SHOPMIND: A MENTAL WORLD MODEL FOR PERSONALIZED AND REALISTIC E-COMMERCE USER SIMULATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Large language models are widely used in e-commerce user simulation, providing a foundation for understanding users and predicting human decisions. However, existing simulators still struggle to achieve both personalization and behavioral realism. Most methods predict the next action directly from interaction histories and the current context, without accounting for the evolving mental world that underlies user behavior. Theory of Mind suggests that behavior is driven by mental states, including beliefs, uncertainties, and intentions. Users may interpret the same information in distinct ways, leading to divergent behaviors, while an individual's mental state evolves through interaction and shapes subsequent actions. Ignoring this process prevents simulators from faithfully capturing individual differences and realistic behavior patterns, and undermines its effectiveness in downstream applications. To address this, we propose **ShopMind**, a **Mental World Model** for e-commerce user simulation that captures how consumers interpret shopping information and how their mental states evolve, better reflecting individual users and real-world behavior patterns. We further use behavior–mental state pairs synthesized by ShopMind to train a **Mental State Reconstruction (MSR) model**, which infers consumers' latent mental states from observable behavior. Experiments show that ShopMind improves personalization by 18.0% and behavioral realism by 19.6% over the strongest alternatives. Moreover, MSR improves performance on consumption value measurement and CTR prediction, showing the utility of ShopMind for understanding and predicting real user behavior. [1]

1 INTRODUCTION

Large language models (LLMs) have shown strong potential for simulating human online shopping behavior (Aher et al., 2023; Argyle et al., 2023; Park et al., 2023; Horton et al., 2024). LLM-based e-commerce user simulation can provide low-cost evaluation environments for recommender systems (Zhang et al., 2025; Huang et al., 2025) and has broader applications in cognitive science and psychology (Binz & Schulz, 2023; Dillion et al., 2023; Lin, 2026). It offers an important foundation for understanding human decision-making and improving behavioral prediction.

Recent studies have increasingly explored LLM-based e-commerce user simulation. Given a user profile, interaction history, and the current shopping context, an e-commerce user simulator predicts the user's next action. Existing approaches mainly improve simulation quality along two directions: some leverage real shopping trajectories for supervised fine-tuning or reinforcement learning to better align simulated behaviors with real users (Lu et al., 2026; Zhang et al., 2026), while others condition user simulation on personas, preferences, and interaction histories to capture individual behavioral differences (Mansour et al., 2025; Wang et al., 2025; 2026; Ling et al., 2026). However, existing simulators still struggle to achieve both personalized and realistic behavior. A key limitation is that they largely predict the next action from observable signals, without explicitly modeling the evolving mental world of consumers throughout the shopping process. User behavior cannot be fully explained by observable context alone, as decisions are also shaped by how individuals perceive and interpret the information they encounter (Bettman et al., 1998). Theory of Mind provides a natural

[1] Our code is available at `https://anonymous.4open.science/r/ShopMind-0E94`.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

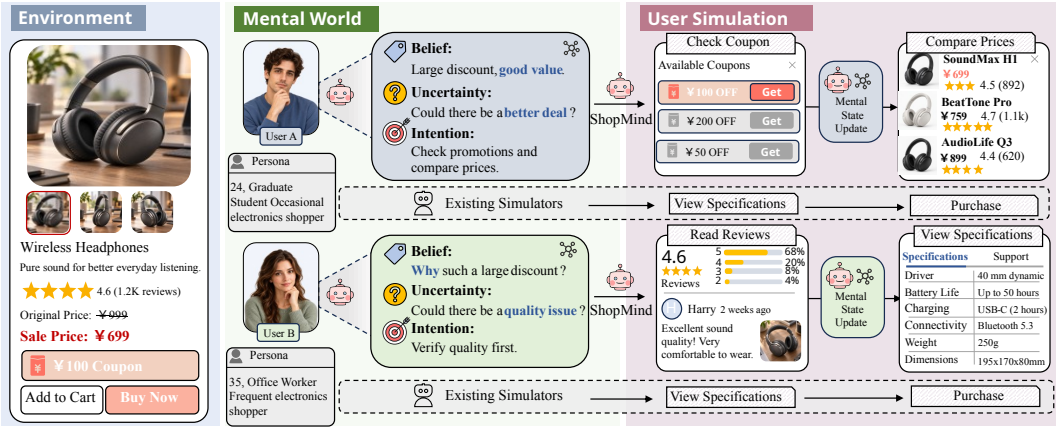


Figure 1: Illustration of the proposed **ShopMind**. Users may form different mental states from the same shopping information, leading to different behaviors. By modeling this evolving mental world, ShopMind enables more personalized and realistic user simulation than other approaches.

perspective for this view: Behavior is driven by mental states, including beliefs, intentions, and uncertainties (Premack & Woodruff, 1978). For example, when exposed to the same discounted product, a price-sensitive user may perceive it as a purchase opportunity, whereas a quality-oriented user may interpret the discount as a signal of lower quality and make a different decision. In addition, the same user's mental state may evolve through browsing, further shaping subsequent behavior. Ignoring the mental world limits both the personalization and realism of simulated behavior and undermines the simulator's effectiveness for user understanding and behavioral prediction.

