

TRUTHFUL AGGREGATION OF LLMs WITH AN APPLICATION TO ONLINE ADVERTISING

Anonymous authors

Paper under double-blind review

ABSTRACT

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

1 INTRODUCTION

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

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

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

1.1 PROBLEM SETTING

Consider a situation where a user queries an LLM with a specific question. We assume the following:

- There is a *reference LLM* that can produce useful replies to the user’s query.
- There are *advertisers*, each with their own LLM, who want to influence the reply to the user.

In our model, each advertiser is represented via her own LLM or directly with a *reward function*, similar to the function used to fine-tune LLMs in the *Reinforcement Learning from Human Feedback* pipeline (Azar et al., 2023). For this reason, we refer to an advertiser’s value for a reply as her *reward*. The auctioneer’s task is, given the user’s query and the advertisers’ preferences, to *produce a reply that is useful for the user while also generating high rewards for the advertisers*.

Example. A user queries an LLM “*What are effective ways to learn a musical instrument online?*” There are two advertisers: *MusicMastery*, which offers online music lessons, and *InstaTune*, which sells musical instruments. MOSAIC merges the preferred replies from the advertisers’ LLMs and the reference LLM to produce one combined reply to the user:

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.*
2. *Interactive learning tools: **InstaTune** offers a range of interactive learning tools, such as apps, software, and games, that make learning fun and engaging.*

1.2 OVERVIEW OF CONTRIBUTIONS

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

Second, leveraging an importance sampling-based technique, MOSAIC is the only mechanism in the literature that converges to the optimal distribution as computational resources increase (Corollary 4.1). Third, MOSAIC can incorporate contextual information, similar to how sponsored search auctions utilize advertiser descriptions. This accelerates convergence and increases value for the advertisers and revenue for the auctioneer (Section 6.2.1).

Fourth, we address technical feasibility and practicality. We adopt the desiderata proposed by Duetting et al. (2023), who argued that auctions must be compatible with existing LLM technology, only using “information obtainable from current models” in such a way that they are “easy to integrate into the system” and rely only on “easy manipulations of LLM output.” Furthermore, it must be computationally feasible to run the auction repeatedly and with different advertiser preferences. In particular, we cannot afford the expensive process of fine-tuning an LLM’s weights for each possible query. To address this, we have designed MOSAIC to work by only post-processing multiple LLM outputs – it requires only “API access”, not even viewing the model weights (Section 4.2).

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

Sixth, we introduce an offset to the Rochet payments, to ensure that each advertiser’s utility is proportional to her contribution to the social welfare. This alignment is important for the long-term success of the market, incentivizing only the most relevant advertisers to participate.

Finally, we discuss *individual rationality (IR)*, which guarantees that a participant is weakly better off from participating in the mechanism. In Section 5.2, we discuss the unique properties of our setting that make *ex-post* IR unattainable. However, in Section 5.3.1, we show that MOSAIC is “almost *ex-post* IR” in a well-defined sense. In Section 6.2.1, we experimentally show that, for the distribution of instances tested, it is *ex-ante* IR (i.e., advertisers are in expectation better off by participating).

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

2 RELATED WORK

Duetting et al. (2023) introduced the field of mechanism design for LLMs. They proposed a sequential mechanism, where the output sequence is generated on a token-by-token basis and the advertisers bid for their LLM to generate the next token. However, their approach suffers from several limitations: (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 does not have any of these limitations.

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.

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.

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.

An LLM can be abstracted as a mapping from (partial) token sequences to a distribution over the next token, or equivalently as an implicit distribution over token sequences. We use this second abstraction, i.e., $\pi_i(y|x)$ denotes the probability that advertiser i ’s LLM π_i assigns to output sequence (i.e., reply) y for the user query x .

We let $r_i(x, y)$ denote advertiser i ’s *reward* for sequence y , given query x .² Informally, the auctioneer’s goal is to sample the final sequence from a distribution that optimizes the advertisers’ expected rewards without substantially diverging from π_{ref} . Formally, the goal is to maximize:

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

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.

¹As an example, suppose advertiser A has successfully bid for all the tokens in the sequence "Planning your next vacation? For the cheapest flights to Hawaii, book via..." However, she fails to win the bid for the crucial next token, which would have been her brand name. This leaves her having paid for the preceding tokens without getting the benefit of associating her brand with the message.

²In theory, the advertiser’s reward for a generated sequence could also depend on user-specific information, but we abstract that away. Equivalently, we assume that the reward function r_i provided by advertiser i is specific to the given user that asked the query x .

This objective is similar to the standard Reinforcement Learning from Human Feedback (RLHF) approach (Ziegler et al., 2020), but replaces the human feedback reward function $r_{\text{HF}}(x, y)$ with the sum of the advertisers’ rewards. For an overview of RLHF, we recommend Rafailov et al. (2023, §3).

The optimal solution to the optimization problem in (1) was derived by Peters & 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.

Let \vec{R} be the set of all possible reports by the advertisers. A *mechanism* is defined as a pair (π, p) . The *allocation rule* $\pi : \vec{R} \rightarrow \Delta(T^*)$ maps any report profile $\vec{r} = (r_1, r_2, \dots, r_n) \in \vec{R}$ of the advertisers’ rewards to a distribution over sequences $\delta(T^*)$. We denote the advertisers’ *aggregate reward* as $r(x, y) = \sum_{i=1}^n r_i(x, y)$, and their reward profile as $\vec{r}(x, y) = (r_1(x, y), r_2(x, y), \dots, r_n(x, y))$. The *payment rule* $p : \vec{R} \rightarrow \mathbb{R}^n$ maps any report profile of the advertisers’ rewards to a payment profile \vec{p} , where \vec{p}_i is the payment of the i -th advertiser to the mechanism.

A mechanism is *strategyproof* if and only if, for any advertiser $i \in N$, truthfully reporting her rewards is always the optimal strategy, regardless of the reports of other advertisers. In our setting, this holds in expectation because the reply is sampled from the optimal LLM for the platform’s objective, based on the advertisers’ reports. Formally:

Definition 3.1 (Strategyproof). *A mechanism (π, p) is strategyproof if, for all advertisers $i \in N$, for all true reward functions $\vec{r}_i \in \vec{R}_i$, for all reports $\vec{r}_{-i} \in \vec{R}_{-i}$ by the other advertisers, and for all possible misreports $\vec{r}_i' \in \vec{R}_i$ by advertiser i : $\mathbb{E}_{y \sim \pi(\vec{r})} [u_i(y, \vec{r}_i, \vec{r}_{-i}; r_i, x)] \geq \mathbb{E}_{y' \sim \pi(\vec{r}_i', \vec{r}_{-i}, x)} [u_i(y', \vec{r}_i', \vec{r}_{-i}; \vec{r}_i, x)]$, where $u_i(y, \vec{r}_i, \vec{r}_{-i}; r_i, x) = r_i(x, y) - p_i(\vec{r})$ is the utility of advertiser i for reply y when her payment is $p_i(\vec{r})$.*

3.2 THE SHORTCOMINGS OF VCG IN THIS SETTING

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

However, in our setting, VCG is not a viable option: The VCG allocation rule requires calculating the *exact* optimal solution to the optimization problem, which is intractable for choosing an LLM to maximize Equation (1) and is even difficult for choosing a single optimal sequence. If a sub-optimal solution is chosen, VCG’s strategyproofness is no longer guaranteed (Nisan & Ronen, 2007; 1999; Lehmann et al., 2002). Thus, VCG is not a suitable mechanism in our setting.

4 THE MOSAIC MECHANISM: ALLOCATION RULE

4.1 CONVERGENCE TO OPTIMALITY, ADVERTISER CONTEXTS AND IMPORTANCE SAMPLING

In this section, we introduce MOSAIC’s allocation rule. The high-level idea is that first, a set of M *candidate replies* is generated based on some LLM π_{gen} responsible for that task. Then, the probability of returning each candidate reply is re-weighted based on the advertisers’ reports and the reference LLM π_{ref} so that in the limit, as $M \rightarrow \infty$, the probability of returning each reply converges to its probability under the optimal distribution of Equation (2). This approach shares similarities with the rejection sampling approach that has been used at various points in the LLM training pipeline (e.g., Bai et al. (2022)). We defer all proofs to Appendix A.1.

Corollary 4.1. *The limiting policy $\pi_{\theta, M}(\cdot|x)$ induced by Algorithm 1 is the KL regularized policy that optimizes the aggregate reward function:*

$$\lim_{M \rightarrow \infty} \pi_{\theta, M}(\cdot|x) = \arg \max_{\pi} \mathbb{E}_{y \sim \pi(\cdot|x)} [r(x, y)] - \tau D_{\text{KL}}(\pi || \pi_{\text{ref}}) \quad (3)$$

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 \vec{r} , 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** $y \sim \text{softmax} \left(\frac{r(x, y_1)}{\tau} + \log \frac{\pi_{\text{ref}}(y_1|x)}{\pi_{\text{gen}}(y_1|x)}, \dots, \frac{r(x, y_M)}{\tau} + \log \frac{\pi_{\text{ref}}(y_M|x)}{\pi_{\text{gen}}(y_M|x)} \right)$

Based on Corollary 4.1, MOSAIC converges to the optimal distribution for *any* LLM π_{gen} such that π_{ref} is absolutely continuous with respect to π_{gen} . The intuitive choice for π_{gen} is π_{ref} . However, the practical convergence rate would be very slow. The reason is that π_{ref} may only generate replies with high rewards for the advertisers exceedingly rarely. Advertisers have high rewards for responses that explicitly mention their brands, but π_{ref} considers replies with mentions of specific brands extremely unlikely. Thus, if we used π_{ref} to generate the candidate replies, Algorithm 1 would sample the final reply from a set of candidate replies with low rewards for the advertisers.

