FROM CONVERSATIONS TO MECHANISMS: ALIGNING ADVERTISER INCENTIVES IN AI-POWERED PRODUCT RECOMMENDATIONS

By

Dirk Bergemann, Marek Bojko, Paul Dütting, Renato Paes Leme, Haifeng Xu and Song Zuo

April 2026

COWLES FOUNDATION DISCUSSION PAPER NO. 2513

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

From Conversations to Mechanisms: Aligning Advertiser Incentives in AI-Powered Product Recommendations

Dirk Bergemann* Marek Bojko† Paul Dütting‡

Renato Paes Leme§ Haifeng Xu¶ Song Zuo‖

April 3, 2026

Abstract

We study the design of efficient dynamic recommendation systems, such as AI shopping assistants, in which a platform interacts with a user over multiple rounds to identify the most suitable product among those offered by advertisers. Advertisers have multi-dimensional private information: their private value from a purchase and private information about the user's preferences. In each round, the platform displays recommendations; the user learns product characteristics of the shown items and then chooses whether to purchase, exit without purchasing, or submit a new query. These actions generate a stream of feedback—purchase, exit, and follow-up queries—that is informative about the user's preferences and can be used both to refine future recommendations and to design contingent transfers. We introduce a class of data-driven dynamic team mechanisms that condition payments on realized user feedback. Our main result shows that data-driven dynamic team mechanisms achieve periodic ex-post implementation of the efficient allocation rule. We then develop variants that guarantee participation and deliver budget surplus, and provide conditions under which these properties can be jointly attained.

*Department of Economics, Yale University, dirk.bergemann@yale.edu

†Department of Economics, Yale University, marek.bojko@yale.edu

‡Google Research, duetting@google.com

§Google Research, renatoppl@google.com

¶Department of Computer Science, University of Chicago and Google Research, haifengxu@uchicago.edu

‖Google Research, szuo@google.com

1 Introduction

1.1 Motivation

The rise of artificial intelligence (AI) is fundamentally reshaping how consumers discover and evaluate products online. Conversational recommendation systems, such as AI-powered shopping assistants, are emerging as a new interface between consumers and digital marketplaces, offering an experience that goes far beyond traditional keyword-based search. Major platforms have already deployed such assistants, including Amazon's Rufus (Amazon, 2025; Mehta, 2024) and Google's AI-assisted shopping features (Google, 2025). OpenAI launched ChatGPT's Instant Checkout in 2025 (OpenAI, 2025), enabling direct purchases within conversational AI interfaces.[1] Rather than presenting a static list of sponsored links, these systems engage users over multiple rounds: the platform proposes options, the user inspects product characteristics, and subsequent responses—purchases, feedback, follow-up queries, or exit—shape the next round of recommendations. This dynamic interaction fundamentally changes the economics of attention allocation and creates new challenges for mechanism design.

In traditional online advertising, platforms auction users' attention to advertisers through mechanisms such as the generalized second-price auction (Edelman et al., 2007; Varian, 2007). The rise of AI shopping assistants introduces qualitatively new features that these classical frameworks do not address. First, the interaction is inherently *dynamic*: the user's choice set evolves over multiple rounds as recommendations are made and refined based on observed behavior. Second, advertisers possess *multidimensional private information*: not only their willingness to pay for a conversion but also proprietary signals about user preferences gleaned from their own data sources. Third, the interaction itself generates a rich stream of *behavioral data*, such as clicks, queries, dwell times, and purchase decisions, that progressively reveals the payoff-relevant state (the user's underlying preferences) that was initially unknown to all parties.

These features create a fundamental tension in the design of efficient recommendation systems that seek to maximize total surplus for advertisers and the user.[2] Efficiency requires aggregating dispersed information, as advertisers may possess valuable signals about user preferences from past interactions or third-party data. At the same time, advertisers have incentives to misreport this information to bias recommendations in their favor. Classical impossibility results by Jehiel et al. (2006) show that with multidimensional information and interdependent values, efficient implementation is generically impossible using standard message-driven mechanisms, in which both allocations and transfers depend solely on agents' reports. This raises a central question: how can a platform dynamically elicit truthful information from strategic advertisers while aggregating advertiser and platform information, as well as user preferences, to maximize total social welfare?

Our key insight is that the dynamic nature of conversational AI provides a natural way to overcome this impossibility. Behavioral signals generated during the interaction, such as

[1] OpenAI also introduced The Agentic Commerce Protocol, which allows interested merchants and developers to build their own integrations with the systems.

[2] While our main model emphasizes product recommendations that maximize joint surplus between advertisers and the user, an alternative interpretation is that the platform first identifies the product that best matches the user's needs and then selects among competing sellers for that product, who differ in prices, fees, or delivery terms. See Section 6 for further discussion.

purchases, exits, follow-up queries, and other forms of feedback, are informative about the user's preferences and can be incorporated into the payment scheme. By conditioning transfers on realized user feedback, the platform can exploit information revealed *after* the recommendation decision to better align advertisers' incentives with social efficiency. Building on the data-driven approach to mechanism design developed by Bergemann et al. (2025) for static environments, we show how the sequential evolution of the conversation and the resulting feedback data can be leveraged to implement efficient outcomes in dynamic settings.

1.2 Framework and Results

We study a dynamic recommendation problem in which a platform interacts with a user over multiple rounds, selecting products from a set of n advertisers to recommend. In each period, the platform makes a recommendation, the user observes the product's characteristics, and then chooses whether to purchase, exit, or continue with a follow-up query according to a known action rule that depends on the history of recommended products, the current recommendation, and the user's taste. Advertisers have two-dimensional private types: a *preference type*, reflecting their value from a purchase, and a *signal*, capturing their private information about the user's underlying taste parameter. We allow both components of private information to evolve dynamically as the interaction unfolds. The platform's objective is to dynamically aggregate advertisers' values with user preferences to maximize expected total surplus.

We seek to implement the efficient recommendation policy in *periodic ex-post equilibrium*, following Bergemann and Välimäki (2010), Athey and Segal (2013), and Liu (2018). This requires truthful reporting to be optimal for each agent after every history—comprising past recommendations, user actions, the evolution of the conversation, and any additional information acquired by advertisers—regardless of other agents' private histories, provided others report truthfully. This equilibrium concept is well suited to our setting, as it does not require agents to form beliefs about others' private information or to engage in complex Bayesian reasoning over time. It is also robust to inefficient information acquisition by advertisers, who have no incentive to engage in costly learning about other advertisers' types or the platform's proprietary conversational data.

Our main contribution is the introduction of *data-driven dynamic team mechanisms*. These mechanisms combine the construction of dynamic team mechanisms of Athey and Segal (2013) with data-driven Vickrey-Clarke-Groves (VCG) mechanisms of Bergemann et al. (2025) developed for static settings to our dynamic setting with interdependent values by conditioning payments on realized user feedback and overall outcomes.[3] Specifically, each advertiser receives a transfer that depends on an estimate of the user's taste parameter constructed from signals observed in the continuation of the conversation. The key insight is that when user feedback provides an unbiased signal of the underlying preference state, conditioning payments on this feedback allows the mechanism to ensure each agent becomes the residual claimant of the expected continuation social surplus under the posterior beliefs obtained by combining the information held by all advertisers and the platform. Our main implementation result, Theorem

[3]These contributions, in turn, build on classical papers by Vickrey (1961); Clarke (1971); Groves (1973); Holmström (1979).

1, shows that data-driven dynamic team mechanisms achieve periodic ex-post implementation of the efficient allocation rule.

We also study variants of our mechanisms that ensure participation and weak budget surplus at the outset of the interaction. First, in Proposition 1, we show how to modify transfers to guarantee *ex-ante* individual rationality and no subsidy: before advertisers receive private information about user tastes, their expected payoff from participation exceeds that of their outside option, and the mechanism does not pay agents more in expectation than their contribution to user welfare. This is achieved through an appropriate choice of the "free term" in the Groves-type transfer, analogous to the pivot mechanism in static settings (Clarke, 1971). Second, we construct pivotal adjustments to the data-driven dynamic team mechanism that ensure participation and budget surplus at the outset *ex-post* with respect to advertisers' initial types and the platform's initial context (Proposition 2). Finally, Corollary 1 shows that when the platform's initial context is sufficient for the unknown user taste relative to the information held by advertisers, participation and budget surplus can be jointly satisfied ex-post at the outset of the interaction.

1.3 Related Literature

This paper contributes to several strands of the mechanism design literature, including efficiency in dynamic mechanism design, mechanism design with interdependent values, advertising auctions, and the emerging literature on auctions and mechanism design with applications to generative AI.

Dynamic Mechanism Design. The literature on efficient dynamic mechanism design has developed powerful tools for implementing efficient allocations over time when agents receive private information dynamically. Foundational contributions by Bergemann and Välimäki (2010) and Athey and Segal (2013) extend VCG mechanisms to dynamic environments through the dynamic pivot mechanism and the team mechanism, respectively, in settings with private values and independent types. Bergemann and Välimäki (2019) provide a comprehensive survey of this literature. Our work extends these frameworks to environments with interdependent values, where agents hold private information not only about their own preferences but also about a common payoff-relevant state.

Efficiency in Mechanism Design with Interdependent Values. The literature on efficient implementation in settings with interdependent values and multi-dimensional types contains several well-known impossibility results when both the allocation and transfers condition only on messages reported by the agents to the mechanism (Maskin, 1992; Dasgupta and Maskin, 2000; Jehiel and Moldovanu, 2001; Jehiel et al., 2006).