To address this limitation, we propose **ShopMind**, a **mental world model** for e-commerce user simulation. As shown in Figure 1, ShopMind continuously models and updates consumers' beliefs, uncertainties, and intentions to capture how they interpret shopping information and make decisions, enabling more realistic and personalized simulation. It consists of three core modules: **Mental State Modeling and Transition** tracks the consumer's evolving mental state; **Mental-World Rollout** explores candidate actions by simulating their subsequent mental-state and behavioral trajectories; and **Human-Aligned Behavior Selection** selects the next action by balancing individual consistency and behavioral realism, with realism assessed by a Behavior Critic trained on real shopping trajectories. The selected action drives the next interaction, which in turn updates both the shopping environment and the consumer's mental state, yielding shopping trajectories that are both behaviorally plausible and tailored to individual users. Throughout this process, ShopMind continuously aligns evolving mental states with generated behaviors, naturally producing behavior–mental-state pairs that are largely unavailable in real-world interaction logs. We use these synthesized pairs to train a **Mental State Reconstruction (MSR) model**. While ShopMind generates behavior from mental states, MSR tackles the inverse problem of inferring latent mental states from observed behavior, supporting more accurate user understanding and behavior prediction on real-world data.

Experiments show that ShopMind achieves gains of 18.0% in personalization and 19.6% in behavioral realism over the strongest baselines. Moreover, the mental states reconstructed by MSR improve consumption value measurement and CTR prediction, indicating that the mental-state supervision produced by ShopMind facilitates the understanding and prediction of real user behavior.

In summary, the main contributions of this work can be summarized as follows:

- We propose **ShopMind**, a mental world model for e-commerce user simulation. ShopMind models consumers' mental states and their evolution, capturing how users interpret shopping information and make decisions to enable personalized and realistic simulation.
- ShopMind synthesizes shopping trajectories paired with evolving mental states. A **Mental State Reconstruction (MSR) model** is trained on these data to recover latent mental states from observable behavior, supporting user understanding and behavior prediction.
- Experiments show that ShopMind improves personalization by 18.0% and behavioral realism by 19.6% over the strongest baselines, while MSR-reconstructed mental states significantly improve consumption value measurement and future behavior prediction.

2 Related Work

E-commerce User Simulation. LLM-based virtual shoppers have recently emerged as a promising approach to e-commerce user simulation. PAARS constructs persona-conditioned shopper populations from historical data to better match real user distributions (Mansour et al., 2025). Lu et al. (2026) show that supervised fine-tuning on real shopping trajectories improves behavioral fidelity, while Shop-R1 further employs reinforcement learning to align generated actions with human behavior (Zhang et al., 2026). Customer-R1 extends this line toward personalized simulation by incorporating explicit user personas (Wang et al., 2025), and OPeRA and ShopperBench further emphasize persona-aware, multi-step shopping behavior (Wang et al., 2026; Ling et al., 2026). Despite these advances, existing approaches still predict behavior largely from observable signals, with limited consideration of how consumers interpret information and how their mental states evolve over time.

World Model. World models learn environment dynamics to anticipate the outcomes of different actions and support planning (Ha & Schmidhuber, 2018). Representative methods such as PlaNet, Dreamer, and MuZero use learned or imagined future trajectories to guide decision-making (Hafner et al., 2019; 2020; Schrittwieser et al., 2020), but primarily model the external world. Recent work extends this idea to the mental world by explicitly modeling latent human states, including beliefs, goals, and intentions, and how they evolve with observations and actions (Liu et al., 2026a; Fei & Zhao, 2026). Fei & Zhao (2026) further formulate mental world modeling as jointly tracking physical and mental states to better predict human decisions. Building on this perspective, ShopMind brings mental world modeling to e-commerce user simulation by continuously tracking consumers' beliefs, uncertainties, and intentions and using their evolution to drive subsequent behavior.

3 Problem Formulation

We study personalized and realistic user behavior simulation in e-commerce. We formulate the task as follows. For each user u, let P_u denote the user profile capturing relatively stable user characteristics, and T_u the shopping task specifying the goal and constraints of the current episode. At step t, O_t denotes the observable shopping context, while H_t summarizes the interaction history and retrieved memory. The goal is to model a user policy π that generates the next action A_t conditioned on the available context, i.e., $A_t \sim \pi(\cdot \mid P_u, T_u, O_t, H_t)$. Rather than directly mapping observable context to actions, ShopMind models user behavior through an evolving mental state M_t, where $M_t = F_\theta(P_u, T_u, O_t, H_t)$ and $A_t \sim \pi_\theta(\cdot \mid M_t, P_u, T_u, O_t, H_t)$. Here, F_θ models mental-state formation and transition, while M_t captures the user's current beliefs, uncertainties, and intentions. After executing A_t, the environment returns a new observation O_{t+1}, which further updates the mental state, forming the sequential process $O_t \to M_t \to A_t \to O_{t+1} \to M_{t+1}$.

