisers</p> <ul style="list-style-type: none"> - LingoLeap: offering online language learning courses and tutoring

- SpeakEasy: providing language learning apps with speech recognition technology

User Query 11

Staying fit without a gym.

Advertisers

- HomeFit: selling home workout equipment and fitness accessories
 - MoveIt: offering online fitness classes and personal training sessions
-

User Query 12

Eco-friendly travel options.

Advertisers

- GreenPath: offering eco-friendly travel packages and sustainable tourism experiences
 - EcoStay: providing listings for green hotels and accommodations
-

User Query 13

Mastering the art of cooking steak.

Advertisers

- GrillMaster: selling premium grills and barbecue accessories
 - SteakPerfection: offering online cooking classes focused on meat preparation
-

User Query 14

Creating a successful YouTube channel.

Advertisers

- VidGrowth: offering video production courses and YouTube growth strategies
 - ChannelDesign: providing custom YouTube channel art and video thumbnails
-

User Query 15

Decorating your home on a budget.

Advertisers

- DecorDeals: selling affordable home decor and furniture
 - StyleSavvy: offering interior design consultations and budget-friendly decorating tips
-

User Query 16

Managing stress and anxiety.

Advertisers

- CalmSpace: offering mindfulness apps and stress reduction tools
 - WellnessWave: providing online therapy sessions and mental health resources
-

User Query 17

How can I make my home more energy efficient?

Advertisers

- EcoWise: offering smart home energy solutions and solar panels
 - GreenInsulate: providing eco-friendly insulation materials and installation services
-

User Query 18

What are the best strategies for investing in the stock market?

Advertisers

- InvestSmart: offering investment advice and financial planning services
 - StockGenius: providing an AI-powered stock market analysis tool
-

User Query 19

How can I improve my mental health?

Advertisers

- MindBalance: offering mental health apps and mindfulness courses
 - SerenityNow: providing online counseling and therapy sessions
-

User Query 20

Where can I find organic and sustainable clothing?

Advertisers

- EcoWear: selling sustainable and organic clothing options
 - GreenThreads: offering eco-friendly fabrics and fair-trade fashion
-

User Query 21

How do I start a vegetable garden in my backyard?

Advertisers

- GardenGuru: selling gardening tools, seeds, and starter kits
 - PlantPerfect: offering online gardening courses and personalized garden planning
-

User Query 22

What are the best tools for remote team collaboration?

Advertisers

- TeamSync: offering collaboration software and project management tools
 - ConnectSuite: providing video conferencing solutions and team communication platforms
-

User Query 23

How can I learn to cook healthy meals quickly?

Advertisers

- QuickChef: offering quick and healthy cooking classes online
 - NutriFast: selling kitchen gadgets and appliances for healthy cooking
-

User Query 24

What are the best ways to save for retirement?

Advertisers

- FutureFund: offering retirement planning and investment services
 - PensionPlus: providing personalized financial advice for retirement savings
-

User Query 25

How can I decorate my apartment on a minimalist budget?

Advertisers

- MinimalSpace: selling minimalist furniture and decor
 - SimplicityDesigns: offering interior design consultation for minimalist living spaces
-

User Query 26

What are the best practices for data security in small businesses?

Advertisers

- SecureIT: offering cybersecurity solutions for small businesses
 - DataGuard: providing data protection services and training
-

User Query 27

How can I reduce my carbon footprint with daily habits?

Advertisers

- EcoHabits: offering products and tips for sustainable living
 - GreenStep: providing carbon offset subscriptions and eco-friendly lifestyle goods
-

User Query 28

What are effective ways to learn a musical instrument online?

Advertisers

- MusicMastery: offering online music lessons and virtual tutoring
 - InstaTune: selling musical instruments and learning aids
-

User Query 29

How can I keep my pets healthy and happy?

Advertisers

- PetWell: offering pet health insurance and wellness products
 - HappyPaws: selling organic pet food and eco-friendly toys
-

User Query 30

What's the best way to start learning coding from scratch?