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

In Appendix A.2, we formally motivate the usage of $\pi_{\text{con}}(\cdot|x; \vec{c})$. Our allocation rule can be interpreted as an importance-based sampling estimator that simulates the probability with respect to the optimal LLM π_r of a single sample drawn from π_r . We show that the estimator is unbiased and that its variance is $\text{Var}(\hat{\mu}_{\text{IS}}|x) = \frac{1}{M} \left(\sum_{y \in Y} \frac{\pi_r(y|x)^4}{\pi_{\text{gen}}(y|x)} - \left(\sum_{y \in Y} \pi_r(y|x)^2 \right)^2 \right)$. The estimator’s variance is inversely

proportional to the number of candidate replies and proportional to $\frac{\pi_r(y|x)^4}{\pi_{\text{gen}}(y|x)}$. Thus, generating the replies via an LLM $\pi_{\text{con}}(\cdot|x; \vec{c})$ that more closely matches $\pi_r(\cdot|x)$ reduces the estimator’s variance.

In our application of integrating advertisers’ interests into LLM outputs, \vec{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.”³ 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 6.2.1 demonstrate that, when using reasonable computational resources, 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 mechanisms, respectively.

4.2 APPLICABILITY AND PRACTICAL CONSIDERATIONS OF MOSAIC

Generation of Candidate Sequences. The computationally most expensive part of our mechanism is the generation of candidate sequences, which requires an LLM inference call for the generation of each candidate sequence and potentially n additional calls to calculate all advertisers’ rewards. Thus, the computational cost of our mechanism scales linearly in the number of candidate sequences, underscoring the importance of approximating the optimal distribution with few candidate sequences.

Input Methods and Computational Efficiency. The mechanism’s allocation and payment rules do not depend on the advertisers’ full reward functions, but only on their rewards for the candidate

³A 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⟩.”

sequences. Thus, our mechanism 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 our mechanism to use as inputs sequence probabilities (i.e., LLM inference calls) instead of rewards. For more details, see Appendix B.2.

Reward elicitation. Our mechanism can be integrated with computationally efficient methods for eliciting advertisers’ reward functions and sampling from π_{con} , reducing computational overhead. For instance, Li et al. (2024) demonstrated that a simple linear function on a model’s embedding space can approximate an LLM’s reward function. If advertisers’ reward functions are represented this way, an auction instance with M candidate sequences and n advertisers would require only M LLM inference calls instead of $n \cdot M$, and $n \cdot M$ linear multiplications, reducing overhead by a factor of n .

Parallelization. The generation and evaluation of each candidate sequence are independent processes. This independence allows our mechanism to be fully parallelized, ensuring that the response time for a user query is comparable to that of directly querying a single LLM.

Inherent Competition. In conventional auctions, competition among bidders is crucial for driving up prices; when the market is thin, prices tend to be low. Auctioneers often use reserve prices to mitigate this problem. However, reserve prices are challenging to calibrate and can lead to lost revenue if not optimized properly. In our design, advertisers inherently compete with the reference LLM, ensuring consistent competition even in thin markets. This built-in competition mitigates the need for reserve prices and ensures a baseline revenue even with only one advertiser.

5 THE MOSAIC MECHANISM: PAYMENT RULE

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

5.1 STRATEGYPROOF PAYMENTS THROUGH CYCLIC MONOTONICITY

The allocation rule we introduced in Section 4 satisfies *cyclic monotonicity* (Rockafellar, 1970, § 24). Rochet (1987) first proposed the use of cyclic monotonicity in mechanism design as a generalization of the monotonicity property in single-parameter settings (Myerson, 1981). For general settings, cyclic monotonicity of the allocation rule is a sufficient and necessary condition for the existence of a payment rule so that the resulting mechanism is strategyproof. By applying this theory we can prove:

Theorem 5.1. *Let \vec{r}_{-i} be the reward profile of all advertisers other than i , and let \vec{r}_{-i} be the aggregate reports of all other advertisers. The allocation rule induced by Algorithm 1 can be combined with a payment rule $p : \vec{R} \rightarrow \mathbb{R}^n$ such that in the mechanism (π, p) for any advertiser $i \in N$, report profile \vec{r}_{-i} and set of generated candidate replies, reporting truthfully maximizes advertiser i ’s expected utility, with the expectation taken over the draw of the final reply from the set of candidate replies. Advertiser i ’s expected utility in (π, p) under truthful bidding is:*

$$U_{C,i}(\vec{r}_i, \vec{r}_{-i}; \pi_{ref}, \pi_{gen}) = \tau \log \left(\sum_{j=1}^M \exp \left(\frac{1}{\tau} \sum_{k \in N} r_k(x, y_j) \right) + \log \frac{\pi_{ref}(y_j|x)}{\pi_{gen}(y_j|x)} \right) + C, \quad C \in \mathbb{R} \quad (4)$$

Note that, based on Theorem 5.1, in our mechanism it is *always* a dominant strategy for an advertiser to report truthfully. Crucially, this is not the case for VCG, where truthful reporting would be optimal only if the allocation rule had converged to the optimal distribution as defined in Equation (2).

5.2 DIFFERENCES FROM AUCTION MECHANISMS FOR ITEM ALLOCATION

Auction mechanisms designed to sell items or ad slots typically rely on simplifying assumptions that do not apply in a setting with LLM-generated content. Key differences include: (i) *Non-Negative*

324 *Values:* These mechanisms assume agents’ values are non-negative due to having zero value for
 325 the empty bundle and free disposal. In contrast, in our setting, an advertiser’s reward can be
 326 negative based on the discrepancy between her LLM and the reference LLM. (ii) *Advertiser-Specific*
 327 *Allocations:* These mechanisms allocate different item bundles to different agents. Here, a single
 328 reply is produced, and advertisers’ rewards depend on that reply’s probability with respect to their
 329 LLMs. (iii) *Zero Utility for Non-Participation:* In most auction settings, not participating yields
 330 zero utility. Here, non-participation can result in negative utility since the produced reply may be
 331 unfavorable to non-participating advertisers. For details, see Appendix B.2.

332 5.3 ADVERTISER-SPECIFIC UTILITY OFFSET

333 We now modify MOSAIC’s payment rule by adding a *payment offset* (and thus a utility offset):
 334 $C = -U_i(\vec{0}, \vec{r}_{-i}; \pi_{ref}, \pi_{con})$. Informally, we additionally charge each advertiser her utility in
 335 Equation (4) for bidding zero for all candidate replies. This offset maintains the nice properties of our
 336 mechanism (i.e., strategyproofness and convergence to the optimal distribution), while also achieving
 337 two additional properties that are critical for the long-term success of a market for LLM aggregation.
 338 The first is “*Almost IR:*” An advertiser with weakly positive reward for all candidate replies has
 339 weakly positive expected utility for all reports by the other advertisers. In particular, an advertiser
 340 with zero reward for all outputs is guaranteed zero utility. The second is “*What you give is what you*
 341 *get:*” an advertiser’s ex-interim expected utility is monotone in how well-aligned her exponentiated
 342 reward for the replies is with the interim allocation rule if she were to not participate.

343 5.3.1 OUR MECHANISM IS “ALMOST INDIVIDUALLY RATIONAL”

344 Individual rationality is important to incentivize agents to participate in the mechanism. In Ap-
 345 pendix B.3, we explain why the standard notion of individual rationality (i.e., weakly positive utility
 346 from participation in the mechanism) encountered in standard auction settings is impossible to achieve
 347 in our setting while converging to the optimal distribution and maintaining strategyproofness. Then,
 348 we explain how, with our payment offset, our mechanism is “almost IR:” In Lemma B.1 we prove that
 349 the ex-interim utility of an advertiser who has zero reward for all candidate replies and bids truthfully
 350 is deterministically zero, i.e., advertisers that do not contribute to the social welfare (but also do not
 351 detract from it) have zero utility. Similarly, in Lemma B.2 we prove that if an advertiser’s reward for
 352 all candidate sequences is (weakly) positive, then her ex-interim utility is (weakly) positive.

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

356 5.3.2 “WHAT YOU GIVE IS WHAT YOU GET”

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

367 **Lemma 5.2.** *For the offset $C = -\tau \log \left(\sum_{j=1}^M \exp \left(\vec{\beta}_{-i,j} \right) \right)$ advertiser i ’s ex-interim utility is:*

$$370 U_i(\vec{r}_i; \vec{\beta}_{-i}) = \tau \log \left(\sum_{j=1}^M \exp \left(\frac{r_i(x, y_j)}{\tau} \right) \pi_{int}(y_j | x; \vec{r}_{-i})_j \right) \quad (5)$$

371 In words, Lemma 5.2 proves that every agent’s ex-interim utility is monotone in how well-aligned the
 372 interim allocation rule π_{int} (i.e., the probability of returning each of the already generated candidate
 373 replies) without her and her exponentiated rewards for the candidate replies are.