4 ShopMind: Mental World Model

Existing methods typically predict user actions from observable shopping context, overlooking the evolving mental state that shapes behavior. ShopMind addresses this limitation by modeling a dynamic mental world that captures the consumer's internal mental state, and uses it to guide subsequent behavior generation. As shown in Figure 2, the framework consists of three stages: **(1) Mental State Modeling and Transition** maintains a structured representation of the consumer's mental state (beliefs, uncertainties, and intentions), and updates it as new information arrives; **(2) Mental-World Rollout** generates candidate actions and simulates their future environment, mental-state, and behavioral trajectories; and **(3) Human-Aligned Behavior Selection** ranks candidates by individual consistency and behavioral realism, combining LLM-based trajectory assessment with a Behavior Critic trained on real consumer trajectories. Together, these stages capture how evolving mental states shape behavior while producing personalized and realistic user behaviors.

4.1 Mental State Modeling and Transition

Theory of Mind (Premack & Woodruff, 1978) suggests that human behavior depends on internal mental states, such as beliefs, uncertainties, and intentions, which evolve as new information is acquired. Inspired by this perspective, ShopMind represents the consumer's mental state in a structured form, making its evolution explicit and consistently trackable throughout the shopping process.

Guided by Theory of Mind, we represent the consumer's mental state at time step t as

$$M_t = (B_t, U_t, I_t, E_t)\,,$$

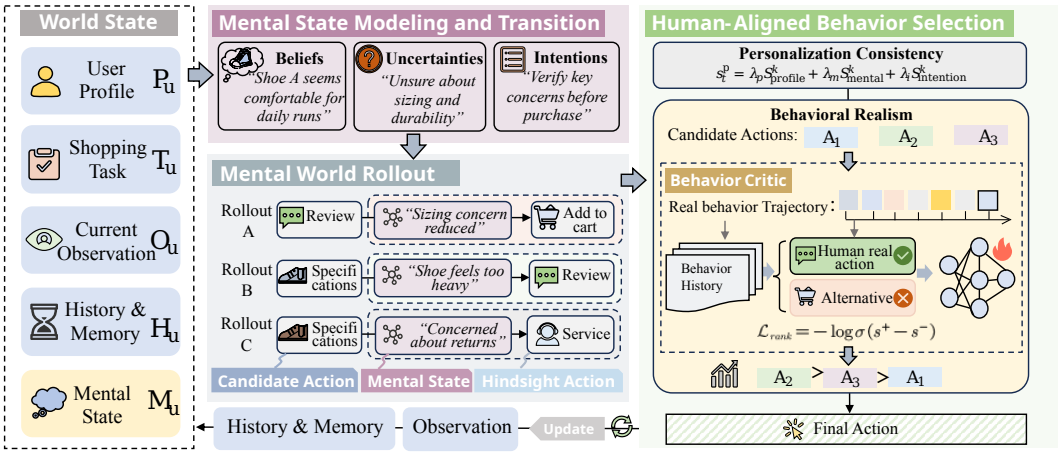


Figure 2: Overview of the **ShopMind** framework. The model maintains an evolving mental state, performs mental-world rollouts over candidate behaviors, and selects the final action by jointly considering personalization consistency and behavioral realism through a learned Behavior Critic.

where B_t summarizes the beliefs formed from observed information, including the user's understanding and judgments of the shopping context and candidate products; U_t captures unresolved uncertainties that may influence subsequent decisions; I_t specifies the user's current intention and immediate goal; and E_t stores the supporting evidence for these mental states.

Since the mental state depends on both the current observation and interaction history, ShopMind maintains a memory H_t of decision-relevant past behaviors and observations. The mental state is initialized from the user profile, shopping task, and initial observation, and is subsequently updated using the previous state, interaction memory, and latest action and observation:

$$M_0 = F_\theta^{\text{init}}(P_u, T_u, O_0), \qquad M_{t+1} = F_\theta(M_t, P_u, T_u, H_t, A_t, O_{t+1}).$$

A key challenge in modeling mental-state dynamics is that users' mental states do not necessarily change with every behavioral interaction. Treating every low-level action as a mental-state transition would introduce spurious fluctuations and undermine the temporal coherence of the inferred mental trajectory. To address this, ShopMind introduces an update gate that triggers a state transition only when the newly observed information is sufficiently informative to update the current mental state; otherwise, $M_{t+1} = M_t$. This mechanism preserves stable mental states across behaviorally trivial interactions while allowing meaningful cognitive changes to emerge when new evidence is acquired.

4.2 Mental World Rollout

Given the mental state, ShopMind first generates a set of candidate actions. However, evaluating candidates only at the current step can be myopic, as actions with similar immediate plausibility may lead to different future trajectories. For example, when a user is comparing two products, both examining the current product further and switching to another candidate may seem reasonable at the current step, yet lead to different trajectories: one may reduce uncertainty and clarify preference, while the other may provide little new information and prolong the comparison. Accordingly, ShopMind rolls out each candidate independently, simulating subsequent environment changes, mental-state transitions, and behaviors, thereby promoting more coherent and realistic decision trajectories.