Advertisers

- CodeBegin: offering coding bootcamps and online tutorials for beginners
 - DevPath: providing interactive coding courses and learning platforms
-

User Query 31

How to find affordable eco-friendly travel destinations?

Advertisers

- GreenTraveler: offering eco-friendly travel guides and affordable green vacation packages
 - EcoTripAdvisor: providing reviews and bookings for sustainable travel options
-

User Query 32

What are the best ways to improve indoor air quality?

Advertisers

- AirPure: selling air purifiers and natural air freshening solutions
 - BreatheEasy: offering indoor plant subscriptions and ventilation solutions
-

User Query 33

How to make my own natural beauty products?

Advertisers

- NatureGlow: offering DIY beauty product kits and organic ingredients
 - EcoBeauty: selling guides and supplies for creating homemade natural cosmetics
-

User Query 34

What are the best techniques for effective online teaching?

Advertisers

- EduTech: offering online teaching tools and educational technology solutions
 - TeachMaster: providing courses on online pedagogy and digital classroom management
-

User Query 35

How to create a successful podcast from home?

Advertisers

- PodCastify: offering podcast creation courses and audio equipment
 - SoundWave: selling podcast editing software and hosting services
-

User Query 36

What are the most effective ways to market a small business locally?

Advertisers

- LocalBoost: offering local SEO services and marketing strategies for small businesses
 - MarketTown: providing targeted advertising solutions and local business networking opportunities
-

User Query 37

How can I optimize my home office for productivity?

Advertisers

- WorkZen: offering ergonomic office furniture and productivity tools
 - SpaceCraft: providing home office design services and organization solutions
-

User Query 38

What are the best platforms for freelance writers to find work?

Advertisers

- WriteFreely: offering a job board and resources for freelance writers
 - ContentConnect: providing a platform to connect writers with businesses
-

User Query 39

How to get started with meditation and mindfulness?

Advertisers

- MindfulStart: offering guided meditation apps and mindfulness training
 - ZenSpace: selling meditation cushions and mindfulness journals
-

User Query 40

Where can I find unique handmade gifts?

Advertisers

- CraftCorner: selling handmade gifts and artisan crafts
 - MakerMarket: offering a marketplace for unique handmade items from local artists
-

User Query 41

How can I improve my credit score quickly?

Advertisers

- CreditBoost: offering credit monitoring services and financial advice
 - ScoreSurge: providing tools and tips for improving your credit score
-

User Query 42

What are the best apps for tracking fitness and nutrition?

Advertisers

- FitTrack: offering comprehensive fitness and nutrition tracking apps
 - NutriSync: selling personalized nutrition plans and diet tracking tools
-

User Query 43

How to plan an environmentally friendly wedding?

Advertisers

- GreenBride: offering eco-friendly wedding planning services and supplies
 - EcoWed: providing sustainable wedding attire and decor
-

User Query 44

What are the safest ways to travel during a pandemic?

Advertisers

- SafeJourney: offering travel safety kits and pandemic travel advice
 - HealthTravel: providing information on safe destinations and travel insurance
-

User Query 45

How can I learn DIY home repairs and improvements?

Advertisers

- FixItYourself: offering online courses and tutorials for home repair
 - DIYHome: selling DIY home improvement tools and kits
-

User Query 46

What are the best resources for starting a plant-based diet?

Advertisers

- PlantEats: offering guides and meal plans for starting a plant-based diet
 - VeggieVibe: selling plant-based cookbooks and kitchen gadgets
-

User Query 47

Tips for running a marathon.

Advertisers

- RunFastGear: specializing in high-performance running shoes
 - HydraFuel: offering electrolyte drinks and energy bars for athletes
-

User Query 48

Best practices for organic gardening.

Advertisers

- GreenThumb Solutions: providing organic fertilizers and pest control products
 - EcoGrow: selling heirloom seeds and sustainable gardening tools
-

User Query 49

How to improve home WiFi?

Advertisers

- SignalBoost: which offers advanced WiFi routers and extenders
 - NetWizard: providing network optimization services and support
-

User Query 50

Ways to save on travel.