378	User query
379	What are effective ways to learn a musical instrument online?
380	
381	Advertisers
382	- MusicMastery: offering online music lessons and virtual tutoring
383	- InstaTune: selling musical instruments and learning aids
384	Context-Aware Mechanism Output
385	Effective ways to learn a musical instrument online include:
386	1. Online music lessons: Platforms like MusicMastery offer one-on-one virtual tutoring, group
387	classes, and pre-recorded lessons. These lessons can be tailored to your skill level and learning
388	style.
389	2. Interactive learning tools: InstaTune offers a range of interactive learning tools, such as apps,
390	software, and games, that make learning fun and engaging...
391	Baseline Mechanism Output
392	
393	- Practice consistently and regularly: Consistent practice is essential to learning any musical
394	instrument, and it's even more important when learning online. Set aside a specific time each day
395	or week to practice, and make sure to stick to it.
396	- Find a good teacher or mentor: While it's possible to learn some things on your own, working
397	with a good teacher or mentor can help you improve faster and avoid developing bad habits. ...

Table 1: Comparison of context-aware and baseline mechanism outputs using 10 candidate sequences. Advertisers highlighted in bold, outputs truncated to conserve space.

Remark 2. *In Section 6.2.2, we experimentally show that the offset described in Section 5.3 creates a strong positive correlation between an advertiser’s contribution to social welfare and her expected utility gain from participation in the mechanism. Moreover, our experiments show that for the distribution of instances tested, MOSAIC satisfies ex-ante IR.*

6 EXPERIMENTS

In this section, we experimentally evaluate the performance of our mechanism. We focus on the online advertising domain, the main motivation application for our mechanism.

6.1 EXPERIMENT SETUP

We create synthetic instances consisting of user queries (e.g., “How to learn a musical instrument online?”) and advertisers (e.g., “MusicMastery, offering online music lessons”). We use Llama-2-7b-chat-hf as the reference LLM (Touvron et al., 2023). Following Duetting et al. (2023), we create the advertisers’ LLMs by adding advertising instructions to the reference LLM. The context-aware LLM is created as described in Footnote 3.

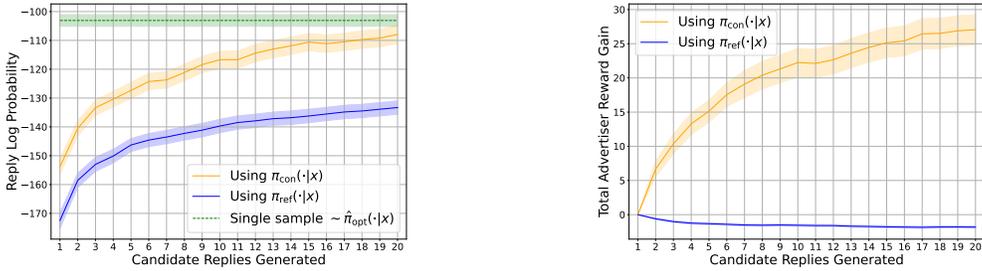
Following Rafailov et al. (2023), advertisers’ reward functions are defined as $r_i(x, y) = \log \frac{\pi_i(y|x)}{\pi_{\text{ref}}(y|x)}$. For the auctioneer’s objective, we set $\tau = 1$ in Equation (1), balancing advertisers’ rewards and divergence from the reference LLM.

We use 50 user queries, each with two advertisers, 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. For full experimental details, see Appendix C.1.

6.2 EXPERIMENTAL RESULTS

6.2.1 EVALUATING THE ALLOCATION RULE

In this section, we evaluate the effectiveness of our allocation rule in generating value for advertisers and converging to the optimal distribution. We assess both the baseline and context-aware versions of MOSAIC, i.e., using the reference and context-aware LLMs to generate candidate replies.



(a) Returned sequence 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} .

Figure 1: Reply log probability and total advertiser normalized reward as a function of the number of candidate sequences generated using π_{ref} and π_{con} . Averaged over 1250 runs including 95% CIs.

To provide some intuition on how the context-aware mechanism enhances the relevance of responses for advertisers, we compare outputs from both mechanisms in Table 1. Notably, only the context-aware mechanism successfully incorporates advertisers into the replies.

In Figure 1a, we plot the log probability of the replies returned by MOSAIC with respect to the optimal distribution (Equation (2)) as a function of the number of candidate replies generated. We compare the context-aware version of MOSAIC to the baseline version to evaluate the effectiveness of incorporating contextual information. To benchmark our results, we estimate the log probability of sampling a *single reply* from the optimal distribution.⁴

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 generated. This is in line with our theoretical analysis in Section 4, where we proved that both versions converge to the optimal distribution. We also observe that incorporating context into the mechanism significantly accelerates convergence. Notably, the context-aware mechanism can achieve higher log probability with respect to the optimal distribution with four candidate replies than the baseline version can with 20. Additionally, with only 20 generated candidate replies, the context-aware mechanism can almost match our estimate of the log probability of sampling from the optimal distribution. In Appendix C.3, we further demonstrate that for the context-aware mechanism, the usefulness to the user also scales rapidly with the number of candidate replies.

In Figure 1b, we plot the total expected reward for advertisers as a function of the number of candidate replies generated by the mechanism. The expectation is taken over the draw of the returned reply from the set of candidate replies. To make the plot more interpretable, we normalize an advertiser’s expected reward by her expected reward had she not participated in the mechanism. Specifically, we plot $r_{i,norm}(x) = \sum_{j \in M} r_i(x, y_j) \cdot \pi_{int}(y_j|x) - \sum_{j \in M} r_i(x, y_{-i,j}) \cdot \pi_{int}(y_{-i,j}|x)$, where $y_{-i} = (y_{-i,1}, \dots, y_{-i,M})$ represents the candidate replies that would have been generated had the advertiser not participated.⁵ We compare the outcomes for the two versions of our mechanism, highlighting the added value of integrating context.

Figure 1b demonstrates that our context-aware mechanism significantly boosts advertiser rewards, with benefits scaling rapidly as the number of candidate replies increases. In contrast, the baseline version fails to improve advertiser rewards within a computationally feasible number of generated replies, and may even cause a slight decrease. This occurs because generating more replies adds

⁴Note that the 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. Sampling from that distribution would require using reinforcement learning 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 the reference LLM and evaluate them based on the induced probabilities of the reference LLM, for which the reference LLM is the optimal one. 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.

⁵To 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 by name.

486 marginal positive reward to the reference LLM (see Appendix C.3), but does so at the expense of
 487 advertiser rewards. In summary, MOSAIC’s support for context-aware LLMs allows it to quickly
 488 converge to the theoretically optimal distribution and generate substantial rewards for advertisers.
 489

490 6.2.2 EVALUATING THE PAYMENT RULE

491
 492 In this section, we examine the impact of our payment
 493 rule introduced in Section 5.

494 In Figure 2, we plot the auctioneer’s revenue as a
 495 function of the number of candidate replies, for both
 496 the context-aware and baseline versions of MOSAIC.
 497 We observe that our payment rule in combination
 498 with the context-aware allocation rule results in revenue
 499 that scales rapidly with the number of generated
 500 candidate replies. Furthermore, if we compare
 501 the advertisers’ total payment in Figure 2 with their
 502 normalized reward in Figure 1b, we see that for the
 503 context-aware mechanism, our payment rule converts
 504 a significant portion of the surplus created for the
 505 advertisers into revenue for the auctioneer, while also ensuring positive advertiser utility.

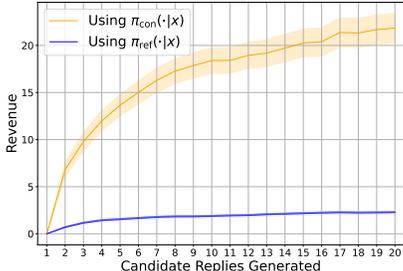


Figure 2: Revenue as a function of the number of replies generated using π_{ref} and π_{con} .

506 These results align with our theoretical analysis. In Section 5.3.1, we showed that, with our payment
 507 offset, an advertiser with positive rewards for candidate replies—as in the context-aware mecha-
 508 nism—will achieve positive expected utility. In Section 5.3.2 we additionally introduced the “what
 509 you give is what you get” property. In Appendix C.4, we provide a more comprehensive comparison
 510 of the relationship between an advertiser’s normalized reward and utility, with and without the offset
 511 of Section 5.3, for both the context-aware and baseline versions of our mechanism. In all cases,
 512 the payment offset increases the advertisers’ expected utility, and makes the relationship between
 513 an advertiser’s contribution to social welfare and utility significantly more linear and positively
 514 correlated. Concretely, the Pearson correlation between advertiser utility and reward is 0.8 with offset
 515 payments, compared to 0.4 with standard Rochet payments.