To overcome the implementation impossibility, several papers propose contingent transfer schemes. Riordan and Sappington (1988) showed how a principal can implement the first-best in a setting with an ex-post signal correlated with the agent's private type using a lottery mechanism akin to Crémer and McLean (1988) surplus extraction mechanisms. Mezzetti (2004) showed that these impossibilities can be circumvented by employing two-stage mechanisms, whereby agents report their private types in the first stage, based on which the allocation is determined. Agents are then assumed to observe their ex-post payoffs and report them in the second stage. Finally, transfers are determined using a generalized VCG scheme based on the reported utilities.

Liu (2018) develops efficient mechanisms for dynamic environments with interdependent values. The core results show how a designer can exploit correlations between an agent's current type and other agents' future types to construct lottery payments that replicate dynamic team payments in expectation. The paper also identifies unbiased estimation of agents' ex-post payoffs as a sufficient condition for implementation. We propose a related mechanism in our applied setting by focusing on estimating the payoff of a non-strategic user, using the natural flow of information generated during the conversation. We show how we can utilize such estimates and other realized outcomes to align incentives.

Bergemann et al. (2025) recently introduced the concept of data-driven mechanism design in a static setting, showing how conditioning payments on additional information, such as estimated click-through rates in online advertising auctions, can achieve implementation when standard message-driven mechanisms fail. Our paper extends this framework to dynamic environments and demonstrates how the sequential revelation of information through user behavior can be leveraged to maintain incentive compatibility over multiple rounds.

Advertising Auctions and Digital Markets. The economics of online advertising has been extensively studied since the seminal analyses of generalized second-price auctions by Edelman et al. (2007) and Varian (2007). More recent work examines topics such as auto-bidding (Balseiro et al., 2021), reserve price optimization (Ostrovsky and Schwarz, 2023), and the welfare properties of alternative auction formats in advertising markets. The rise of AI-generated content and conversational interfaces has spurred new research on mechanism design for large language models (Dütting et al., 2024), token auctions for coordinating multiple LLM agents, and advertising within AI-generated content (Feizi et al., 2025; Soumalias et al., 2024; Dubey et al., 2024; Hajiaghayi et al., 2024). Banchio et al. (2024) study the optimal timing of ad placement for conversational AI assistants.

Our work complements this emerging literature by developing a dynamic mechanism design framework for AI shopping assistants that accounts for both the sequential nature of interaction and informational externalities across advertisers. In particular, we focus on efficient product recommendations in environments where advertisers hold private information about user preferences.

Recommender Systems and Platform Economics. There is a growing literature on the economics of recommender systems, examining their effects on consumer search (Dinerstein et al., 2018), supplier competition (Fletcher et al., 2023), and market concentration (Calvano et al., 2025). The algorithmic design of recommendation systems has implications for welfare that depend critically on whether systems are designed to be "customer-centric" or to serve platform objectives (De Corniere and Taylor, 2019; Hagiu and Jullien, 2011). Our model provides a framework for thinking about how to align the incentives of platforms, advertisers, and consumers in dynamic recommendation environments.

Consumer Search and Information Design. Finally, our paper connects to the literature on consumer search (Weitzman, 1979; Armstrong and Zhou, 2011) and information design (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019). The sequential nature of conversational AI naturally creates a search problem for users, while the platform's control over recommendations can be viewed through the lens of information design. Our framework bridges these literatures by explicitly modeling the dynamic interaction between information revelation through recommendations and information acquisition through user feedback.

1.4 Outline

The remainder of the paper is organized as follows. Section 2 introduces a simplified version of the dynamic model to illustrate the construction of our mechanisms. Section 3 presents the full dynamic model. Section 4 defines data-driven dynamic team mechanisms, introduces our implementation concept, and states the main implementation result. Section 5 studies participation and budget properties. Section 6 discusses limitations and open questions, and concludes.

2 Illustrative Example

In this section, we first present a simplified static version of the model and then extend it to two periods to highlight the paper's main conceptual contributions in a simple setting. The discussion is intentionally informal to streamline the exposition. A formal dynamic mechanism design model is introduced in the next section.

A one-shot environment. A user seeks a product and queries an AI shopping assistant. Two advertisers offer differentiated products described by characteristic vectors $\gamma_i \in \mathbb{R}^k$, $i = 1, 2$. The outside option has characteristic vector normalized to zero. The user's taste is summarized by $\omega \in \Omega \subseteq \mathbb{R}^k$, and her consideration set initially contains only the outside option. A recommendation adds a product to the consideration set, after which the user observes its characteristics. Advertiser i obtains payoff θ_i if her product is purchased, where θ_i is private information. Advertisers and the platform share a common prior $F \in \Delta(\Omega)$ over ω. Each advertiser i and the platform observe noisy signals

$$\mathbf{S}_i = \omega + \epsilon_i, \qquad \mathbf{S}_p = \omega + \epsilon_p,$$

where noise terms are potentially correlated. Let us collect the noise term in a profile ϵ.

In a single round, the platform recommends one product and the user either purchases or exits. The user purchases product i if and only if

$$\langle \gamma_i, \omega \rangle \geq 0.$$

Let

$$y_i(\omega) = \mathbb{1}_{\langle \gamma_i, \omega \rangle \geq 0}$$

denote this decision rule, and also the realized purchase outcome.

The platform chooses recommendation probabilities $x = (x_1, x_2)$ to maximize expected total surplus,

$$\max_x \sum_{i=1}^{2} \mathbb{E}\big[(\theta_i + \langle \gamma_i, \omega \rangle) y_i(\omega) x_i \mid \mathbf{S} = s\big],$$

where $\mathbf{S}$ collects all signals. Let x^* denote the efficient (deterministic) rule.

After the recommendation, user feedback generates an estimator

$$\widehat{\omega} = \omega + \epsilon_{p1}(x, y),$$

observed by the platform and verifiable by advertisers before transfers, where ϵ_{p1} is mean zero and independent of ϵ. That is, $\widehat{\omega}$ is an unbiased estimator,

$$\mathbb{E}[\widehat{\omega} \mid \omega, x, y] = \omega,$$

and it is independent of agents' signal profile $\mathbf{S}$ conditionally on the payoff-relevant state ω.

To implement the efficient allocation, we use the *data-driven VCG* mechanism introduced by Bergemann et al. (2025). Fix an advertiser i and let $j \neq i$. The mechanism uses the efficient allocation rule and assigns payment

$$p_i = -\langle \gamma_i, \widehat{\omega} \rangle y_i - (\theta_j + \langle \gamma_j, \widehat{\omega} \rangle) y_j + \phi_i(\theta_j, \mathbf{S}_j, s_j)$$

to advertiser i, for an arbitrary function ϕ_i. Theorem 1 of Bergemann et al. (2025) implies that truthful reporting constitutes an ex-post equilibrium. To see this directly, the expected payoff of agent i under a report $(\theta_i', \mathbf{S}_i', s_i')$ when the true profile is $(\theta, \mathbf{S}, s)$ equals, up to an additive term independent of the agent's report,

$$\sum_{j=1}^{2} \mathbb{E}\Big[(\theta_j + \langle \gamma_j, \omega \rangle) y_j(\omega)\, x_j^*(\theta_i', \theta_{-i}, \mathbf{S}_i', \mathbf{S}_{-i}, s_i', s_{-i}) \;\Big|\; \mathbf{S} = s\Big],$$

which is maximized by truthful reporting because x^* maximizes expected total surplus. Hence, the efficient allocation is implemented. The free terms ϕ_i can be used to ensure participation and budget properties. Consider the pivotal-style payments

$$\phi_i^+ = \max_{s_i \in \mathcal{S}_i} \mathbb{E}\Big[(\theta_j + \langle \gamma_j, \widehat{\omega} \rangle)\, y_j(\omega) \;\Big|\; \mathbf{S}_i = s_i, \mathbf{S}_{-i} = s_{-i}\Big],$$

and

$$\phi_i^- = \min_{s_i \in \mathcal{S}_i} \mathbb{E}\Big[(\theta_j + \langle \gamma_j, \widehat{\omega} \rangle)\, y_j(\omega) \;\Big|\; \mathbf{S}_i = s_i, \mathbf{S}_{-i} = s_{-i}\Big].$$

Under ϕ^+, advertiser i's participation constraint holds in expectation given the combined information held by the advertisers and the platform, while under ϕ^- she never receives more than her marginal contribution to user surplus in expectation. When the information of advertiser i is redundant, the two coincide and both properties hold simultaneously. An ex-ante version,

$$\phi_i = \mathbb{E}\Big[(\theta_j + \langle \gamma_j, \widehat{\omega} \rangle)\, y_j(\omega)\Big],$$

ensures ex-ante participation and no subsidy simultaneously. We show these statements formally in the context of our dynamic model in Section 5.

Two rounds. Suppose now the user may submit a follow-up query after the first recommendation. Let

$$y_q(i, \omega) \in (0, 1)$$

denote the probability that the user continues with a follow-up query after being recommended product i. We assume that the user is less likely to continue searching when the recommended product better matches her preferences. Formally,

$$\langle \gamma_i, \omega \rangle \geq \langle \gamma_j, \omega \rangle \;\Rightarrow\; y_q(i, \omega) \leq y_q(j, \omega),$$

so products that yield higher utility to the user lead to a lower probability of further search.

A purchase in period 1 occurs with probability

$$(1 - y_q(i, \omega))\mathbb{1}_{\langle \gamma_i, \omega \rangle \geq 0},$$

while otherwise either the interaction continues or the user selects the outside option. In period 2, after observing both products, the user selects her preferred option or the outside alternative. Purchase or exit are terminal actions; if either occurs in period 1, the interaction ends immediately.