Advertisers

- BudgetJourneys: specializing in affordable travel packages and deals
 - StayLocal: offering discounts on boutique hotels and unique accommodations
-

NeurIPS Paper Checklist

1. Claims

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

Answer: [\[Yes\]](#)

Justification: Our abstract and introduction make theoretical claims, which we prove, and describe the experiments we ran.

Guidelines:

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

2. Limitations

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

Answer: [\[Yes\]](#)

Justification: We clearly state the (quite weak) assumptions required for our key mechanism design properties to hold in Section 4 and Section 5. We describe the scope of the experiments (multiple language models, ablation on aspects of the mechanism, etc.) in Section 6.

Guidelines:

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

3. Theory assumptions and proofs

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

Answer: [Yes]

Justification: All proofs (e.g., of strategyproofness, convergence rates, etc. in Sections 4 and 5) have their assumptions clearly stated and discussed.

Guidelines:

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

4. Experimental result reproducibility

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

Answer: [Yes]

Justification: All relevant prompts, the list of fictitious advertisers, and hyperparameters required for running the experiments are included.

Guidelines:

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

5. Open access to data and code

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

Answer: [Yes]

Justification: Our generated data and code are included in the supplemental material.

Guidelines:

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

6. Experimental setting/details

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

Answer: [Yes]

Justification: As in question 4, all relevant hyperparameters are included. We also clearly describe how we estimated performance metrics to compare performance under different conditions.

Guidelines:

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

7. Experiment statistical significance

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

Answer: [Yes]

Justification: We report 95% CIs (over experiment runs) around our key performance metrics (revenue, advertiser welfare gain, etc.) in Section 6 and in all relevant appendices.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

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

8. Experiments compute resources

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

Answer: [Yes]

Justification: We describe the compute nodes used in Appendix D.

Guidelines:

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

9. Code of ethics

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

Answer: [Yes]

Justification: There are no human subjects or sensitive data. The primary contribution of our paper is mechanism design theory as well as proof of concept experiments. Negative social impacts of our work are unlikely. See also checklist item 10.

Guidelines:

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

10. Broader impacts

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

Answer: [Yes]

Justification: Our paper is primarily fundamental research, consisting of theoretical results and proof of concept experiments, so there is no direct societal impact. Our experiments (Section 6) do consider shifts in welfare between advertisers, the user, and the platform.

Guidelines:

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

11. Safeguards

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

Answer: [NA]

Justification: We don't train any new models nor does any of our released code have a high risk of misuse.

Guidelines:

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

12. Licenses for existing assets

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

Answer: [Yes]

Justification: For the open-source LLMs we use in our experiments, creators are credited, the terms of use are respected, and the licenses named (Appendix D).

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.

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

13. New assets

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

Answer: [NA]

Justification: We don't release any new assets other than the code and synthetic data required to reproduce the experiments.

Guidelines:

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

14. Crowdsourcing and research with human subjects

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

Answer: [NA]

Justification: We don't have any human subjects experiments.

Guidelines:

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

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

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

Answer: [NA]

Justification: We don't have any human subjects experiments.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- 1309 • Depending on the country in which research is conducted, IRB approval (or equivalent)
1310 may be required for any human subjects research. If you obtained IRB approval, you
1311 should clearly state this in the paper.
- 1312 • We recognize that the procedures for this may vary significantly between institutions
1313 and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the
1314 guidelines for their institution.
- 1315 • For initial submissions, do not include any information that would break anonymity (if
1316 applicable), such as the institution conducting the review.

1317 16. **Declaration of LLM usage**

1318 Question: Does the paper describe the usage of LLMs if it is an important, original, or
1319 non-standard component of the core methods in this research? Note that if the LLM is used
1320 only for writing, editing, or formatting purposes and does not impact the core methodology,
1321 scientific rigorousness, or originality of the research, declaration is not required.

1322 Answer: [Yes]

1323 Justification: The entire paper, especially the experiments section, describes our usage of
1324 LLMs, and we mention them even in the title.

1325 Guidelines:

- 1326 • The answer NA means that the core method development in this research does not
1327 involve LLMs as any important, original, or non-standard components.
- 1328 • Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>)
1329 for what should or should not be described.