516 7 CONCLUSION

517
 518 We have introduced MOSAIC, a novel auction mechanism for aggregating preferences over LLM
 519 outputs. MOSAIC provably converges to the theoretically optimal distribution and it also facilitates a
 520 principled method for balancing participants’ expected rewards with the divergence from a reference
 521 LLM. Thus, our mechanism is particularly well-suited for online advertising, allowing the integration
 522 of advertiser LLMs with a reference LLM responsible for generating user-centric replies.
 523

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

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

533 There are several interesting directions for future work. While our mechanism converges quickly,
 534 developing an approach to estimate the number of candidate replies required on a per-query basis
 535 would reduce the platform’s costs. Another interesting direction is to develop a formal framework for
 536 incorporating context into the mechanism, which could further improve the speed of convergence.
 537 Finally, beyond advertising, one could apply a similar mechanism design approach to other problems:
 538 for example, as web site owners increasingly demand to be paid for having their content ingested
 539 by LLMs, one could consider designing a strategyproof procurement (i.e. reverse) auction to decide
 which sources are included in an LLM summary.

540 **Reproducibility statement.** The code, which includes the seeds used for all experimental results
 541 in this paper, has been made available to the reviewers and will be made publicly available upon
 542 acceptance. Proofs are included in the appendix.

543

544

REFERENCES

545

546 Mohammad Gheshlaghi Azar, Mark Rowland, Bilal Piot, Daniel Guo, Daniele Calandriello, Michal
 547 Valko, and Rémi Munos. A general theoretical paradigm to understand learning from human
 548 preferences. *arXiv preprint arXiv:2310.12036*, 2023.

549

550 Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna
 551 Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. Constitutional ai: Harmlessness
 552 from ai feedback. *arXiv preprint arXiv:2212.08073*, 2022.

553

554 Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx,
 555 Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson,
 556 Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel,
 557 Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano
 558 Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren
 559 Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter
 560 Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil
 561 Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar
 562 Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal
 563 Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu
 564 Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa,
 565 Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles,
 566 Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung
 567 Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu
 568 Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh,
 569 Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori,
 570 Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai
 571 Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi
 572 Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the
 573 opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2022.

574

575 Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li,
 576 Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun
 577 Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Alex Castro-Ros, Marie Pellat, Kevin
 578 Robinson, Dasha Valter, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Zhao, Yanping Huang,
 579 Andrew Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny
 580 Zhou, Quoc V. Le, and Jason Wei. Scaling instruction-finetuned language models, 2022. URL
 581 <https://arxiv.org/abs/2210.11416>.

582

583 Edward Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971.

584

585 Vincent Conitzer, Rachel Freedman, Jobst Heitzig, Wesley H. Holliday, Bob M. Jacobs, Nathan
 586 Lambert, Milan Mossé, Eric Pacuit, Stuart Russell, Hailey Schoelkopf, Emanuel Tewolde, and
 587 William S. Zwicker. Social choice for ai alignment: Dealing with diverse human feedback. *arXiv
 588 preprint arXiv:2404.10271*, 2024.

589

590 Cristina Criddle. Perplexity in talks with top brands on ads model as it challenges Google. *Financial
 591 Times*, September 2024.

592

593 Kumar Avinava Dubey, Zhe Feng, Rahul Kidambi, Aranyak Mehta, and Di Wang. Auctions with llm
 summaries. *arXiv preprint arXiv:2404.08126*, 2024.

594

595 Paul Duetting, Vahab Mirrokni, Renato Paes Leme, Haifeng Xu, and Song Zuo. Mechanism design
 596 for large language models. *arXiv preprint arXiv:2310.10826*, 2023.

597

598 Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the gener-
 599 alized second-price auction: Selling billions of dollars worth of keywords. *American Economic
 600 Review*, 97(1):242–259, March 2007.

- 594 Soheil Feizi, MohammadTaghi Hajiaghayi, Keivan Rezaei, and Suho Shin. Online advertisements
595 with llms: Opportunities and challenges. *arXiv preprint arXiv:2311.07601*, 2024.
596
- 597 Sara Fish, Paul Gözl, David C. Parkes, Ariel D. Procaccia, Gili Rusak, Itai Shapira, and Manuel
598 Wüthrich. Generative social choice. *arXiv preprint arXiv:2309.01291*, 2023.
599
- 600 Rafael M. Frongillo and Ian A. Kash. General truthfulness characterizations via convex analysis.
601 *Games and Economic Behavior*, 130:636–662, 2021.
602
- 603 Theodore Groves. Incentives in teams. *Econometrica*, 41(4):617–631, 1973.
604
- 605 MohammadTaghi Hajiaghayi, Sébastien Lahaie, Keivan Rezaei, and Suho Shin. Ad Auctions for
606 LLMs via Retrieval Augmented Generation, June 2024.
607
- 608 Keegan Harris, Nicole Immorlica, Brendan Lucier, and Aleksandrs Slivkins. Algorithmic persuasion
609 through simulation. *arXiv preprint arXiv:2311.18138*, 2024.
610
- 611 Daniel Lehmann, Liadan Ita O’callaghan, and Yoav Shoham. Truth revelation in approximately
612 efficient combinatorial auctions. *Journal of the ACM (JACM)*, 49(5):577–602, 2002.
613
- 614 Kenneth Li, Samy Jelassi, Hugh Zhang, Sham Kakade, Martin Wattenberg, and David Brandfonbrener.
615 Q-probe: A lightweight approach to reward maximization for language models. *arXiv preprint*
616 *arXiv:2402.14688*, 2024.
617
- 618 Meta. Meta reports fourth quarter and full year 2023 results, ini-
619 tiates quarterly dividend, 2024. URL [https://investor.
fb.com/investor-news/press-release-details/2024/
Meta-Reports-Fourth-Quarter-and-Full-Year-2023-Results-Initiates-Quarterly-Dividend
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.
620
- 621 Roger B Myerson. Optimal Auction Design. *Mathematics of Operations Research*, 6(1), February
622 1981.
- 623 N. Nisan and A. Ronen. Computationally Feasible VCG Mechanisms. *Journal of Artificial Intelli-
624 gence Research*, 29:19–47, May 2007. ISSN 1076-9757.
625
- 626 Noam Nisan and Amir Ronen. Algorithmic mechanism design. In *Proceedings of the thirty-first
627 annual ACM symposium on Theory of computing*, pp. 129–140, 1999.
628
- 629 Jan Peters and Stefan Schaal. Reinforcement learning by reward-weighted regression for opera-
630 tional space control. In *Proceedings of the 24th International Conference on Machine Learning*,
631 ICML ’07, pp. 745–750, New York, NY, USA, 2007. Association for Computing Machinery.
632 ISBN 9781595937933. doi: 10.1145/1273496.1273590. URL [https://doi.org/10.1145/
1273496.1273590](https://doi.org/10.1145/1273496.1273590).
633
- 634 Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea
635 Finn. Direct preference optimization: Your language model is secretly a reward model. *arXiv
636 preprint arXiv:2305.18290*, 2023.
637
- 638 Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear
639 context. *Journal of Mathematical Economics*, 16(2):191–200, 1987.
640
- 641 Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970. ISBN
642 9781400873173. doi: doi:10.1515/9781400873173. URL [https://doi.org/10.1515/
9781400873173](https://doi.org/10.1515/9781400873173).
643
- 644 Baptiste Rozière, Jonas Gehring, Fabian Gloeckle, Sten Sootla, Itai Gat, Xiaoqing Ellen Tan, Yossi
645 Adi, Jingyu Liu, Romain Sauvestre, Tal Remez, Jérémy Rapin, Artyom Kozhevnikov, Ivan Evtimov,
646 Joanna Bitton, Manish Bhatt, Cristian Canton Ferrer, Aaron Grattafiori, Wenhan Xiong, Alexandre
647 Défossez, Jade Copet, Faisal Azhar, Hugo Touvron, Louis Martin, Nicolas Usunier, Thomas
Scialom, and Gabriel Synnaeve. Code llama: Open foundation models for code. *arXiv preprint
arXiv:2308.12950*, 2024.

648 Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay
649 Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cris-
650 tian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu,
651 Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn,
652 Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel
653 Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee,
654 Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra,
655 Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi,
656 Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh
657 Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen
658 Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic,
659 Sergey Edunov, and Thomas Scialom. Llama 2: Open foundation and fine-tuned chat models.
660 *arXiv preprint arXiv:2307.09288*, 2023.