After period 1, the platform observes an estimator $\widehat{\omega}_1$. If the interaction continues, a second estimator $\widehat{\omega}_2$ is generated at the end of period 2.

The platform's objective is to maximize expected total surplus across both periods. In this simple example, the second product will be recommended in period 2 whenever the interaction continues, independently of reports. Thus, the only allocative decision concerns which product to recommend first. Recommending product i in period 1 yields expected surplus

$$\mathbb{E}\Big[(\theta_i + \langle \gamma_i, \omega \rangle) y_i(i, \omega) + y_q(i, \omega)(\theta_j + \langle \gamma_j, \omega \rangle) y_j(j, \omega) \;\Big|\; \mathbf{S} = s\Big],$$

where $j \neq i$. The first term captures surplus if product i is purchased immediately, while the second term captures continuation surplus if the user proceeds to period 2 and then purchases product j. The monotonicity of y_q implies that better matches in period 1 reduce the likelihood of continuation but increase the likelihood of immediate trade. It follows that, since the second-period recommendation is fixed, the efficient first-period recommendation is the product i that maximizes the above expected total surplus conditional on signals $\mathbf{S} = s$.

A dynamic extension of the data-driven VCG mechanism, based on the dynamic team mechanism of Athey and Segal (2013), assigns in each period

$$p_{it} = -\langle \gamma_i, \widehat{\omega}_t \rangle y_{it} - (\theta_j + \langle \gamma_j, \widehat{\omega}_t \rangle) y_{jt} + \phi_{it}(\cdot),$$

so each agent is a residual claimant on realized flow surplus. (Periodic) ex-post implementation follows from unbiasedness by an analogous argument, which we develop in our general dynamic model in Theorem 1.

Toward the general model. The full model extends this example along several dimensions. We allow for an arbitrary finite set of advertisers and an arbitrary horizon of interaction between the user and the AI assistant. Over time, the platform gradually learns about the user's preferences from the sequence of recommendations, follow-up queries, purchases, and exits. Early rounds may therefore serve an exploratory role, as the platform trades off immediate surplus against the value of information for future recommendations.

At each point in the interaction, the user's choice set consists of the products recommended so far and the outside option. In every round, the user may purchase one of the available products, exit without purchase, or submit a follow-up query and continue the conversation. Each such action generates feedback data that are informative about the underlying taste parameter. The platform uses this evolving information both to improve subsequent recommendations and to construct estimators of user preferences that enter contingent transfers.

On the mechanism side, we generalize the data-driven dynamic team mechanisms introduced above. Allocations in each period depend only on the information available at the time of recommendation, while transfers may condition on the feedback observed over the course of the interaction. The formal analysis in the next section shows that this structure permits periodic ex-post implementation of the efficient dynamic recommendation policy in a rich class of environments.

3 Dynamic Model

3.1 Set-up

The mechanism design environment consists of a finite set of advertisers, $N \equiv \{1, \dots, n\}$, whom we also refer to as agents, and a (non-strategic) user who seeks to purchase a product. The platform interacts with both advertisers and the user over multiple rounds (time periods). Time is discrete and indexed by $t \in \{1, \dots, T\}$, where $T \leq \infty$.

Payoffs. Each advertiser $i \in N$ offers a single product that is fixed throughout the interaction. Product i is characterized by a finite-dimensional vector of observable attributes $\gamma_i \in \mathbb{R}^k$. The user's preferences are summarized by a taste parameter $\omega \in \Omega = \mathbb{R}^k_+$. If the user purchases product i, she derives utility

$$\langle \gamma_i, \omega \rangle.$$

We assume there is an outside option with characteristic vector $\gamma_0 = \mathbf{0} \in \mathbb{R}^k$, which yields zero utility. Hence, product utilities are interpreted relative to the outside option. Advertiser i receives a payoff $\theta_{it} \in \Theta_{it} \subseteq \mathbb{R}_+$ if the user purchases product i in period t, where Θ_{it} is measurable. Let us denote by $\Theta_t = \prod_{i \in N} \Theta_{it}$ the product space.

In each period $t \geq 1$, the platform chooses whether to recommend a product and, if so, which product to recommend. The set of feasible allocations is the simplex

$$X_t \equiv \Delta(N \cup \{0\}),$$

where x_{it} denotes the probability of recommending product $i \in N$ and x_{0t} denotes the probability of making no recommendation in period t. We denote the realized recommendation by

$$r_t \in R \equiv N \cup \{0\},$$

where $r_t = 0$ corresponds to no recommendation.

In each period $t \geq 1$, after observing the recommendation, the user chooses whether to purchase one of the products she has observed up to that point, permanently exit the platform, or make no purchase and continue with a follow-up query; these choices may be made probabilistically. The set of pure actions available to the user is

$$A \equiv N \cup \{0\} \cup \{q\},$$

where q denotes the action of submitting a follow-up query and 0 denotes exit. The corresponding set of mixed actions is

$$Y \equiv \Delta(A).$$

An outcome history o_t is the sequence of realized recommendations and user actions up to period t:

$$o_t = (r_1, a_1, \ldots, r_{t-1}, a_{t-1}, r_t, a_t) \in \mathcal{O}_t,$$

where the set of feasible histories is

$$\mathcal{O}_t \equiv R^t \times A^t.$$

We write $r^t \in R^t$ for the sequence of recommendations up to t and $a^t \in A^t$ for the corresponding sequence of user actions.

Given an action sequence $a^T \in A^T$, advertiser i's total ex-post payoff is

$$\sum_{t=1}^{T} \theta_{it} \, \mathbb{1}_{a_t = i},$$

and the user's total ex-post utility is

$$\sum_{i=1}^{n} \sum_{t=1}^{T} \langle \gamma_i, \omega \rangle \, \mathbb{1}_{a_t = i}.$$

User search behavior. We model the user's search behavior via a commonly known *action rule*. Given a past outcome history $o_{t-1} \in \mathcal{O}_{t-1}$, a current recommendation $r_t \in R$, and the user's taste parameter $\omega \in \Omega$, the user's (mixed) action is governed by a map

$$y_t : \mathcal{O}_{t-1} \times R \times \Omega \to Y.$$

We write $y_{it}(o_{t-1}, r_t, \omega)$ for the probability that the user purchases product i in period t, $y_{0t}(o_{t-1}, r_t, \omega)$ for the probability of exit, and $y_{qt}(o_{t-1}, r_t, \omega)$ for the probability of submitting a follow-up query. We impose the following minimal restrictions.

Assumption 1 (Properties of the user's action rule)**.** *The sequence* $\{y_t\}_{t=1}^{T}$ *satisfies:*

1. Purchase and exit are terminal actions. *With a slight abuse of notation,*

$$(N \cup \{0\}) \cap a^{t-1} \neq \emptyset \;\Rightarrow\; y_{0t}(o_{t-1}, r_t, \omega) = 1.$$

2. Choice among observed products. *The user may only purchase products that have been recommended so far:*

$$y_{it}(o_{t-1}, r_t, \omega) > 0 \;\Rightarrow\; i \in r^t.$$

3. Finite termination. *Define the (random) termination time*

$$\tau \equiv \inf\{t \geq 1 : a_t \in N \cup \{0\}\},$$

with the convention $\inf \emptyset = T$. *The interaction terminates almost surely:*

$$\mathbb{P}(\tau < \infty) = 1.$$

The assumption accommodates a wide range of user search behaviors. For example, the user may exhibit perfect recall, retaining all previously recommended products and ultimately selecting the most preferred option from the resulting consideration set. Alternatively, the user may exhibit imperfect recall, placing greater weight on more recently displayed products. An extreme case of this behavior is one in which, whenever a purchase occurs, the user selects only the most recently recommended product. Our framework allows for all of these interpretations.

Given an outcome history o_{t-1}, taste parameter ω, and allocation $x_t \in X_t$, the probability that product i is purchased in period t is

$$z_{it}(o_{t-1}, \omega, x_t) \equiv \sum_{r_t \in R} y_{it}(o_{t-1}, r_t, \omega)\, x_{r_t t}.$$

Accordingly, the expected flow payoff in period t is

$$\theta_{it} \cdot z_{it}(o_{t-1}, \omega, x_t)$$

for advertiser i, and

$$\sum_{i=1}^{n} \langle \gamma_i, \omega \rangle \cdot z_{it}(o_{t-1}, \omega, x_t)$$

for the user.

Information. We allow advertisers' values from purchase to evolve stochastically over time. There is a common prior $G_0 \in \Delta(\Theta_1)$. Further, there is a commonly known stochastic kernel

$$G_{it}(\cdot \mid o_t, \theta_{it}) \in \Delta(\Theta_{i,t+1})$$

that governs the evolution of advertiser i's value in the next period as a function of the current outcome history o_t and current value θ_{it}.

The profile of product characteristics $\gamma \equiv (\gamma_1, \ldots, \gamma_n)$ is commonly known among advertisers.

The user observes her taste parameter ω, whereas the advertisers and the platform are uncertain about it. At the outset, advertisers and the platform share a common prior

$$F_0 \in \Delta(\Omega)$$

over ω.

We allow both the platform and advertisers to possess private information about ω. Each advertiser i is endowed with an initial signal

$$\mathbf{S}_{i1} = \omega + \epsilon_{i1},$$

with realization $s_{i1} \in \mathcal{S}_{i1} \subseteq \mathbb{R}^k$, where the noise terms ϵ_{i1} may be correlated across advertisers. Such signals may represent, for example, third-party purchase data or information collected from interactions on the advertiser's own tool, interface, or website.