We define the current decision context as $J_t = (P_u, T_u, O_t, H_t, M_t)$, consisting of the user profile, shopping task, current observation, interaction memory, and mental state. Given J_t, ShopMind generates K candidate actions: $\mathcal{C}_t = \{a_t^1, \ldots, a_t^K\}$. To ensure executability, candidate generation is constrained by the valid action space under the environment state. Each candidate a_t^k is then rolled out for L steps in an independent copy of the environment, producing an imagined trajectory

$$\hat{\tau}_t^k = \left(a_t^k, \hat{O}_{t+1}^k, \hat{M}_{t+1}^k, \ldots, \hat{O}_{t+L}^k, \hat{M}_{t+L}^k\right).$$

The rollout jointly simulates how each candidate may reshape the environment and the user's mental state, allowing the current action to be evaluated through its possible future trajectory.

4.3 HUMAN-ALIGNED BEHAVIOR SELECTION

For each candidate trajectory, ShopMind evaluates it along two dimensions: Personalized Consistency and Behavioral Realism. The former assesses alignment with the user's profile and mental state. Behavioral Realism, in contrast, measures real-world plausibility, distinguishing among individually consistent actions and favoring those that better reflect real-world behavioral patterns.

Personalized Consistency. It considers three aspects of the rollout trajectory: profile alignment S^k_{profile} with the user's characteristics, mental-state coherence S^k_{mental} throughout the rollout, and intention consistency $S^k_{\text{intention}}$ between the current intention and subsequent behaviors. An LLM-based evaluator scores these dimensions, and the resulting scores are aggregated as

$$s_t^{k,\text{pc}} = \lambda_p S^k_{\text{profile}} + \lambda_m S^k_{\text{mental}} + \lambda_i S^k_{\text{intention}},$$

to quantify the consistency of the trajectory with the user's mental state and behavioral tendencies.

Behavioral Realism. Multiple candidate actions may be similarly plausible for a given user, yet differ in behavioral realism. When uncertain about fit, for instance, the user may check the size guide, read reviews, or contact customer service, although these actions vary in how closely they reflect real-world behavioral patterns. Therefore, we train a **Behavior Critic** on real consumer trajectories to assign each candidate a scalar behavioral plausibility score, indicating how well the action matches real user behavior under the current shopping context: $c_t^k = C_\phi(H_t, O_t, a_t^k)$, where H_t and O_t denote the interaction history and observation, respectively. Higher scores indicate greater behavioral realism. The critic is trained with a pairwise ranking objective, encouraging the observed next action a^+ to score higher than an alternative a^-: $C_\phi(h_t, a^+) > C_\phi(h_t, a^-)$. We optimize

$$\mathcal{L}_{\text{critic}} = -\log \sigma\big(C_\phi(h_t, a^+) - C_\phi(h_t, a^-)\big)\,.$$

For final selection, we rank candidates by personalized consistency $s_t^{k,\text{pc}}$ and define the top-two margin as $\Delta_t = s_t^{(1),\text{pc}} - s_t^{(2),\text{pc}}$. If $\Delta_t \geq \tau_{\text{margin}}$, we select the top-ranked candidate unless strongly rejected by the Behavior Critic. Otherwise, candidates with $s_t^{k,\text{pc}} \geq \tau_{\text{compat}}$ form $\mathcal{C}_t^{\text{pc}}$, from which the Behavior Critic selects the most realistic action:

$$A_t = \arg\max_{a_t^k \in \mathcal{C}_t^{\text{pc}}} C_\phi(H_t, O_t, a_t^k).$$

This design preserves clear user-specific behavioral tendencies while using real consumer behavior to resolve ambiguous choices, balancing personalization and behavioral realism.

5 LEARNING MENTAL STATE RECONSTRUCTION FROM SHOPMIND

Beyond behavior simulation, ShopMind provides scalable supervision for reconstructing latent mental states from observed behavior. Real-world e-commerce logs capture user interaction histories, but rarely record how mental states evolve during shopping. These latent states provide a richer representation of users for understanding their values and predicting future behavior, yet obtaining such annotations through surveys or interviews is costly and difficult to scale. ShopMind instead jointly synthesizes behavioral trajectories and aligned mental-state transitions, enabling large-scale behavior–mental-state supervision. After quality filtering, we use these data to train a Mental State Reconstruction (MSR) model that infers latent mental states from real behavioral traces and supports downstream consumer value measurement and behavior prediction.

5.1 DATA CONSTRUCTION

Given the lack of continuous mental-state annotations in real-world logs, we leverage ShopMind to synthesize complete shopping trajectories together with the mental-state sequences that evolve throughout the interaction. Each trajectory starts from the initial shopping context and ends with a purchase or session termination. We organize each trajectory as a single MSR training instance, pairing the complete observable interaction with its aligned mental-state sequence:

$$\mathcal{D}_{\text{MSR}} = \left\{ \left(X^{(i)}_{1:T_i}, M^{(i)}_{1:K_i} \right) \right\}_{i=1}^{N},$$