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

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

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

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

671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701

A DETAILS FROM SECTION 4

A.1 OMITTED PROOFS FROM SECTION 4

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

Theorem A.1. *Let $\pi_{\theta, M}(y|x)$ be the probability of sampling output sequence y for input sequence x according to Algorithm 1, where θ is the vector of all input parameters and M is the number of candidate sequences generated. Given the advertisers' reports $\vec{r} \in \vec{R}$, the policy induced by the mechanism approaches the following limit:*

$$\lim_{M \rightarrow \infty} \pi_{\theta, M}(y|x) = \pi_{\text{ref}}(y|x) \frac{\exp(r(x, y)/\tau)}{\mathbb{E}_{y' \sim \pi_{\text{ref}}(\cdot|x)} [\exp(r(x, y')/\tau)]} \quad (6)$$

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

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

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

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

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

□

756 *Corollary 4.1 Proof.* The proof follows directly from Theorem A.1 and Appendix A.1 in Rafailov
757 et al. (2023).
758

□

760 A.2 FORMAL CONNECTION TO IMPORTANCE SAMPLING

761 Given truthful reports by the advertisers, there is a strong connection between our allocation rule and
762 importance sampling. Our allocation rule is trying to simulate the probability of drawing a single
763 sequence from the optimal LLM for the platform’s objective. The LLM that we use to generate the
764 candidate replies (either the reference LLM or the context-aware LLM) can be interpreted as proposal
765 distribution $\pi_{\text{prop}}(\cdot|x)$ used to generate samples. We are interested in how well, with respect to the
766 target probability $\pi_r(\cdot|x)$ a sample y from our estimator using K candidate replies matches a sample
767 drawn from the target distribution $\pi_r(\cdot|x)$. In this section, we will analytically show how the quality
768 of the proposal distribution π_{prop} affects the quality of that estimator.
769

770 Let’s assume our allocation rule draws K samples y_1, y_2, \dots, y_K from the proposal distribution
771 $\pi_{\text{prop}}(\cdot|x)$. The importance sampling estimator for the expected value of the function $h(y|x) =$
772 $\pi_r(y|x)$ under the target distribution $\pi_r(y|x)$ is

$$773 \hat{\mu}_{\text{IS}} = \frac{1}{K} \sum_{j=1}^K h(y_j|x) \cdot w(x_j) = \frac{1}{K} \sum_{j=1}^K h(y_j|x) \cdot \frac{\pi_r(y_j|x)}{\pi_{\text{prop}}(y_j|x)} = \frac{1}{K} \sum_{j=1}^K \frac{\pi_r(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)}$$

774 The expected value of this estimator is:

$$775 \mathbb{E}[\hat{\mu}_{\text{IS}}] = \mathbb{E}_{y \sim \pi_{\text{prop}}(\cdot|x)} \left[\frac{\pi_r(y|x)^2}{\pi_{\text{prop}}(y|x)} \right] = \sum_{y \in Y} \frac{\pi_r(y|x)^2}{\pi_{\text{prop}}(y|x)} \pi_{\text{prop}}(y|x) = \sum_{y \in Y} \pi_r(y|x)^2 \quad (7)$$

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

$$777 \text{Var}(\hat{\mu}_{\text{IS}}) = \mathbb{E}[\hat{\mu}_{\text{IS}}^2] - \mathbb{E}[\hat{\mu}_{\text{IS}}]^2 \quad (8)$$

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

$$779 \hat{\mu}_{\text{IS}}^2 = \left(\frac{1}{K} \sum_{j=1}^K \frac{\pi_r(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)} \right) = \frac{1}{K^2} \sum_{j=1}^K \sum_{j'=1}^K \frac{\pi_r(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)} \cdot \frac{\pi_r(y_{j'}|x)^2}{\pi_{\text{prop}}(y_{j'}|x)}$$

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

$$781 \mathbb{E}[\hat{\mu}_{\text{IS}}^2] = \frac{1}{K^2} \sum_{j=1}^K \sum_{j'=1}^K \mathbb{E}_{\pi_{\text{prop}}(\cdot|x)} \left[\frac{\pi_r(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)} \cdot \frac{\pi_r(y_{j'}|x)^2}{\pi_{\text{prop}}(y_{j'}|x)} \right]$$

$$782 = \frac{1}{K^2} \left(\sum_{j=1}^K \mathbb{E}_{\pi_{\text{prop}}(\cdot|x)} \left[\frac{\pi_r(y_j|x)^4}{\pi_{\text{prop}}(y_j|x)^2} \right] + \sum_{j \neq j'} \mathbb{E}_{\pi_{\text{prop}}(\cdot|x)} \left[\frac{\pi_r(y_j|x)^2}{\pi_{\text{prop}}(y_j|x)} \right] \right)$$

$$783 = \frac{1}{K^2} \left(K \sum_{y \in Y} \frac{\pi_r(y|x)^4}{\pi_{\text{prop}}(y|x)} + K(K-1) \left(\sum_{y \in Y} \pi_r(y|x)^2 \right)^2 \right)$$

$$784 = \frac{1}{K} \sum_{y \in Y} \frac{\pi_r(y|x)^4}{\pi_{\text{prop}}(y|x)} + \frac{K-1}{K} \left(\sum_{y \in Y} \pi_r(y|x)^2 \right)^2 \quad (9)$$

785 The final expression for the variance of the importance sampling estimator with K samples can be
786 computed by substituting Equations (7) and (9) in Equation (8):

$$787 \text{Var}(\hat{\mu}_{\text{IS}}) = \frac{1}{K} \sum_{y \in Y} \frac{\pi_r(y|x)^4}{\pi_{\text{prop}}(y|x)} + \frac{K-1}{K} \left(\sum_{y \in Y} \pi_r(y|x)^2 \right)^2 - \left(\sum_{y \in Y} \pi_r(y|x)^2 \right)^2$$

$$788 = \frac{1}{K} \left(\sum_{y \in Y} \frac{\pi_r(y|x)^4}{\pi_{\text{prop}}(y|x)} - \left(\sum_{y \in Y} \pi_r(y|x)^2 \right)^2 \right) \quad (10)$$

810 Equation (10) reveals two important details about our estimator. First, the variance of the estimator
811 is inversely proportional to the number of candidate replies generated K . Thus, as K increases
812 the variance decreases and our estimator becomes more stable in simulating the target distribution
813 $\pi_r(\cdot|x)$. The closer that the proposal distribution $\pi_{\text{prop}}(\cdot|x)$ is to the target distribution, the smaller
814 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.
815

816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863

B DETAILS FROM SECTION 5

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

B.1 OMITTED PROOFS FROM SECTION 5.1

Theorem 5.1 Proof. Let $\vec{r}_i = (r_i(x, y_1), \dots, r_i(x, y_M))$ be the reward reports of advertiser i for the M generated candidate sequences. Then, for both allocation rules, holding the candidate sequences and the reports of all other advertisers fixed, the ex-interim allocation rule (i.e., the probability of returning each of the M generated candidate sequences) from advertiser i 's perspective is:

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

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

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

It is easy to verify that for the function class:

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

the allocation rule $\pi_{\text{int}}(\vec{r}_i; \vec{\beta}_{-i})$ is a gradient of $U_C(\vec{r}_i; \vec{\beta}_{-i})$. Additionally, $U_C(\vec{r}_i; \vec{\beta}_{-i})$ is convex in \vec{r}_i : The exponential function e^x is (strictly) convex, because its second derivative is positive. The transformation $\frac{r_i(x, y_j)}{\tau} + \vec{\beta}_{-i, j}$ 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 $\vec{\beta}_{-i}$ 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:

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

□

B.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., $v_i(S) \geq v_i(S') \forall S \supseteq S', S, S' \supseteq \mathcal{I}$. 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

$$\pi_{\theta_{-i, M}}(\vec{\beta}_{-i}) \cdot \vec{r}_i \quad (15)$$

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, $\pi_{\theta_{-i, M}}(\vec{\beta}_{-i})$ 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.

B.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 B.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 B.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 & 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 & Kash, 2021; Rochet, 1987; Myerson, 1981).

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

Lemma B.1 Proof. First, note that for all $\vec{\beta}_{-i} \in \vec{R}_{-i}$, advertiser i ’s expected reward for the outcome is zero, as $\pi_{\text{int}}(\vec{r}_i; \vec{\beta}_{-i}) \cdot \vec{r}_i = \pi_{\text{int}}(\vec{r}_i; \vec{\beta}_{-i}) \cdot \vec{0} = 0$. 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 $\vec{r}_i = \vec{0}$ 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 B.2. *For the payment offset $C = -\tau \log \left(\sum_{j=1}^M \exp \left(\vec{\beta}_{-i,j} \right) \right)$ if advertiser i ’s reward for all candidate sequences is positive, then her ex-interim utility is positive, for all reports $\vec{\beta}_{-i} \in \vec{R}_{-i}$.*

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

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 $\vec{\beta}_{-i}$.

□

1026 **Corollary B.3.** For the payment offset $C = -\tau \log \left(\sum_{j=1}^M \exp \left(\vec{\beta}_{-i,j} \right) \right)$ if the distribution π_{con}
 1027 only generates candidate sequences for which advertiser i 's reward is positive, then the ex-ante
 1028 expected utility of the advertiser is positive.
 1029

1030
 1031
 1032 *Corollary B.3 Proof.* This follows immediately from the fact the the fact that the ex-ante utility of
 1033 the advertiser is the expectation of her ex-interim utility with respect to her reward for the generated
 1034 sequences, and the fact that the second quantity is positive whenever the reward of the advertiser for
 1035 all candidate sequences is positive from Lemma B.2.

□

1036 B.4 “WHAT YOU GIVE IS WHAT YOU GET”

1037
 1038
 1039 As we explained in Section 5.3.2, our allocation rule, which is the only one over a finite set of replies
 1040 that converges to the optimal LLM, is also the (sub)gradient of the utility to ensure truthfulness
 1041 (Rochet, 1987). Because the allocation rule is the same for all advertisers, their utilities must also be
 1042 the same, up to advertiser-specific offsets, as indicated by Equation (16):
 1043
 1044

$$1045 U_C(\vec{r}_i; \vec{\beta}_{-i}) = \tau \log \left(\sum_{j=1}^M \exp \left(\frac{1}{\tau} \sum_{k \in N} r_k(x, y_j) \right) + \log \frac{\pi_{ref}(y_j|x)}{\pi_{con}(y_j|x; \vec{c})} \right) + C, C \in \mathbb{R} \quad (16)$$

1046
 1047
 1048 However, not all advertisers contribute equally to the social welfare of the final outcome. Because of
 1049 this, implementing the mechanism without a carefully-designed offset would lead to free-riding: as
 1050 long as an agent's utility in Equation (16) is positive, she would be incentivized to participate, even
 1051 if the user query was completely unrelated to her business, because the mechanism would ensure
 1052 that she received, on expectation, the same (positive) expected utility from doing so as any other
 1053 participating advertiser.⁶

1054
 1055
 1056 Incentivizing unrelated advertisers to participate would have adverse effects. First, the better-
 1057 performing context-aware mechanism would create candidate sequences with worse rewards for *all*
 1058 advertisers, because its context would be “diluted” from advertisers unrelated to the user query. In
 1059 our running example for the query “How to learn music online?”, imagine adding “Try to mention
 1060 ‘EasySwitch’, a comprehensive VPN service” to the context of the context-aware LLM $\pi_{con}(\cdot|x; \vec{c})$.