The platform may also hold private information about user tastes beyond the common prior, obtained for instance from past interactions with similar users or from contextual information at the start of the interaction. We represent this by an initial platform signal

$$\mathbf{S}_{p1} = \omega + \epsilon_{p1},$$

with realization $s_{p1} \in \mathcal{S}_{p1} \subseteq \mathbb{R}^k$, where ϵ_{p1} may be correlated with $\{\epsilon_{i1}\}_{i \in N}$. We denote by

$$\mathbf{S}_1 \equiv (\mathbf{S}_{p1}, \mathbf{S}_{11}, \ldots, \mathbf{S}_{n1})$$

the period-1 profile of signals, and by s_1 the corresponding profile of realizations; we follow this notational convention for each period t. We further denote by $\mathbf{S}_{-pt}$ the corresponding profile excluding the information held by the platform.

As the interaction progresses, additional information about ω is revealed through user behavior. At the beginning of each period t, the platform observes the user's response and obtains a signal

$$\mathbf{S}_{pt}(o_{t-1}) = \omega + \epsilon_{pt}(o_{t-1}),$$

for every outcome history $o_{t-1} \in \mathcal{O}_{t-1}$. We assume that $\epsilon_{pt}(o_{t-1})$ is independent across periods t and histories o_{t-1},[4] and satisfies

$$\mathbb{E}[\epsilon_{pt}(o_{t-1}) \mid \omega, o_{t-1}] = 0.$$

The distribution of $\epsilon_{pt}(o_{t-1})$ may depend on the outcome history. For instance, early responses may be more informative than later responses, which primarily refine previously acquired information. We assume that the signal in period t is obtained at the beginning of the period, once the previous round has concluded with an action of the user. The signal therefore represents feedback generated by the user's interaction with the recommendation. Importantly, we assume such feedback is obtained also after the final interaction with the user.

We also allow advertisers to obtain additional information about the user's preferences as the

[4] In particular, it is independent of the period-1 noise terms.

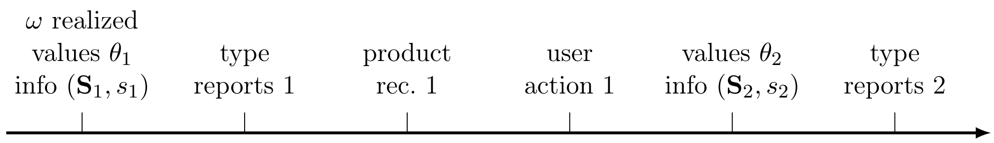

Figure 1: Timeline of the interaction with the user.

interaction unfolds. For example, after observing a recommendation, a user may engage with the advertiser's integrated tool or visit the advertiser's website before returning to the platform with a follow-up query, thereby generating further information for the advertiser. Formally, at the beginning of period t, given outcome history o_{t-1}, advertiser i observes a signal $\mathbf{S}_{it}(o_{t-1})$ with a structure analogous to the signals defined above, with realization $s_{it} \in \mathcal{S}_{it} \subseteq \mathbb{R}^k$. In particular, we assume $\epsilon_{it}(o_{t-1})$ may be correlated with $\epsilon_{pt}(o_{t-1})$. This formulation captures the preceding example as well as richer information structures, such as cases in which users revisit previously viewed products to make comparisons. We denote the corresponding profile of signals of the platform and advertisers by $\mathbf{S}_t$ and its realization by s_t.

Finally, we assume that the information structure (i.e., the underlying random variables) is commonly known, while the signal realizations s_{it} are the private information of advertiser i. By contrast, both the signals and realizations observed by the platform are publicly revealed.[5]

We summarize information up to period t by the information history

$$I_t = (\mathbf{S}_1, s_1, \ldots, \mathbf{S}_t, s_t),$$

and denote the corresponding space by $\mathcal{I}_t$. The public information history is

$$I_t^p = (\mathbf{S}_{p1}, s_{p1}, \ldots, \mathbf{S}_{pt}, s_{pt}),$$

with corresponding space $\mathcal{I}_t^p$. Given I_t and the prior F_0, agents form a posterior $F_t(\cdot \mid I_t)$.

Finally, the type space of advertiser i in period t is

$$\Xi_{it} \equiv \Theta_{it} \times \mathcal{S}_{it}.$$

Let $\Xi_t \equiv \prod_{i \in N} \Xi_{it}$ and $\Xi \equiv \prod_{t=1}^{T} \Xi_t$ be the corresponding product spaces.

The timeline of the interaction, including type reports discussed in the next section, is visualized in Figure 1. We summarize key notation in Table 1.

3.2 Efficiency

We focus on the efficient outcome, which dynamically aggregates agents' private preferences and information, while allowing for residual uncertainty about the payoff-relevant state ω.

[5]Below, we study periodic ex-post implementation, where agents evaluate expected payoffs with respect to the combined information held by advertisers and the platform. Moreover, we allow the platform to condition transfers on information revealed in future periods. Public revelation of platform information is without loss for our implementation objectives; Bergemann et al. (2025) discuss this point further.

Table 1: Summary of Notation

Symbol	**Description**
$i \in N = \{1, \ldots, n\}$	Set of advertisers (agents)
$t \in \{1, \ldots, T\}$	Time period (possibly infinite horizon)
$\gamma_i \in \mathbb{R}^k$	Observable characteristic vector of product i
$\omega \in \Omega \subseteq \mathbb{R}^k_+$	User taste parameter
$\theta_{it} \in \Theta_{it}$	Advertiser i's value from a purchase in period t
$x_t \in X_t = \Delta(N \cup \{0\})$	Set of feasible recommendation lotteries in period t
$r_t \in R = N \cup \{0\}$	Realized recommendation in period t
$a_t \in A_t = N \cup \{0\} \cup \{q\}$	User action set (purchase, exit, or follow-up query)
$o_t \in \mathcal{O}_t = R^t \times A^t$	Outcome history up to period t
$y_t(o_{t-1}, r_t, \omega)$	User action rule in period t
τ	(Random) termination time of the interaction
$\mathbf{S}_{it}$, s_{it}	Advertiser i's signal and realization in period t
$\mathbf{S}_{pt}$, s_{pt}	Platform signal and realization in period t
$I_t \in \mathcal{I}_t$	Full information history up to period t
$I^p_t \in \mathcal{I}^p_t$	Public information history up to period t
$\xi_{it} \in \Xi_{it} = \Theta_{it} \times \mathcal{S}_{it}$	Type space of advertiser i in period t
$m_t \in \mathcal{M}_t$	Report history up to and including period t
$h_t \in \mathcal{H}_t$	Public history up to period t

Definition 1 (Efficient allocation)**.** A sequence of allocation rules

$$\mathbf{x}^* \equiv \{x^*_t : \Theta_t \times \mathcal{O}_{t-1} \times \mathcal{I}_t \to X_t\}_{t=1}^T$$

is an efficient allocation rule if it solves

$$\max_{\{x_t\}_{t=1}^T} \mathbb{E}\left[\sum_{t=1}^T \sum_{i=1}^n (\theta_{it} + \langle \gamma_i, \omega \rangle) \sum_{r_t \in R} y_{it}(o_{t-1}, r_t, \omega)\, x_{r_t t}(\theta_t, o_{t-1}, I_t) \,\middle|\, I_1, \theta_1 \right].$$

The social program admits a recursive representation. Let $W_t(\theta_t, o_{t-1}, I_t)$ denote the maximal expected continuation surplus from period t onward given (θ, o_{t-1}, I_t), and set $W_{T+1} \equiv 0$. Then

$$W_t(\theta_t, o_{t-1}, I_t) = \max_{x_t \in X_t} \left\{ \mathbb{E}\left[\sum_{i=1}^n (\theta_{it} + \langle \gamma_i, \omega \rangle) z_{it}(o_{t-1}, \omega, x_t) + W_{t+1}(\theta_{t+1}, o_t, I_{t+1}) \,\middle|\, x_t, o_{t-1}, I_t \right] \right\}.$$

By Bellman's optimality principle, the solution sets of the recursive and original formulations coincide.

4 Implementation with Data-Driven Dynamic Team Mechanisms

In this section, we formulate the mechanism design problem and define our implementation concept. We then introduce the main class of mechanisms analyzed in this paper: the class of *data-driven dynamic team mechanisms*. Finally, we show that every mechanism in this class implements the efficient allocation.

4.1 Mechanisms and Implementation

Our goal is to implement the efficient allocation rule $\mathbf{x}^*$. We restrict attention to direct public mechanisms and focus on truth-telling equilibria. In each period t, each agent i submits a report

$$\xi'_{it} \in \Xi_{it}.$$

Based on the profile of reports $\xi'_t \in \Xi_t$, together with the history of reports, realized outcomes, and public information, the mechanism determines the allocation x_t and enforces transfers p_t.

The report history in period t consists of all reports submitted up to and including period t:

$$m_t = (\xi'_1, \dots, \xi'_t).$$

We denote the corresponding space of report histories by $\mathcal{M}_t$. The overall public history in period t is

$$h_t = o_{t-1} \cup I^p_t \cup m_{t-1},$$

with associated space of feasible histories $\mathcal{H}_t$.

Data-driven mechanisms. We now formally define the class of mechanisms we consider. Following the data-driven approach of Bergemann et al. (2025), we assume allocations may depend only on the information available at the beginning of period t, while transfers may additionally depend on information revealed after the allocation has been determined: both the realized outcomes and information about user tastes obtained as the conversation unfolds.