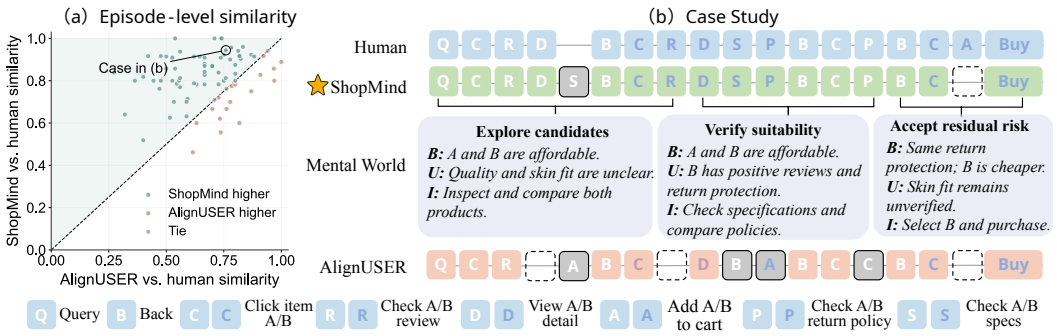


Figure 3: Qualitative comparison of ShopMind and AlignUSER (Bougie et al., 2026b), a strong baseline, on 80 randomly sampled episodes from the Taobao real log dataset. (a) ShopMind shows higher similarity to human trajectories. (b) A case illustrates how mental-world modeling helps ShopMind better follow human decision dynamics, while AlignUSER exhibits larger deviations.

where $X^{(i)}_{1:T_i}$ denotes the user profile and interaction trajectory, and $M^{(i)}_{1:K_i}$ denotes the corresponding sequence of latent mental states, with each mental state potentially spanning multiple consecutive interactions. To ensure dataset quality, we first perform deterministic checks on purchase outcomes, action execution, and mental-state transitions, followed by an LLM-based evaluation of task fit, decision consistency, state–action consistency, and evidence quality. We further conduct human review to validate trajectory quality. After filtering, the final dataset contains 5K high-quality samples for MSR training. Quality-control results are reported in Appendix C.2.

5.2 Mental State Reconstruction Model

We train a Mental State Reconstruction (MSR) model on the quality-controlled trajectories to recover latent mental-state evolution from observed shopping behavior. Given an observable trajectory $X_{1:T}$, MSR reconstructs the corresponding mental-state sequence: $\hat{M}_{1:T} = f_\theta(X_{1:T})$, where each $\hat{M}_t = \{\hat{B}_t, \hat{U}_t, \hat{I}_t\}$ represents the reconstructed beliefs, uncertainties, and intentions at step t. We use Qwen3-8B (Yang et al., 2025a) as the backbone and fine-tune it on the 5K quality-filtered trajectories using supervised fine-tuning (SFT). MSR is formulated as a conditional generation task, producing the same structured mental-state representation as ShopMind. The training objective is

$$\mathcal{L}_{\text{MSR}} = -\sum_{i=1}^{N} \log p_\theta \left(M^{(i)}_{1:T_i} \mid X^{(i)}_{1:T_i} \right).$$

This objective enables MSR to reconstruct how latent mental states evolve throughout an observed shopping trajectory. Additional training details are provided in Appendix C.

6 Experiments

Our experiments focus on two research questions: *(1) RQ1: Can ShopMind generate more personalized and realistic shopping behaviors than existing e-commerce user simulators? (2) RQ2: Can MSR trained on ShopMind-synthesized data improve real-user understanding and behavior prediction, demonstrating the real-world transferability of the synthesized mental-state supervision?* We further conduct ablation studies to quantify the contribution of each key component.

6.1 Experimental Setup

Implementation Details. ShopMind uses DeepSeek-V4-Pro (DeepSeek-AI, 2026) as its backbone, with $K = 4$ candidates and an $L = 2$ rollout horizon. The Behavior Critic is initialized from Qwen3-4B (Yang et al., 2025a) and fine-tuned with LoRA using LLaMA-Factory (Zheng et al., 2024) on two NVIDIA A100 80GB GPUs. More details are provided in Appendix A.

Datasets and Models. For RQ1, we evaluate ShopMind on next-action prediction and multi-step trajectory reproduction using OPeRA (Wang et al., 2026), ShopCART (Lu et al., 2026), and Taobao

Table 1: Performance comparison of individual consistency. ShopMind achieves the best performance across all metrics. ⋆, †, and ‡ denote the first-, second-, and third-best results, respectively.

		OPeRA				Taobao			
Method	**Method Type**	**Ind. JS ↓**	**Spearman ↑**	**Gap Err. ↓**	**Δ JS ↑**	**Ind. JS ↓**	**Spearman ↑**	**Gap Err. ↓**	**Δ JS ↑**
Zero-shot	Prompting	31.02	0.334	0.054	−0.44	34.28	0.281	0.061	−0.62
SFT	SFT	26.48	0.330	0.051	−1.41	29.82	0.308	0.051	−1.94
SFT+Rationale	SFT	24.31	0.363	0.039	−2.61	25.16‡	0.337	0.042‡	−0.91
SimUSER v2	Agent	32.24	0.238	0.055	+0.29	31.82	0.254	0.057	+1.14
DGDPO	Agent	30.22	0.296	0.037	+1.74‡	29.74	0.303	0.044	+1.86
ContextSim	Agent	32.04	0.214	0.054	+2.44†	28.96	0.319	0.049	+2.17‡
IToM	Agent	30.25	0.315	0.054	+1.25	27.83	0.326	0.047	+2.35†
AlignUSER	Agent	20.36†	0.376†	0.035†	−0.36	23.47†	0.348‡	0.039†	+0.18
Shop-R1	RL	22.60‡	0.368‡	0.036‡	−1.27	25.72	0.341	0.045	−1.52
Customer-R1	RL	27.46	0.347	0.043	−1.53	26.41	0.352†	0.043	−0.94
ShopMind	**Mental World**	**17.21** ⋆	**0.438** ⋆	**0.028** ⋆	**+4.12** ⋆	**18.63** ⋆	**0.421** ⋆	**0.031** ⋆	**+4.36** ⋆

Table 2: Performance comparison of local behavior prediction. ShopMind achieves the best performance on most metrics. ⋆, †, and ‡ denote the first-, second-, and third-best results, respectively.