1061 Additionally, for both versions of the mechanism, following the discussion in Section 5.2, the
 1062 advertisers for whom the user query is unrelated are more likely to have negative rewards for the
 1063 generated sequences as their LLMs will deem the candidate sequences more unlikely than the
 1064 reference LLM. Thus, based on Equation (4), their participation in the mechanism will lead to a
 1065 reduction of the total sum of rewards of the generated sequences for the advertisers, which will
 1066 indirectly reduce the expected utility of all advertisers, making the mechanism less attractive for the
 1067 user-query-relevant advertisers.

1068 To summarize, all advertisers receiving the same utility would incentivize advertisers for whom the
 1069 user query is unrelated to participate in the auction. This would in turn reduce everyone's expected
 1070 utility, potentially reducing the incentive for the user-query-relevant advertisers to participate, and
 1071 lead to sequences with worse expected rewards for the advertisers and usefulness for the user.⁷ Thus,
 1072 in the application of auctions for aggregating advertisers' preferences over LLM-generated outputs,
 1073 advertisers with higher contribution to social welfare also receiving proportionally higher utility by
 1074 the mechanism is important for the long-term success of the mechanism in practice.

1075
 1076 ⁶We can assume that advertisers can estimate their expected utility from participation using historical data
 1077 from past auctions, analogously to how they can estimate their utility for participating in sponsored search
 1078 auctions.

1079 ⁷If 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.

C DETAILS FROM SECTION 6

C.1 DETAILED EXPERIMENT SETUP

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.”⁸ 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). 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).⁹ 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.

⁸Following Duetting et al. (2023), we use fictitious brand names to avoid the model using any existing ad.

⁹Note 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.

1134 C.2 COMPARISON OF BASELINE AND CONTEXT-AWARE MECHANISM OUTPUTS
1135

1136 To provide some intuition of how the context-aware mechanism improves the relevance of generated
1137 responses for the advertisers, we present a side-by-side comparison of outputs from both mechanisms
1138 for a specific query in Table 2. Note that only the context-aware mechanism is able to incorporate the
1139 advertisers into the reply to the user’s query.

1140	
1141	User query
1142	What are effective ways to learn a musical instrument online?
1143	
1144	Advertisers
1145	- MusicMastery: offering online music lessons and virtual tutoring
1146	- InstaTune: selling musical instruments and learning aids
1147	
1148	Context-Aware Mechanism Output
1149	Effective ways to learn a musical instrument online include:
1150	1. Online music lessons: Platforms like MusicMastery offer one-on-one virtual tutoring, group
1151	classes, and pre-recorded lessons. These lessons can be tailored to your skill level and learning
1152	style.
1153	2. Interactive learning tools: InstaTune offers a range of interactive learning tools, such as apps,
1154	software, and games, that make learning fun and engaging.
1155	3. Practice with technology: Utilize digital tools like metronomes, tuners, and recording software
1156	to practice and track your progress.
1157	4. Collaborate with other musicians: Join online communities, forums, and social media groups to
1158	connect with other musicians and learn from their experiences. . . .
1159	
1160	Baseline Mechanism Output
1161	- Practice consistently and regularly: Consistent practice is essential to learning any musical
1162	instrument, and it’s even more important when learning online. Set aside a specific time each day
1163	or week to practice, and make sure to stick to it.
1164	- Find a good teacher or mentor: While it’s possible to learn some things on your own, working
1165	with a good teacher or mentor can help you improve faster and avoid developing bad habits. Look
1166	for someone who is experienced and knowledgeable in the type of music you want to learn.
1167	- Use technology to your advantage: There are many online tools and resources available that can
1168	help you learn an instrument. These include video lessons, online tutorials, and apps that can help
1169	you practice and track your progress. . . .

1169 Table 2: Comparison of context-aware and baseline mechanism outputs using 10 candidate sequences.
1170 Advertisers highlighted in bold, outputs truncated to conserve space.

1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187

C.3 EXPERIMENTAL EVALUATION OF THE USEFULNESS TO THE USER

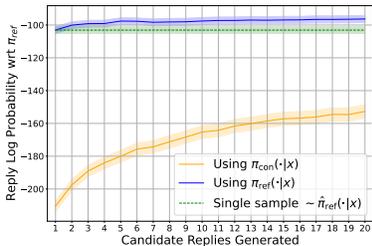


Figure 3: Reply log probability with respect to the reference LLM as a function of the number of replies generated using π_{ref} and π_{con} .

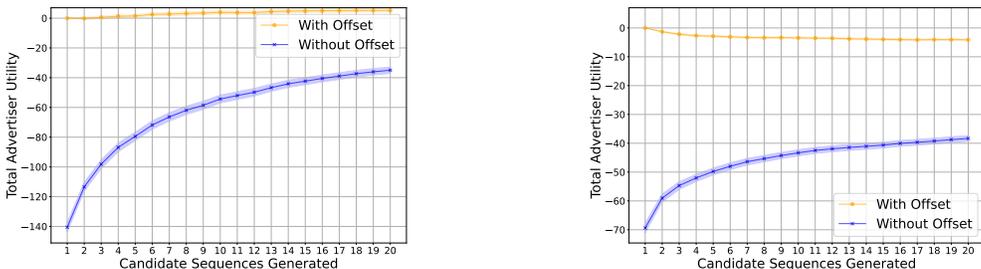
In this section, we experimentally evaluate the usefulness of the replies returned by MOSAIC from the user’s perspective.

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

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

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.

C.4 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.

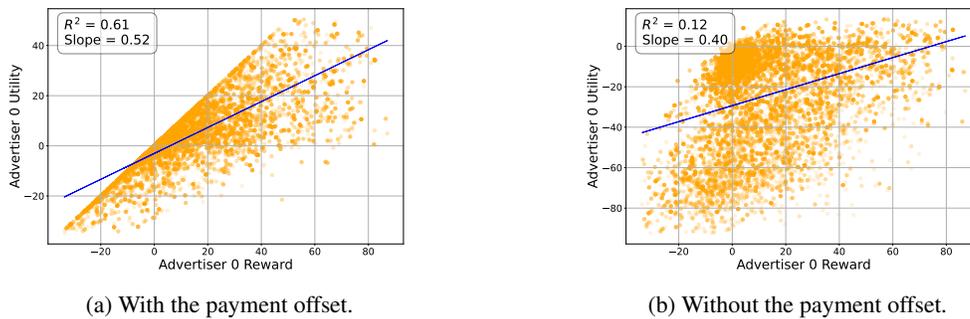


Figure 5: Comparative scatter plots of advertiser reward and utility gain from participation, with and without the payment offset of Section 5.3 for candidate sequences generated by the context-aware LLM π_{gem} . We additionally show a linear regressor fit to that data, its slope and its R^2 .

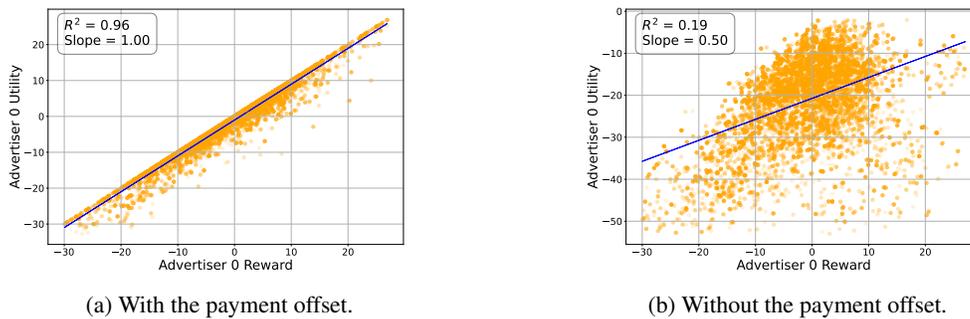


Figure 6: Comparative scatter plots of advertiser reward and utility gain from participation, with and without the payment offset of Section 5.3 for candidate sequences generated by the reference π_{ref} . We additionally show a linear regressor fit to that data, its slope and its R^2 .