One interpretation is that the platform settles transfers at the end of the interaction with the user, in which case all period-t transfers are computed and enforced using the full information available at that time. Alternatively, the platform may enforce transfers at the beginning of the next round of interaction. In the exposition below, we adopt the latter convention: the platform observes new information at the beginning of each period and enforces transfers at that stage. Adapting our main results to the former timing assumption is straightforward.

Assumption 2 (Estimator)**.** *For each period $t \leq \tau$ and each outcome history $o_t \in \mathcal{O}_t$, the platform and advertisers observe the realization of a random variable $\widehat{\omega}(o_t) \in \Omega$, interpreted as an estimator of the state. The realization is observed after the period-t recommendation and user action (r_t, a_t) are realized (i.e., after o_t is formed), but before the period-t transfers are determined.*

Thus, while allocations depend only on reports and information available at the beginning of the period, transfers may additionally condition on the realized outcome in that period and on an estimate of the payoff-relevant state obtained after the outcome is realized.

Definition 2 (Data-driven direct revelation mechanism)**.** A data-driven direct revelation mechanism is a tuple $(\mathbf{x}, \mathbf{p})$, where

$$\mathbf{x} = \{x_t : \Xi_t \times \mathcal{H}_t \to X_t\}_{t=1}^T \quad \text{and} \quad \mathbf{p} = \{p_t : \Xi_t \times \mathcal{H}_t \times R \times A \times \Omega \to \mathbb{R}^n\}_{t=1}^T$$

denote the sequences of allocation and transfer functions, respectively.

As above, we compress notation and write

$$z_{it}(\xi_t, h_t, \omega) = \sum_{r_t \in R_t} y_{it}(o_{t-1}, r_t, \omega) \cdot x_{r_t t}(\xi_t, h_t)$$

for the probability that product i is purchased in period t.

Implementation concept. Following Bergemann and Välimäki (2010), Athey and Segal (2013), and Liu (2018), we focus on *periodic ex-post implementation* of the efficient policy $\mathbf{x}^*$. Under this notion, truthful reporting must be optimal for each agent after every history and for every realization of the agent's current type, provided all other agents report truthfully. Agents evaluate payoffs in expectation using the true type profile and beliefs formed from the full history of signals about the state. To define this concept formally, let

$$V_{it}(\xi_t, h_t, I_t) = \max_{\xi'_{it} \in \Xi_{it}} \Bigg\{ \mathbb{E}\Big[\theta_{it} \cdot z_{it}(\xi'_{it}, \xi_{-it}, h_t, \omega) \\ - p_{it}(\xi'_{it}, \xi_{-it}, h_t, r_t, a_t, \widehat{\omega}(o_t)) + V_{it+1}(\xi_{t+1}, h_{t+1}, I_{t+1}) \,\Big|\, x_t(\cdot), \xi_t, h_t, I_t\Big]\Bigg\}$$

denote the continuation value for agent i in period t.

Definition 3 (Periodic ex-post implementation)**.** A data-driven mechanism $(\mathbf{x}, \mathbf{p})$ permits periodic ex-post implementation if, for all $i \in N$, $t \geq 1$, $h_t \in \mathcal{H}_t$, $\xi_t \in \Xi_t$, and $I_t \in \mathcal{I}_t$,

$$\xi_{it} \in \arg\max_{\xi'_{it} \in \Xi_{it}} \Bigg\{ \mathbb{E}\Big[\theta_{it} \cdot z_{it}(\xi'_{it}, \xi_{-it}, h_t, \omega) \\ - p_{it}(\xi'_{it}, \xi_{-it}, h_t, r_t, a_t, \widehat{\omega}(o_t)) + V_{it+1}(\xi_{t+1}, h_{t+1}, I_{t+1}) \,\Big|\, x_t(\cdot), \xi_t, h_t, I_t\Big]\Bigg\}.$$

4.2 Data-Driven Dynamic Team Mechanisms

To achieve periodic ex-post implementation of the efficient allocation, we consider the following class of mechanisms.

Definition 4 (Data-driven dynamic team mechanism)**.** A direct revelation mechanism $(\mathbf{x}^*, \mathbf{p})$ is a *data-driven dynamic team mechanism* if, for every agent $i \in N$, period t, report profile $\xi_t \in \Xi_t$, and public history $h_t \in \mathcal{H}_t$, transfers take the form

$$p_{it}(\xi_t, h_t, r_t, a_t, \widehat{\omega}) = -\langle \gamma_i, \widehat{\omega} \rangle \mathbb{1}_{a_t = i} - \sum_{j \neq i} (\theta_{jt} + \langle \gamma_j, \widehat{\omega} \rangle) \mathbb{1}_{a_t = j} + \phi_{it}(\xi_{-it}),$$

for some arbitrary function ϕ_{it}.

Observe that transfers in the data-driven dynamic team mechanism take a particularly simple form. An estimate of the user's taste parameter is substituted into the transfer formula, and the core component of the payment is nonzero if and only if a purchase occurs. The Groves-type term ϕ_{it} can then be used to ensure individual rationality and revenue properties, which we analyze in the next section.

At this point, we have not specified how the estimator of the state is constructed, a question to which we now turn. In particular, realizations of the conversation in future periods can be used to align incentives. Consider the realization of the feedback signal of the platform in the next period:

$$\widehat{\omega} = s_{pt+1},$$

where s_{pt+1} is the realization of $\mathbf{S}_{pt+1}(o_t)$.[6] Accordingly, the estimator is given by

$$\widehat{\omega}(o_t) = \mathbf{S}_{pt+1}(o_t).$$

By construction of the signal structure, such estimators satisfy two key properties. First, they are conditionally independent of the prior information history I_t given the outcome history o_t and the true state ω. Second, they are unbiased:

$$\mathbb{E}[\widehat{\omega}(o_t) \mid o_t, \omega] = \omega$$

for all $o_t \in \mathcal{O}_t$ and $\omega \in \Omega$. Our main result below holds for any estimator satisfying these two properties.

Importantly, the platform need not acquire any external data: future realizations of the conversation within the current environment are sufficient to align incentives. Moreover, the component of transfers that depends on the estimate of the state ω is nonzero only when a purchase occurs. If we assume that one additional round of feedback follows a purchase before the user exits the platform, conditioning the estimator on this feedback exploits the full amount of feedback information available within the current conversation for incentive alignment.

Assumption 3 (Unbiased estimator)**.** *The sequence of estimators $\widehat{\boldsymbol{\omega}} = \{\widehat{\omega}(o_t)\}_{o_t \in \mathcal{O}_t,\, t \geq 1}$ introduced in Assumption 2 satisfies:*

1. *For any t, outcome history o_t, and information history I_t, the estimator $\widehat{\omega}(o_t)$ is conditionally independent of I_t given o_t and ω; and*
2. *For any t, outcome history o_t, and state ω, $\mathbb{E}[\widehat{\omega}(o_t) \mid o_t, \omega] = \omega$.*

Theorem 1 (Implementation with the data-driven dynamic team mechanism)**.** *Under Assumptions 1-3, every data-driven dynamic team mechanism permits periodic ex-post implementation.*

The implementation result follows by an argument analogous to that of Athey and Segal (2013) for dynamic team mechanisms in private-value environments, combined with the insight of Bergemann et al. (2025) developed for static settings. By conditioning transfers on an unbiased estimator of the payoff-relevant state ω, constructed from user feedback, and on realized outcomes, the mechanism ensures that in each period an agent's expected payment equals the social externality induced by her report. Consequently, each agent becomes a residual claimant

[6] Recall the platform observes feedback data after each round, including T. Hence, this expression is well-defined.

of the expected social surplus generated by the current allocation and its continuation value. Because agents revert to truthful reporting in all future periods, any deviation affects only the current-period allocation. Truthful reporting is therefore optimal precisely because the allocation rule is efficient.

Proof of Theorem 1. Fix an agent $i \in N$, a period $t \geq 1$, a public history $h_t \in \mathcal{H}_t$, a type profile $\xi_t \in \Xi_t$, and an information history $I_t \in \mathcal{I}_t$. Consider an arbitrary one-shot deviation in period t to a (possibly non-truthful) report $\xi'_{it} \in \Xi_{it}$, while all other agents report truthfully in all periods and agent i reports truthfully in all subsequent periods. By the principle of optimality, it suffices to show that no such deviation can be profitable.

Let $x_t^*(\cdot)$ denote the efficient period-t allocation rule evaluated at the report profile (ξ'_{it}, ξ_{-it}) and history h_t, i.e. $x_t^*(\cdot) = x_t^*(\xi'_{it}, \xi_{-it}, h_t) \in X_t$. Given $x_t^*(\cdot)$ and the user action rule, define the induced probability that product j is purchased in period t as

$$z_{jt}(\xi'_{it}, \xi_{-it}, h_t, \omega) \equiv \sum_{r_t \in R} y_{jt}(o_{t-1}, r_t, \omega)\, x^*_{r_t t}(\xi'_{it}, \xi_{-it}, h_t).$$

All conditional expectations below are taken with respect to the probability measure induced by the true environment, the efficient policy x^*, and the deviation ξ'_{it} at time t, conditional on $(x_t^*(\cdot), h_t, \xi_t, I_t)$. We suppress this conditioning whenever it is clear.