	OPeRA					ShopCART				
Method	**Exact Acc. ↑**	**Non-Term. Acc. ↑**	**Macro F1 ↑**	**Top-3 Acc. ↑**	**Buy F1 ↑**	**Exact Acc. ↑**	**Macro F1 ↑**	**Top-3 Acc. ↑**	**Buy F1 ↑**	**Exit F1 ↑**
Zero-shot	6.10	4.85	32.50	8.98	67.31†	6.96	17.60	9.23	36.76	0.62
SFT	29.38	27.50	45.45	42.35	20.69	20.83	40.62	41.28	35.46	33.71
SFT+Rationale	33.04	29.89	42.32	46.23	52.05	15.14	39.09	43.60	59.70†	14.25
SimUSER v2	5.32	3.55	35.81	12.31	42.67	4.26	27.98	1.23	29.51	1.23
DGDPO	5.99	5.14	34.33	9.20	67.31†	6.41	20.72	10.85	41.63	3.17
ContextSim	16.19	12.96	35.46	28.49	42.67	11.05	35.79	35.51	20.93	36.45‡
IToM	6.21	4.70	31.21	8.76	**68.63** ⋆	7.57	18.08	9.49	37.16	2.45
AlignUSER	35.92‡	33.62†	47.81‡	47.89†	51.85	22.98‡	48.40†	44.13‡	31.37	58.77†
Shop-R1	37.18†	32.95‡	49.26†	47.31‡	52.74	16.45	41.49‡	44.47†	**60.61** ⋆	22.83
Customer-R1	27.82	24.63	18.74	31.86	0.00	28.02†	15.11	30.46	0.00	**94.85** ⋆
ShopMind	**41.22** ⋆	**37.10** ⋆	**52.89** ⋆	**60.85** ⋆	54.23‡	**31.29** ⋆	**49.14** ⋆	**45.94** ⋆	58.53‡	27.45

real shopping logs. We compare against representative LLM-based user simulation methods, including SimUSER (Bougie & Watanabe, 2025), DGDPO (Liu et al., 2026b), AlignUSER (Bougie et al., 2026b), IToM (Chen et al., 2026), Shop-R1 (Zhang et al., 2026), ContextSim (Bougie et al., 2026a), and Customer-R1 (Wang et al., 2025) (see Appendix D). For RQ2, we evaluate MSR on two tasks: consumption value measurement and CTR prediction. On B2V-Bench (Hou et al., 2026), we compare against 10 strong LLMs and B2V-Verifier to assess whether mental states improve value measurement. For CTR prediction, we construct a test set from real Taobao logs, using DeepSeek-V4-Pro (DeepSeek-AI, 2026) and Qwen2.5-72B (Yang et al., 2025b) as backbones (see Appendix E).

Evaluation Metrics. RQ1 is evaluated using Individual Action Distribution JS, User-level Spearman ρ, Persona Shuffle Δ, and Group Gap Error for individual consistency; Exact Action Accuracy, Action Type Macro-F1, Top-k Action Accuracy, Non-Terminal Action Accuracy, Purchase F1, and Exit F1 for local behavioral fidelity; and Action Transition JS, Session Length Error, Purchase Rate Error, and Action JS for sequential realism. Jensen–Shannon divergence is used for distributional comparison (Lin, 1991). RQ2 adopts the evaluation protocol of B2V-Bench (Hou et al., 2026) for consumption value measurement, and uses AUC, GAUC, and PR-AUC for CTR prediction.

6.2 RQ1: Personalization and Behavioral Realism

ShopMind is evaluated along two dimensions: **Individual Consistency** and **Behavioral Realism**. Behavioral realism is further examined at both local decision and long-horizon trajectory levels. Figure 3 qualitatively illustrates closer alignment between ShopMind and human trajectories.

Table 3: Performance comparison of long-horizon behavioral realism. ShopMind achieves consistent gains across all metrics. ⋆, †, and ‡ denote the first-, second-, and third-best results, respectively.