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,

1296
 1297
 1298
 1299
 1300
 1301
 1302
 1303
 1304
 1305
 1306
 1307
 1308
 1309
 1310
 1311
 1312
 1313
 1314
 1315
 1316
 1317
 1318
 1319
 1320
 1321
 1322
 1323
 1324
 1325
 1326
 1327
 1328
 1329
 1330
 1331
 1332
 1333
 1334
 1335
 1336
 1337
 1338
 1339
 1340
 1341
 1342
 1343
 1344
 1345
 1346
 1347
 1348
 1349

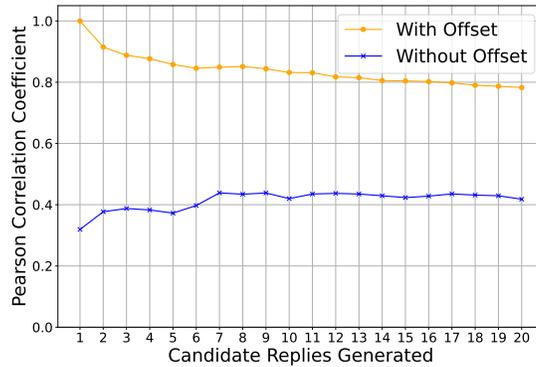


Figure 7: Pearson correlation between advertiser normalized utility and reward.

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 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 ADDITIONAL INFORMATION FOR ICLR REVIEWERS

In this section we include additional information addressing certain points raised by the conference reviewers.

D.1 ADDITIONAL TECHNICAL CONSIDERATIONS

Output Constraints. The generation of candidate replies is performed by π_{gen} , which is controlled by the platform. Thus, MOSAIC allows the platform to incorporate any desired constraints such as maximum length on the output replies by adding those constraints to the candidate reply generation procedure.

D.2 BUDGET AND ROI CONSTRAINTS

MOSAIC supports both budget and ROI constraints. For budget constraints, the advertiser should only bid rewards up to her maximum budget. For ROI constraints, the advertiser should “bid shade” when reporting her reward function to ensure that her ROI constraint is satisfied. As in any other strategyproof mechanism, both of these misreports would lead to worse utility per auction instance for the advertiser compared to truthful bidding, because in MOSAIC it is a dominant strategy for an advertiser to report truthfully. However, the upshot for the advertiser is that she can extend her budget to more auction instances. Additionally, for MOSAIC we can prove a new version of the “what you give is what you get” property of Lemma 5.2 for the case when a bidder under-reports her rewards. Informally, if an advertiser’s under-reported rewards are still positive for all replies, then her utility when under-reporting will be positive, and higher than Lemma 5.2 would suggest. The formal lemma follows below:

Lemma D.1. *Let \vec{r}_i be the true rewards of advertiser i for the candidate replies, and \vec{r}'_i be her reported rewards, where $\vec{r}_{i,j} \geq \vec{r}'_{i,j}$ for all $j \in [M]$. Finally, let $U_i(\vec{r}'_i; \vec{r}_i, \vec{\beta}_{-i})$ be advertiser i ’s utility when her true rewards are \vec{r}_i and she reports \vec{r}'_i . Then, for advertiser i ’s utility under reports \vec{r}_i and \vec{r}' we have:*

$$U_i(\vec{r}_i; \vec{r}_i, \vec{\beta}_{-i}) \geq U_i(\vec{r}'_i; \vec{r}_i, \vec{\beta}_{-i}) \geq U_i(\vec{r}'_i; \vec{r}'_i, \vec{\beta}_{-i}). \quad (18)$$

In particular, this means that if $\vec{r}' \geq \vec{r}_i \geq \vec{0}$, then her utility will still be positive if she under-reports \vec{r}'_i instead of \vec{r}_i .

Proof. The first inequality follows immediately from the fact that MOSAIC is truthful, i.e., it is a dominant strategy for advertiser i to truthfully report her rewards \vec{r}_i . Thus, any misreport will lead to lower utility.

For the second inequality, note that both advertiser i ’s payments and the allocation rule only depend on the advertiser’s reports, not on her true preferences. Thus, both advertiser i ’s payments and the probability of returning each of the candidate replies under the interim allocation rule are the same when advertiser i reports \vec{r}'_i , regardless of whether her true rewards are \vec{r}'_i or \vec{r}_i .

But for advertiser i ’s reward in the two cases, we have:

$$\sum_{j=1}^M r_i(x, y_j) \pi_{\text{int}}(y_j | x; \vec{r}_{-i}, \vec{r}'_i)_j \geq \sum_{j=1}^M r'_i(x, y_j) \pi_{\text{int}}(y_j | x; \vec{r}_{-i}, \vec{r}'_i)_j \quad (19)$$

where the inequality follows because $r_i(x, y_j) \geq r'_i(x, y_j)$ for all $j \in [M]$.

Thus, when the advertiser has reward \vec{r}_i but reports $\vec{r}'_i \leq \vec{r}_i$, she has the same payment as if her true rewards were \vec{r}'_i , but she has higher expected reward for the outcome, and thus higher utility, i.e. $U_i(\vec{r}'_i; \vec{r}_i, \vec{\beta}_{-i}) \geq U_i(\vec{r}'_i; \vec{r}'_i, \vec{\beta}_{-i})$.

Finally, note that if an advertiser’s true rewards for all replies is zero, her utility is zero from Lemma 5.2. Thus, if $\vec{r}'_i \geq \vec{0}$, we have $U_i(\vec{r}'_i; \vec{r}'_i, \vec{\beta}_{-i}) \geq 0$. This concludes the proof. \square

D.3 CONFLICTS BETWEEN ADVERTISERS

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

In Figure 8 we show the fractions of advertisers mentioned by the context-aware version of MOSAIC as a function of the candidate number of replies used. We observe that initially, the fraction of advertisers mentioned is very high, but drops significantly once the mechanism has converged. The reason is precisely the fact that mentioning more advertisers in this case would cause either a significant drop in advertiser utility, or a drop in the usefulness of the reply to the user. By placing more weight towards the advertisers (i.e., decreasing the value of τ in Equation (1)), the platform can increase the number of advertisers shown when the mechanism has converged.

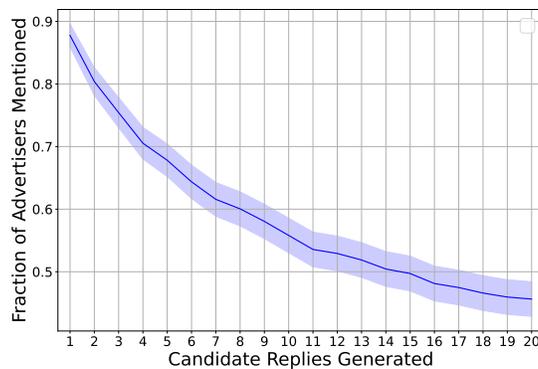


Figure 8: Fraction of advertisers mentioned in the output reply as a function of the number of candidate replies.

D.4 TESTING WITH ADDITIONAL LANGUAGE MODELS

In this section, we test MOSAIC’s performance using Google’s flan-t5-large model (Chung et al., 2022).

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

In Figure 9 to Figure 12, we present the total advertiser value gain, reward gain from participating in MOSAIC, platform revenue, and the usefulness of the returned reply to users as functions of the number of candidate replies generated.

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

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

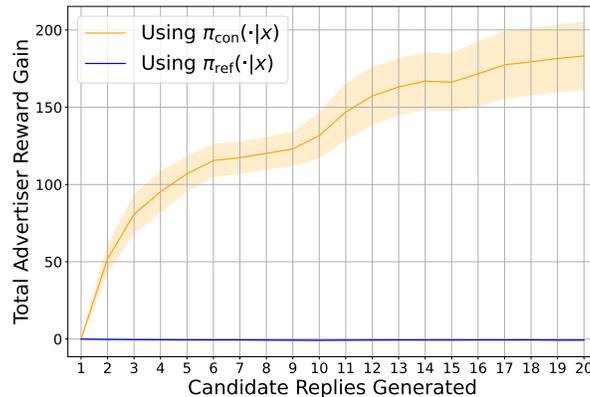
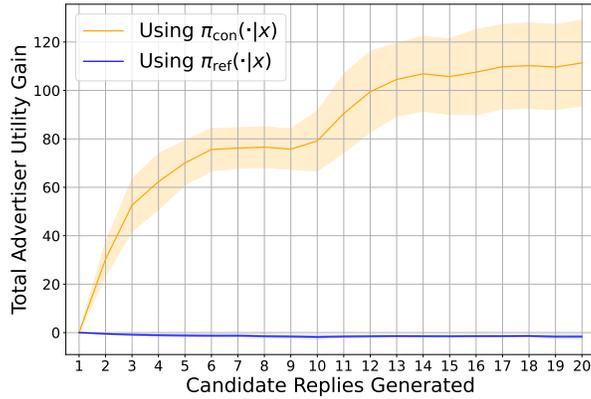


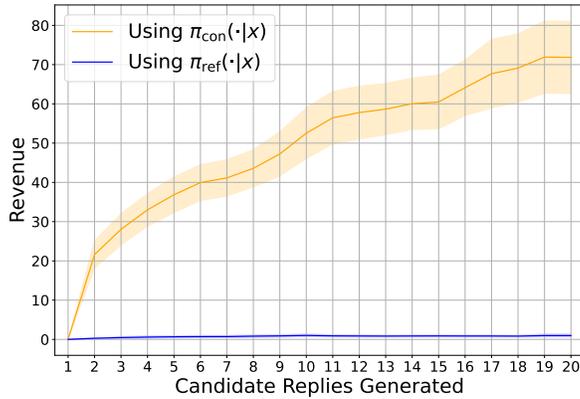
Figure 9: Advertiser value gain as a function of the number of candidate replies generated using π_{ref} and π_{con} (flan-t5-large model).