Under a data-driven dynamic team mechanism, the period-t transfer to agent i is

$$p_{it} = -\langle \gamma_i, \widehat{\omega}(o_t) \rangle \mathbb{1}_{a_t=i} - \sum_{j \neq i} (\theta_{jt} + \langle \gamma_j, \widehat{\omega}(o_t) \rangle) \mathbb{1}_{a_t=j} + \phi_{it}(\xi_{-it}).$$

Taking conditional expectations yields

$$\mathbb{E}[p_{it}] = -\mathbb{E}[\langle \gamma_i, \widehat{\omega}(o_t) \rangle \mathbb{1}_{a_t=i}] - \sum_{j \neq i} \mathbb{E}[\theta_{jt} \mathbb{1}_{a_t=j}] - \sum_{j \neq i} \mathbb{E}[\langle \gamma_j, \widehat{\omega}(o_t) \rangle \mathbb{1}_{a_t=j}] + \phi_{it}(\xi_{-it}). \quad (1)$$

By Lemma 1, for every $j \in N$ (including $j = i$),

$$\mathbb{E}[\langle \gamma_j, \widehat{\omega}(o_t) \rangle \mathbb{1}_{a_t=j}] = \mathbb{E}\left[\langle \gamma_j, \omega \rangle\, z_{jt}(\xi'_{it}, \xi_{-it}, h_t, \omega)\right].$$

Moreover, by definition of z_{jt} and the action rule, $\mathbb{E}[\mathbb{1}_{a_t=j}] = \mathbb{E}[z_{jt}(\xi'_{it}, \xi_{-it}, h_t, \omega)]$. Substituting these identities into (1) gives

$$\mathbb{E}[\theta_{it} z_{it} - p_{it}] = \mathbb{E}\left[\sum_{j=1}^{n} (\theta_{jt} + \langle \gamma_j, \omega \rangle)\, z_{jt}(\xi'_{it}, \xi_{-it}, h_t, \omega)\right] - \phi_{it}(\xi_{-it}),$$

Thus, up to the additive constant $-\phi_{it}(\xi_{-it})$, agent i's expected current payoff equals the expected *social flow surplus* generated in period t by the allocation $x_t^*(\xi'_{it}, \xi_{-it}, h_t)$.

Because we consider a one-shot deviation at time t and truthful play thereafter, the continuation allocation from period $t+1$ onward is efficient and does not depend on ξ'_{it} except through the realized outcome history o_t induced in period t. Under these assumptions, the continuation value of agent i under truthful reporting in future periods and the stream of unbiased estimators equals, up to an additive function independent of the reports of the focal agent,

$$\mathbb{E}[W_{t+1}(\theta_{t+1}, o_t, I_{t+1})]\,. \tag{2}$$

Indeed, the expected continuation value is given by the expectation of the sum of future flow utilities. In each period, given truthful reporting and properties of estimators, Lemma 2 shows the expected flow utility in period $s > t$ is given by

$$\mathbb{E}\left[\sum_{j=1}^{n}(\theta_{js} + \langle\gamma_j, \omega\rangle)\, z_{js}(\xi_{is}, \xi_{-is}, h_s, \omega)\right],$$

again up to addition of a function independent of the reports of the focal agent. Summing over such flow payoffs yields (2).

Therefore, agent i's expected total payoff from period t onward can be written as

$$\begin{aligned}&\mathbb{E}[\theta_{it}\mathbb{1}_{a_t=i} - p_{it} + V_{i,t+1}(\xi_{t+1}, h_{t+1}, I_{t+1})]\\ &\quad= \mathbb{E}\left[\sum_{j=1}^{n}(\theta_{jt} + \langle\gamma_j, \omega\rangle)\, z_{jt}(\xi'_{it}, \xi_{-it}, h_t, \omega)\right] + \mathbb{E}[W_{t+1}(\theta_{t+1}, o_t, I_{t+1})]\,. \end{aligned} \tag{3}$$

up to an additive term that does not depend on the report of the focal agent.

But by definition of the efficient allocation rule x_t^* (equivalently, by the Bellman recursion for W_t), for every realized (θ_t, o_{t-1}, I_t) the policy $x_t^*(\xi_{it}, \xi_{-it}, h_t)$ maximizes exactly the expected period-t social flow surplus plus the expected continuation surplus, given the information available at the beginning of period t. Therefore, no report ξ'_{it} can yield a strictly larger value of the above objective than truthful reporting ξ_{it}. □

5 Participation and Budget Surplus using Pivotal Mechanisms

The implementation result holds for any choice of the Groves term ϕ_{it}. In this section, we design this term in the spirit of the pivotal mechanism (Clarke, 1971) to ensure participation and budget surplus at the outset of the conversation. Focusing on these constraints at the initial stage is natural in our setting, as advertisers plausibly commit to participating in the recommendation environment before the interaction with the user unfolds and before subsequent feedback is realized. We study participation and budget surplus both ex ante, prior to the realization of advertisers' information about user tastes, and ex-post with respect to this information at the outset of the conversation.

5.1 Ex-Ante Properties

We begin with ex-ante properties. We say a data-driven dynamic mechanism $(\mathbf{x}, \mathbf{p})$ is *ex-ante individually rational* if each advertiser obtains a non-negative expected value from participating in the mechanism in the truthful periodic ex-post equilibrium. The mechanism satisfies *ex-ante no subsidy* if no agent receives an expected payoff exceeding the contribution of the agent to the welfare of the user.

Definition 5 (Ex-ante individual rationality)**.** Data-driven mechanism $(\mathbf{x}, \mathbf{p})$ is ex-ante individually rational if for any advertiser $i \in N$,[7]

$$\mathbb{E}\left[V_{i1}(\xi_1, I_1)\right] \geq 0.$$

Definition 6 (Ex-ante no subsidy)**.** Data-driven mechanism $(\mathbf{x}, \mathbf{p})$ gives no subsidy ex-ante if for any advertiser $i \in N$,

$$\mathbb{E}\left[\sum_{t=1}^{T} (p_{it}(\xi_t, h_t, r_t, a_t, \widehat{\omega}(o_t)) + \langle \gamma_i, \omega \rangle \mathbb{1}_{a_t=i})\right] \geq 0.$$

We define the following adjustments to the data-driven dynamic team mechanism resembling the pivot mechanism. Fix an advertiser $i \in N$, initial information $I_1 \in \mathcal{I}_1$, and profile of advertiser values $\theta_{-i1} \in \Theta_{-i1}$. We define the social problem excluding agent i:

$$W_1^{-i}(\theta_{-i1}, I_1) \equiv \max_{\{x_t \in X_t^{-i}\}_{t=1}^{T}} \mathbb{E}\left[\sum_{t=1}^{T} \sum_{j \neq i} (\theta_{jt} + \langle \gamma_j, \omega \rangle) \sum_{r_t \in R} y_{jt}(o_{t-1}, r_t, \omega)\, x_{r_t t}(\theta_{-it}, o_{t-1}, I_t) \mid I_1, \theta_{-i1}\right],$$

where $X_t^{-i} = \Delta((N \setminus \{i\}) \cup \{0\})$. Note that, while we exclude advertiser i's payoffs, we condition on the advertiser's initial information in the history I_1.

We consider Groves terms in the data-driven dynamic team mechanism that satisfy

$$\sum_{t=1}^{T} \phi_{it} = \mathbb{E}\left[W_1^{-i}(\theta_1, I_1)\right].$$

Let us denote any such sequence of terms by ϕ. We obtain the following result.

Proposition 1 (Ex-ante participation and no subsidies)**.** *Under Assumptions 1-3, the data-driven dynamic team mechanism with Groves adjustment ϕ is individually rational and gives no subsidies ex-ante.*

Proof. Fix an arbitrary advertiser $i \in N$. Under the Groves adjustment ϕ, the total ex-ante expected payoff of the advertiser is:

$$\mathbb{E}\left[V_{i1}(\xi_1, I_1)\right] = \mathbb{E}\left[W_1(\theta_1, I_1)\right] - \mathbb{E}\left[W_1^{-i}(\theta_{-i1}, I_1)\right] \geq 0,$$

which follows since for any θ_1 and I_1,

$$W_1(\theta_1, I_1) \geq W_1^{-i}(\theta_{-i1}, I_1),$$

since the inclusion of product i cannot decrease the expected social surplus when the same initial information about user tastes is used to compute the optimal allocation. Thus, the mechanism

[7] We slightly abuse notation in this expression by suppressing the redundant dependence on the public history h_1. The expectation is taken with respect to the initial types of the advertisers and the initial context observed by the platform.

is ex-ante individually rational. The no subsidy condition follows since, for any θ_{-i1} and I_1,

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{j\neq i}(\theta_{jt}+\langle\gamma_j,\omega\rangle)\sum_{r_t\in R}y_{jt}(o_{t-1},r_t,\omega)\,x^*_{r_t t}(\theta_{-it},o_{t-1},I_t)\mid I_1,\theta_{-i1}\right]\leq W_1^{-i}(\theta_{-i1},I_1),$$

by construction, where $\mathbf{x}^*$ denotes the efficient allocation in the environment including agent i. □

5.2 Ex-Post Properties at the Outset

We now turn to ex-post properties at the outset.