	OPeRA				ShopCART				Taobao			
Method	**Trans. JS ↓**	**Len. Err. ↓**	**Buy Err. ↓**	**Action JS ↓**	**Trans. JS ↓**	**Len. Err. ↓**	**Buy Err. ↓**	**Action JS ↓**	**Trans. JS ↓**	**Len. Err. ↓**	**Buy Err. ↓**	**Action JS ↓**
Zero-shot	38.06	7.47	31.11	23.53	51.73	2.80	48.18	31.20	44.82	6.95	27.64	25.81
SFT	18.05	4.78	43.33	14.65	23.84	2.64	6.37	12.76	17.83‡	4.35	12.77	13.26
SFT+Rationale	14.45‡	5.13	15.56	10.84†	33.05	4.27	1.40‡	15.30	20.31	4.88	10.42	13.95
SimUSER v2	32.74	7.21	12.22	25.04	28.07	3.62	19.89	21.78	29.63	6.38	14.21	22.16
DGDPO	39.14	7.11	34.44	22.16	41.36	3.18	31.27	22.43	34.88	6.17	24.93	20.42
ContextSim	17.56	7.20	12.22	18.32	20.95‡	2.38‡	8.40	11.59‡	18.94	5.22	9.18	14.03
IToM	37.85	7.54	34.44	21.43	48.08	2.78	43.42	26.00	40.26	6.84	29.77	23.65
AlignUSER	13.14†	4.37†	4.44†	11.38‡	19.73†	2.41	0.28†	6.18†	16.72†	4.12‡	5.36‡	12.47‡
Shop-R1	15.82	5.06	9.74‡	12.63	32.25	4.14	1.68	13.82	27.14	4.56	3.92†	11.84†
Customer-R1	25.68	4.61‡	14.83	16.27	50.66	2.21†	9.80	17.56	31.77	3.91†	7.13	14.28
ShopMind	**10.92** ⋆	**3.86** ⋆	**3.71** ⋆	**8.93** ⋆	**16.84** ⋆	**1.98** ⋆	**0.21** ⋆	**5.24** ⋆	**12.18** ⋆	**3.62** ⋆	**2.84** ⋆	**9.41** ⋆

Table 4: Consumption value measurement results. Best, second, and third results are marked by ⋆, †, and ‡. MSR-reconstructed mental states provide the strongest gains for value measurement.

Type	**Model**	**NDCG @1 ↑**	**NDCG @2 ↑**	**NDCG @3 ↑**	**Primary Acc. ↑**	**Primary Top-2 Acc. ↑**
General LLM	Claude-Haiku-4 (Anthropic, 2025)	0.7135	0.6721	0.6716	0.5743	0.7434
	DeepSeek-V3.2 (DeepSeek-AI et al., 2025)	0.7211	0.6967	0.6917	0.5613	0.7565
	Gemini-2.5-Flash (Gemini Team, Google, 2025)	0.7750	0.7026	0.6746	0.5780	0.7174
	Gemini-2.5-Pro (Gemini Team, Google, 2025)	0.5353	0.5170	0.5142	0.4442	0.5576
	GLM-5 (GLM-5 Team et al., 2026)	0.7044	0.6729	0.6594	0.5780	0.7137
	GPT-4o (OpenAI, 2024)	0.7211	0.6417	0.6345	0.5260	0.6449
	GPT-5 (OpenAI, 2025)	0.6858	0.6873	0.7052	0.5483	0.7590
	Qwen3-235B (Yang et al., 2025a)	0.7304	0.6998	0.6975	0.5743	0.7416
Task-specific	B2V-Verifier Hou et al. (2026)	0.8236†	0.8069†	0.7977†	0.6518†	0.8420†
	B2V + Zero-shot Mental State	0.7931‡	0.7662‡	0.7849‡	0.6123‡	0.7943‡
	B2V + MSR Mental State (Ours)	**0.8474** ⋆	**0.8714** ⋆	**0.8590** ⋆	**0.7868** ⋆	**0.9154** ⋆
	Δ vs. B2V-Verifier	**+2.89%**	**+7.99%**	**+7.68%**	**+20.71%**	**+8.72%**

Individual Consistency. We assess individual consistency by comparing simulated behaviors with real user behavior patterns. Given each user's profile and shopping task, the model generates shopping trajectories, from which these distributions are computed. As shown in Table 1, ShopMind achieves an average 18.0% improvement in personalization, as measured by Individual JS across OPeRA and Taobao. We also perform a persona-shuffle test by randomly permuting user profiles while keeping all other inputs unchanged. ShopMind exhibits the largest ΔJS, indicating a stronger dependence on user-specific information and greater sensitivity to individual differences.

Behavioral Realism. We first evaluate next-action prediction given the user profile, current observation, and interaction history. As shown in Table 2, ShopMind achieves the best performance on most metrics across both datasets. Although some baselines achieve higher F1 on individual terminal actions, their predictions are often biased toward a specific outcome (e.g., Customer-R1 attains high Exit F1 but zero Buy F1). Furthermore, we evaluate behavioral realism over complete shopping sessions. Starting only from the initial user information and environment, each model acts autonomously until termination. Table 3 shows that ShopMind achieves the best performance on all 12 metrics. Using Transition JS as the representative metric of behavioral realism, ShopMind achieves an average relative improvement of 19.6% over the strongest baselines. These results suggest that modeling evolving mental world and future action consequences helps distinguish among plausible actions, improving both local prediction accuracy and long-horizon behavioral realism.

6.3 RQ2: Real-World Utility of MSR

We use consumption value measurement to test whether MSR provides a deeper understanding of users, and CTR prediction to assess its utility for subsequent behavior prediction.