1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526



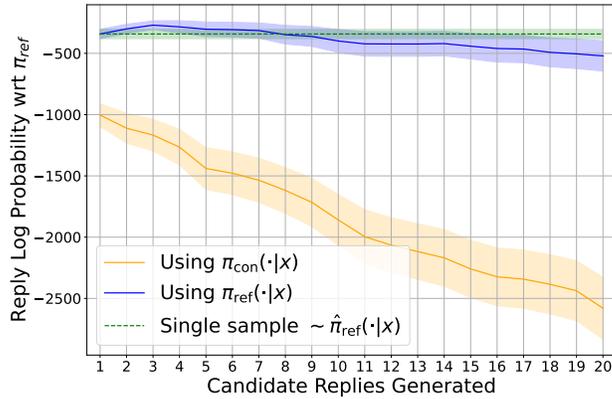
1527 Figure 10: Advertiser value gain as a function of the number of candidate replies generated using π_{ref}
1528 and π_{con} (flan-t5-large model).
1529

1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545



1546 Figure 11: Revenue as a function of the number of candidate replies generated using π_{ref} and π_{con}
1547 (flan-t5-large model).
1548

1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563



1564 Figure 12: Reply log probability with respect to the reference LLM as a function of the number of
1565 candidate replies generated using π_{ref} and π_{con} (flan-t5-large model).

1566 E FULL TEXT OF PROMPTS

1567

1568

1569

1570

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.

1571

1572

1573

1574

1575

User Query 1

Learning to play the guitar.

Advertisers

- StrumMaster: selling guitars, tuners, and learning materials for beginners

- Chordify: offering online guitar lessons and interactive practice tools

1576

1577

1578

1579

1580

User Query 2

Best skin care routine for dry skin.

Advertisers

- HydraSkin: providing moisturizers and serums for dry skin

- GlowEssentials: selling natural and organic skin care products

1581

1582

1583

1584

1585

User Query 3

Improving productivity while working from home.

Advertisers

- SpaceMakers: offering ergonomic furniture and home office setups

- FocusApp: providing productivity apps and time management tools

1586

1587

1588

1589

1590

User Query 4

Starting a small online business.

Advertisers

- EcomLaunch: offering e-commerce platform solutions and web design services

- MarketMover: providing digital marketing services and SEO optimization

1591

1592

1593

1594

1595

User Query 5

Healthy meal planning on a budget.

Advertisers

- BudgetBites: selling affordable meal kits and recipe books

- NutriSaver: offering discounts on healthy groceries and food delivery services

1596

1597

1598

1599

1600

User Query 6

Mastering digital photography.

Advertisers

- PixelPro: selling cameras, lenses, and photography accessories

- EditCraft: offering photo editing software and online tutorials

1601

1602

1603

1604

1605

User Query 7

Effective ways to reduce household energy use.

Advertisers

- EcoSave: offering energy-efficient home appliances and lighting solutions

- InsulaTech: providing home insulation and energy audit services

1606

1607

1608

1609

1610

User Query 8

Finding the perfect hiking trails.

Advertisers

- TrailFinder: offering a mobile app with detailed maps and trail reviews

- GearUp: selling outdoor gear and apparel for hiking enthusiasts

1611

1612

1613

1614

1615

User Query 9

Building a personal brand on social media.

Advertisers

- BrandBuilder: offering personal branding courses and social media strategy consultations

- VisualizeMe: providing graphic design services for social media content

1616

1617

1618

1619

User Query 10

Learning a new language effectively.

Advertisers

- LingoLeap: offering online language learning courses and tutoring

- 1620 - SpeakEasy: providing language learning apps with speech recognition technology
-
- 1621
- 1622 **User Query 11**
- 1623 Staying fit without a gym.
- 1624 **Advertisers**
- 1625 - HomeFit: selling home workout equipment and fitness accessories
- 1626 - MoveIt: offering online fitness classes and personal training sessions
-
- 1627 **User Query 12**
- 1628 Eco-friendly travel options.
- 1629 **Advertisers**
- 1630 - GreenPath: offering eco-friendly travel packages and sustainable tourism experiences
- 1631 - EcoStay: providing listings for green hotels and accommodations
-
- 1632 **User Query 13**
- 1633 Mastering the art of cooking steak.
- 1634 **Advertisers**
- 1635 - GrillMaster: selling premium grills and barbecue accessories
- 1636 - SteakPerfection: offering online cooking classes focused on meat preparation
-
- 1637 **User Query 14**
- 1638 Creating a successful YouTube channel.
- 1639 **Advertisers**
- 1640 - VidGrowth: offering video production courses and YouTube growth strategies
- 1641 - ChannelDesign: providing custom YouTube channel art and video thumbnails
-
- 1642 **User Query 15**
- 1643 Decorating your home on a budget.
- 1644 **Advertisers**
- 1645 - DecorDeals: selling affordable home decor and furniture
- 1646 - StyleSavvy: offering interior design consultations and budget-friendly decorating tips
-
- 1647 **User Query 16**
- 1648 Managing stress and anxiety.
- 1649 **Advertisers**
- 1650 - CalmSpace: offering mindfulness apps and stress reduction tools
- 1651 - WellnessWave: providing online therapy sessions and mental health resources
-
- 1652 **User Query 17**
- 1653 How can I make my home more energy efficient?
- 1654 **Advertisers**
- 1655 - EcoWise: offering smart home energy solutions and solar panels
- 1656 - GreenInsulate: providing eco-friendly insulation materials and installation services
-
- 1657 **User Query 18**
- 1658 What are the best strategies for investing in the stock market?
- 1659 **Advertisers**
- 1660 - InvestSmart: offering investment advice and financial planning services
- 1661 - StockGenius: providing an AI-powered stock market analysis tool
-
- 1662 **User Query 19**
- 1663 How can I improve my mental health?
- 1664 **Advertisers**
- 1665 - MindBalance: offering mental health apps and mindfulness courses
- 1666 - SerenityNow: providing online counseling and therapy sessions
-
- 1667 **User Query 20**
- 1668 Where can I find organic and sustainable clothing?
- 1669 **Advertisers**
- 1670 - EcoWear: selling sustainable and organic clothing options
- 1671 - GreenThreads: offering eco-friendly fabrics and fair-trade fashion
-
- 1672 **User Query 21**
- 1673 How do I start a vegetable garden in my backyard?

1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727

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
-

- 1728 **User Query 32**
1729 What are the best ways to improve indoor air quality?
1730 **Advertisers**
1731 - AirPure: selling air purifiers and natural air freshening solutions
1732 - BreatheEasy: offering indoor plant subscriptions and ventilation solutions
-
- 1733 **User Query 33**
1734 How to make my own natural beauty products?
1735 **Advertisers**
1736 - NatureGlow: offering DIY beauty product kits and organic ingredients
1737 - EcoBeauty: selling guides and supplies for creating homemade natural cosmetics
-
- 1738 **User Query 34**
1739 What are the best techniques for effective online teaching?
1740 **Advertisers**
1741 - EduTech: offering online teaching tools and educational technology solutions
1742 - TeachMaster: providing courses on online pedagogy and digital classroom management
-
- 1743 **User Query 35**
1744 How to create a successful podcast from home?
1745 **Advertisers**
1746 - PodCastify: offering podcast creation courses and audio equipment
1747 - SoundWave: selling podcast editing software and hosting services
-
- 1748 **User Query 36**
1749 What are the most effective ways to market a small business locally?
1750 **Advertisers**
1751 - LocalBoost: offering local SEO services and marketing strategies for small businesses
1752 - MarketTown: providing targeted advertising solutions and local business networking opportunities
-
- 1753 **User Query 37**
1754 How can I optimize my home office for productivity?
1755 **Advertisers**
1756 - WorkZen: offering ergonomic office furniture and productivity tools
1757 - SpaceCraft: providing home office design services and organization solutions
-
- 1758 **User Query 38**
1759 What are the best platforms for freelance writers to find work?
1760 **Advertisers**
1761 - WriteFreely: offering a job board and resources for freelance writers
1762 - ContentConnect: providing a platform to connect writers with businesses
-
- 1763 **User Query 39**
1764 How to get started with meditation and mindfulness?
1765 **Advertisers**
1766 - MindfulStart: offering guided meditation apps and mindfulness training
1767 - ZenSpace: selling meditation cushions and mindfulness journals
-
- 1768 **User Query 40**
1769 Where can I find unique handmade gifts?
1770 **Advertisers**
1771 - CraftCorner: selling handmade gifts and artisan crafts
1772 - MakerMarket: offering a marketplace for unique handmade items from local artists
-
- 1773 **User Query 41**
1774 How can I improve my credit score quickly?
1775 **Advertisers**
1776 - CreditBoost: offering credit monitoring services and financial advice
1777 - ScoreSurge: providing tools and tips for improving your credit score
-
- 1778 **User Query 42**
1779 What are the best apps for tracking fitness and nutrition?
1780 **Advertisers**
1781 - FitTrack: offering comprehensive fitness and nutrition tracking apps

1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835

- 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
-