Definition 7 (Ex-post individual rationality at the outset)**.** Data-driven mechanism $(\mathbf{x},\mathbf{p})$ is ex-post individually rational at the outset if for any advertiser $i\in N$, types $\xi_1\in\Xi_1$, and initial information I_1,

$$V_{i1}(\xi_1,I_1)\geq 0.$$

Definition 8 (Ex-post no subsidy at the outset)**.** Data-driven mechanism $(\mathbf{x},\mathbf{p})$ gives no subsidy ex-post at the outset if for any advertiser $i\in N$, types $\xi_1\in\Xi_1$, and initial information I_1,

$$\mathbb{E}\left[\sum_{t=1}^{T}(p_{it}(\xi_t,h_t,r_t,a_t,\widehat{\omega}(o_t))+\langle\gamma_i,\omega\rangle\mathbb{1}_{a_t=i})\mid\theta_1,I_1\right]\geq 0.$$

We define the following pivotal adjustments to the data-driven dynamic team mechanism. Fix an advertiser $i\in N$, reports $\xi_1\in\Xi_1$, and initial information I_1. We now consider Groves terms in the data-driven dynamic team mechanism that satisfy, with a slight abuse of notation,

$$\sum_{t=1}^{T}\phi^+_{it}=\sup_{(\mathbf{S}_{i1},s_{i1})}W_1^{-i}(\theta_{-i1},(\mathbf{S}_{i1},s_{i1}),(\mathbf{S}_{-i1},s_{-i1}))$$

and

$$\sum_{t=1}^{T}\phi^-_{it}=\inf_{(\mathbf{S}_{i1},s_{i1})}W_1^{-i}(\theta_{-i1},(\mathbf{S}_{i1},s_{i1}),(\mathbf{S}_{-i1},s_{-i1})).$$

Denote any corresponding Groves adjustments by ϕ^+ and ϕ^-, respectively. We obtain the following result.

Proposition 2 (Participation and subsidies ex-post at the outset)**.** *Under Assumptions 1-3, the data-driven dynamic team mechanism with Groves adjustment ϕ^+ gives no subsidies ex-post at the outset. With the Groves adjustment ϕ^-, it is ex-post individually rational at the outset.*

Proof. Fix an arbitrary agent i, types $\xi_1\in\Xi_1$ and initial information I_1. Under the Groves adjustment ϕ^+, the expected transfer satisfies

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{j\neq i}(\theta_{jt}+\langle\gamma_j,\omega\rangle)\sum_{r_t\in R}y_{jt}(o_{t-1},r_t,\omega)\,x^*_{r_t t}(\theta_{-it},o_{t-1},I_t)\mid I_1,\theta_{-i1}\right]\leq W_1^{-i}(\theta_{-i1},I_1)\leq\sum_{t=1}^{T}\phi^+_{it}.$$

The claim regarding no subsidies follows.

Further, we have

$$W_1(\theta_1, I_1) \geq W_1^{-i}(\theta_{-i1}, I_1) \geq \sum_{t=1}^{T} \phi_{it}^{-},$$

and the claim regarding individual rationality now also follows. □

It may be infeasible to satisfy the no-subsidy and individual rationality conditions simultaneously using a pivotal adjustment. In environments with interdependent values, a pivotal mechanism may reward an agent for providing valuable information that improves the joint prediction problem faced by the mechanism. When this informational contribution is sufficiently large, the resulting informational rent can exceed the payment required to internalize the agent's allocative externality; see Bergemann et al. (2025) for a detailed discussion for the static setting. Nevertheless, there is an important case in which this tension disappears. Suppose that the platform's initial context is sufficiently informative about the user's taste parameter ω that an individual advertiser's private information does not affect posterior beliefs, and hence does not affect the efficient allocation:

$$F_1(\cdot|\mathbf{S}_{p1}, \mathbf{S}_{-p1}) = F_1(\cdot|\mathbf{S}_{p1}).$$

In this case, advertiser i has zero informational impact on the continuation welfare of the remaining agents. Formally, the continuation value for the environment without agent i is invariant to i's report, so that

$$\sup_{(\mathbf{S}_{i1}, s_{i1})} W_1^{-i}(\theta_{-i1}, (\mathbf{S}_{i1}, s_{i1}), (\mathbf{S}_{-i1}, s_{-i1})) = \inf_{(\mathbf{S}_{i1}, s_{i1})} W_1^{-i}(\theta_{-i1}, (\mathbf{S}_{i1}, s_{i1}), (\mathbf{S}_{-i1}, s_{-i1})).$$

Denoting the corresponding Groves terms by $\bar{\phi}$, the following result then follows immediately.

Corollary 1 (Initial context is sufficient)**.** *Suppose* $\mathbf{S}_{p1}$ *is sufficient for* ω *with respect to* $\mathbf{S}_{-p1}$ *and Assumptions 1-3 are satisfied. Then the data-driven dynamic team mechanism with Groves adjustment* $\bar{\phi}$ *is individually rational and gives no subsidies ex-post at the outset.*

6 Discussion

This paper studies a tractable model of dynamic interaction between a conversational AI shopping assistant and a user who seeks to purchase a product from a set of advertisers. Advertisers compete to be recommended and hold private information both about their conversion values and about the user's underlying preferences. We propose a class of data-driven dynamic mechanisms that leverage the information revealed during the conversation to simultaneously maximize expected social surplus and align advertisers' incentives.

While our main exposition models advertisers as offering distinct products, the framework also admits an alternative interpretation. The platform's interaction with the user, namely, determining *which product* best matches the user's needs, may be driven purely by user welfare and thus independent of advertiser incentives. Conditional on a product being selected, the platform may then choose *which seller* to recommend for that product. Sellers offering the same product may differ in price, delivery speed, fees, or return policies, generating a distinct dimension of strategic interaction. A closely related example is Amazon's Buy Box algorithm, which selects among competing sellers for the same product; see Lee and Musolff (2025) for an economic analysis.

Our approach has several limitations that suggest directions for future research. First, for tractability we assume the user is non-strategic and follows a commonly known action rule. In practice, users may search strategically, and microfounding the action rule from an explicit model of user optimization would be valuable. In canonical search models such as Weitzman (1979), user behavior depends on the distribution of future rewards, which in our setting is endogenous to the recommendation policy. Endogenizing user search would therefore introduce a fixed-point problem between user behavior and efficient recommendations, which lies beyond the scope of this paper.

Second, we focus on periodic ex-post implementation, which yields a robust implementation result independent of whether conversational data are public or proprietary. For other objectives, such as revenue maximization, it would be natural to analyze how disclosure versus non-disclosure of the history of the conversation affects incentives and market outcomes.

References

Amazon (2025). Amazon's new ai-powered shopping feature 'Help Me Decide' makes it easy to quickly pick the right product. `https://www.aboutamazon.com/news/retail/amazon-things-to-buy-help-me-decide-gen-ai` (accessed 2025-12-21).

Armstrong, M. and Zhou, J. (2011). Paying for prominence. *The Economic Journal*, 121(556):F368–F395.

Athey, S. and Segal, I. (2013). An efficient dynamic mechanism. *Econometrica*, 81(6):2463–2485.

Balseiro, S. R., Deng, Y., Mao, J., Mirrokni, V. S., and Zuo, S. (2021). The landscape of auto-bidding auctions: Value versus utility maximization. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 132–133.

Banchio, M., Mehta, A., and Perlroth, A. (2024). Ads in conversations. *arXiv preprint arXiv:2403.11022*.

Bergemann, D., Bojko, M., Duetting, P., Paes Leme, R., Xu, H., and Zuo, S. (2025). Data-driven mechanism design: Jointly eliciting preferences and information. In *Proceedings of the 26th ACM Conference on Economics and Computation*, pages 507–507.

Bergemann, D. and Morris, S. (2019). Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95.

Bergemann, D. and Välimäki, J. (2010). The dynamic pivot mechanism. *Econometrica*, 78(2):771–789.

Bergemann, D. and Välimäki, J. (2019). Dynamic mechanism design: An introduction. *Journal of Economic Literature*, 57(2):235–274.

Calvano, E., Calzolari, G., Denicolo, V., and Pastorello, S. (2025). Artificial intelligence, algorithmic recommendations and competition. *Algorithmic Recommendations and Competition (February 19, 2025).*

Clarke, E. H. (1971). Multipart pricing of public goods. *Public choice*, pages 17–33.

Crémer, J. and McLean, R. P. (1988). Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica: Journal of the Econometric Society*, pages 1247–1257.

Dasgupta, P. and Maskin, E. (2000). Efficient auctions. *The Quarterly Journal of Economics*, 115(2):341–388.

De Corniere, A. and Taylor, G. (2019). A model of biased intermediation. *The RAND Journal of Economics*, 50(4):854–882.

Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N. (2018). Consumer price search and platform design in internet commerce. *American Economic Review*, 108(7):1820–1859.

Dubey, K. A., Feng, Z., Kidambi, R., Mehta, A., and Wang, D. (2024). Auctions with llm summaries. *arXiv preprint arXiv:2404.08126.*

Dütting, P., Mirrokni, V., Paes Leme, R., Xu, H., and Zuo, S. (2024). Mechanism design for large language models. In *Proceedings of the ACM on Web Conference 2024*, pages 144–155.

Edelman, B., Ostrovsky, M., and Schwarz, M. (2007). Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American economic review*, 97(1):242–259.

Feizi, S., Hajiaghayi, M., Rezaei, K., and Shin, S. (2025). Online Advertisements with LLMs: Opportunities and Challenges. *ACM SIGecom Exchanges*, 22(2).

Fletcher, A., Ormosi, P. L., and Savani, R. (2023). Recommender systems and supplier competition on platforms. *Journal of Competition Law & Economics*, 19(3):397–426.

Google (2025). Ai is transforming shopping in search: Google Shopping's AI Mode and virtual try-on updates. https://business.google.com/us/think/search-and-video/google-shopping-ai-mode-virtual-try-on-update/ (accessed 2026-02-06).

Groves, T. (1973). Incentives in teams. *Econometrica: Journal of the Econometric Society*, pages 617–631.

Hagiu, A. and Jullien, B. (2011). Why do intermediaries divert search? *The RAND Journal of Economics*, 42(2):337–362.

Hajiaghayi, M., Lahaie, S., Rezaei, K., and Shin, S. (2024). Ad auctions for llms via retrieval augmented generation. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024.*

Holmström, B. (1979). Groves' scheme on restricted domains. *Econometrica: Journal of the Econometric Society*, pages 1137–1144.

Jehiel, P., Meyer-ter Vehn, M., Moldovanu, B., and Zame, W. R. (2006). The limits of ex post implementation. *Econometrica*, 74(3):585–610.

Jehiel, P. and Moldovanu, B. (2001). Efficient design with interdependent valuations. *Econometrica*, 69(5):1237–1259.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Lee, K. H. and Musolff, L. (2025). Two-sided markets shaped by platform-guided search.'.

Liu, H. (2018). Efficient dynamic mechanisms in environments with interdependent valuations: The role of contingent transfers. *Theoretical Economics*, 13(2):795–829.

Maskin, E. (1992). Auctions and privatization. *Privatization*, 1:15–136.

Mehta, R. (2024). How customers are making more informed shopping decisions with rufus, amazon's generative ai-powered shopping assistant. https://www.aboutamazon.com/news/retail/how-to-use-amazon-rufus (accessed 2025-12-21).

Mezzetti, C. (2004). Mechanism design with interdependent valuations: Efficiency. *Econometrica*, 72(5):1617–1626.

OpenAI (2025). Buy it in chatgpt: Instant checkout and the agentic commerce protocol. https://openai.com/index/buy-it-in-chatgpt/ (accessed 2026-02-06).

Ostrovsky, M. and Schwarz, M. (2023). Reserve prices in internet advertising auctions: A field experiment. *Journal of Political Economy*, 131(12):3352–3376.

Riordan, M. H. and Sappington, D. E. (1988). Optimal contracts with public ex post information. *Journal of Economic Theory*, 45(1):189–199.

Soumalias, E., Curry, M. J., and Seuken, S. (2024). Truthful aggregation of LLMs with an application to online advertising. In *Agentic Markets Workshop at ICML 2024*.

Varian, H. R. (2007). Position auctions. *international Journal of industrial Organization*, 25(6):1163–1178.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica*, 47(3).

A Derivations and Proofs

A.1 Proof Details for Theorem 1

Lemma 1. *Suppose Assumptions 1-3 hold. Fix a period* t*, an advertiser* j*, a report profile* $\xi_t \in \Xi_t$*, a public history* $h_t \in \mathcal{H}_t$*, and an information history* $I_t \in \mathcal{I}_t$*. Let*

$$z_{jt}(\xi_t, h_t, \omega) \;\equiv\; \sum_{r_t \in R} y_{jt}(o_{t-1}, r_t, \omega)\, x^*_{r_t t}(\xi_t, h_t).$$

Then

$$\mathbb{E}[\langle \gamma_j, \widehat{\omega}(o_t)\rangle\, \mathbb{1}_{a_t=j} \mid x^*_t(\cdot), h_t, \xi_t, I_t] = \mathbb{E}[\langle \gamma_j, \omega\rangle\, z_{jt}(\xi_t, h_t, \omega) \mid x^*_t(\cdot), h_t, \xi_t, I_t]\,.$$

Proof. Let $\mathcal{G}_t \equiv \sigma(x^*_t(\cdot), h_t, \xi_t, I_t)$. By the tower property,

$$\mathbb{E}[\langle \gamma_j, \widehat{\omega}\rangle \mathbb{1}_{a_t=j} \mid \mathcal{G}_t] = \mathbb{E}[\mathbb{E}[\langle \gamma_j, \widehat{\omega}\rangle \mathbb{1}_{a_t=j} \mid \omega, o_t, \mathcal{G}_t] \mid \mathcal{G}_t]\,.$$

Conditional on (ω, o_t), $\mathbb{1}_{a_t=j}$ is measurable, hence

$$\mathbb{E}[\langle \gamma_j, \widehat{\omega}\rangle \mathbb{1}_{a_t=j} \mid \omega, o_t, \mathcal{G}_t] = \mathbb{1}_{a_t=j}\, \langle \gamma_j, \mathbb{E}[\widehat{\omega} \mid \omega, o_t, \mathcal{G}_t]\rangle.$$

By Assumption 3(i), $\mathbb{E}[\widehat{\omega} \mid \omega, o_t, \mathcal{G}_t] = \mathbb{E}[\widehat{\omega} \mid \omega, o_t]$, and by Assumption 3(ii), $\mathbb{E}[\widehat{\omega} \mid \omega, o_t] = \omega$. Therefore

$$\mathbb{E}[\langle \gamma_j, \widehat{\omega}\rangle \mathbb{1}_{a_t=j} \mid \mathcal{G}_t] = \mathbb{E}[\langle \gamma_j, \omega\rangle \mathbb{1}_{a_t=j} \mid \mathcal{G}_t].$$

Apply the tower property again, conditioning on (ω, o_{t-1}, r_t):

$$\mathbb{E}[\langle \gamma_j, \omega\rangle \mathbb{1}_{a_t=j} \mid \mathcal{G}_t] = \mathbb{E}[\langle \gamma_j, \omega\rangle\ \mathbb{P}(a_t = j \mid \omega, o_{t-1}, r_t, \mathcal{G}_t) \mid \mathcal{G}_t]\,.$$

By the action rule, $\mathbb{P}(a_t = j \mid \omega, o_{t-1}, r_t, \mathcal{G}_t) = y_{jt}(o_{t-1}, r_t, \omega)$. Taking expectations over r_t induced by $x^*_t(\cdot)$ yields

$$\mathbb{E}[\mathbb{1}_{a_t=j} \mid \omega, o_{t-1}, \mathcal{G}_t] = \sum_{r_t \in R} y_{jt}(o_{t-1}, r_t, \omega)\, x^*_{r_t t}(\xi_t, h_t) = z_{jt}(\xi_t, h_t, \omega),$$

and the claim follows. □

Lemma 2. *Suppose Assumptions 1-3 hold. Fix periods* $t \le s \le T$*, an advertiser* j*, a report profile* $\xi_t \in \Xi_t$*, a public history* $h_t \in \mathcal{H}_t$*, and an information history* $I_t \in \mathcal{I}_t$*. Let* (ξ_s, h_s) *denote the (random) report and public histories in period* s *induced by truthful play from period* t *onward under the mechanism with allocation rule* $\mathbf{x}^*$*. Define, for every realization of* (ξ_s, h_s, ω)*,*

$$z_{js}(\xi_s, h_s, \omega) \;\equiv\; \sum_{r_s \in R} y_{js}(o_{s-1}, r_s, \omega)\, x^*_{r_s s}(\xi_s, h_s).$$

Then

$$\mathbb{E}[\langle \gamma_j, \widehat{\omega}(o_s)\rangle\, \mathbb{1}_{a_s=j} \mid x^*_t(\cdot), h_t, \xi_t, I_t] = \mathbb{E}[\langle \gamma_j, \omega\rangle\, z_{js}(\xi_s, h_s, \omega) \mid x^*_t(\cdot), h_t, \xi_t, I_t]\,.$$

Proof. Let $\mathcal{G}_t \equiv \sigma(x_t^*(\cdot), h_t, \xi_t, I_t)$. Let $\mathcal{G}_s \equiv \sigma(x_s^*(\cdot), h_s, \xi_s, I_s)$ denote the corresponding period-s sigma-field under truthful play from period t onward. By construction, $\mathcal{G}_t \subseteq \mathcal{G}_s$.

We first establish the identity conditional on $\mathcal{G}_s$ and then take conditional expectations with respect to $\mathcal{G}_t$.

By Lemma 1, we have

$$\mathbb{E}[\langle \gamma_j, \widehat{\omega}(o_s)\rangle \mathbb{1}_{a_s=j} \mid \mathcal{G}_s] = \mathbb{E}[\langle \gamma_j, \omega\rangle z_{js}(\xi_s, h_s, \omega) \mid \mathcal{G}_s].$$

Since $\mathcal{G}_t \subseteq \mathcal{G}_s$, apply the tower property:

$$\mathbb{E}[\langle \gamma_j, \widehat{\omega}(o_s)\rangle \mathbb{1}_{a_s=j} \mid \mathcal{G}_t] = \mathbb{E}[\mathbb{E}[\langle \gamma_j, \widehat{\omega}(o_s)\rangle \mathbb{1}_{a_s=j} \mid \mathcal{G}_s] \mid \mathcal{G}_t]\,.$$

Substituting the period-s identity from Step 1 yields

$$\mathbb{E}[\langle \gamma_j, \widehat{\omega}(o_s)\rangle \mathbb{1}_{a_s=j} \mid \mathcal{G}_t] = \mathbb{E}[\mathbb{E}[\langle \gamma_j, \omega\rangle z_{js}(\xi_s, h_s, \omega) \mid \mathcal{G}_s] \mid \mathcal{G}_t] = \mathbb{E}[\langle \gamma_j, \omega\rangle z_{js}(\xi_s, h_s, \omega) \mid \mathcal{G}_t],$$

where the last equality uses that $\langle \gamma_j, \omega\rangle z_{js}(\xi_s, h_s, \omega)$ is $\mathcal{G}_s$-measurable. This is exactly the claimed equality with conditioning on $(x_t^*(\cdot), h_t, \xi_t, I_t)$. □