Consumption Value Measurement. For each real shopping trajectory, MSR reconstructs the latent mental state from observed behavior, which is integrated with the original behavior to predict the user's consumption value constructs. As shown in Table 4, B2V + MSR Mental State consistently

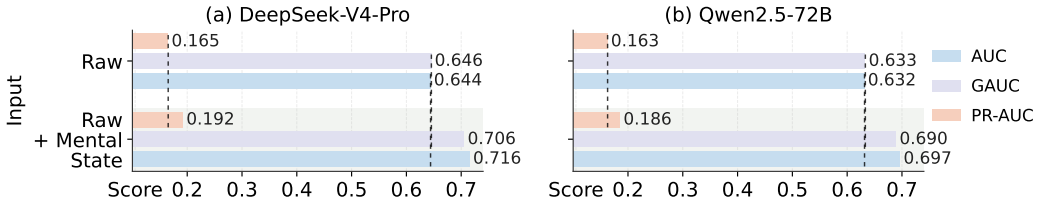


Figure 4: CTR prediction performance with and without MSR-reconstructed mental states. Adding mental states consistently improves AUC, GAUC, and PR-AUC across both models.

Table 5: Ablation study of ShopMind on Taobao logs across behavioral fidelity and personalization.

	Local Fidelity			Sequential Realism				Individual Consistency		
Variant	**Exact Acc.** ↑	**Macro F1** ↑	**Top-3 Acc.** ↑	**Trans. JS** ↓	**Len. Err.** ↓	**Buy Err.** ↓	**Action JS** ↓	**Indiv. JS** ↓	**Spearman** ρ ↑	**Δ JS** ↑
ShopMind	**53.58**	**59.02**	**72.4**	**12.18**	**3.62**	**2.84**	**9.41**	**18.63**	**0.421**	**+4.36**
w/o Mental World	13.27	38.61	21.9	30.84	6.37	11.26	16.92	28.41	0.301	+0.84
w/o Rollout	50.84	56.47	68.9	22.63	5.41	7.38	15.27	20.12	0.397	+3.72
w/o Individual Consistency	50.96	55.83	68.2	13.86	3.91	3.31	10.24	29.16	0.318	+1.43
w/o Behavior Critic	44.72	49.86	62.5	18.41	4.67	6.14	14.38	20.37	0.389	+3.58

outperforms both general LLMs and B2V-Verifier, with the largest gain on Primary Accuracy. In contrast, B2V + Zero-shot Mental State performs worse than B2V-Verifier, indicating that the gains stem from accurate mental-state reconstruction rather than simply adding inferred mental information, and further validating the effectiveness of the ShopMind-synthesized training data.

CTR Prediction. In this task, reconstructed mental states provide additional user-side signals for predicting subsequent click behavior. As shown in Figure 4, incorporating MSR-reconstructed mental states consistently improves performance across both backbones and all metrics, with full results reported in Appendix E.2. Directly using zero-shot mental states, however, degrades performance relative to the raw-behavior input. This comparison provides stronger evidence that the gains stem from ShopMind-synthesized mental-state supervision rather than from simply adding mental-state descriptions, demonstrating its transferability to real-world behavioral forecasting.

6.4 Ablation Study

We ablate each core component of ShopMind to examine its contribution. The ablation results in Table 5 show that each component contributes to a distinct aspect of ShopMind. Removing the Mental World causes the largest degradation, highlighting its central role in maintaining coherent and personalized behavior. Without Rollout, performance declines mainly in sequential realism, suggesting that short-horizon simulation is important for maintaining realistic long-term trajectories. In contrast, ablating Individual Consistency primarily degrades user-level consistency and persona sensitivity, confirming its role in preserving individualized behavioral patterns. The Behavior Critic mainly affects local and sequential realism by filtering behaviorally implausible actions. Overall, all components contribute effectively to ShopMind's performance. Additionally, the hyperparameter sensitivity analysis in Appendix B shows that ShopMind remains robust across a range of hyperparameter settings, with $K = 4$ and $L = 2$ achieving the best overall performance.

7 Conclusion

In this work, we introduce **ShopMind**, a **Mental World Model** for e-commerce user simulation that captures how consumers interpret shopping information and continuously form and update their beliefs, uncertainties, and intentions to drive more personalized and realistic behavior. Using the synthesized behavior–mental-state pairs, we further train a **Mental State Reconstruction model** that infers latent mental states from observed behaviors for real-user understanding and behavior prediction. Experiments show that ShopMind improves personalization and behavioral realism by 18% and 19.6%, respectively, while MSR-reconstructed mental states further enhance consumption value measurement and future behavior prediction, validating the utility of ShopMind-synthesized mental-state supervision. This work takes a step toward establishing Mental World Models as a general framework for modeling human cognition and behavior, with broader implications for recommender systems, user modeling, and computational cognitive science.

AI USE STATEMENT

We used generative AI tools to assist with the generation of synthetic behavior–mental-state data and to polish the writing of the manuscript. All AI-assisted data and text were reviewed and verified by the authors. We take full responsibility for the final content of this work.

ETHICS STATEMENT

This work uses real-world user interaction data that were anonymized and de-identified prior to analysis. No personally identifiable information is included in the data used for our experiments, and all analyses are conducted solely on anonymized behavioral records for research purposes. Accordingly, the study is designed to preserve user privacy and minimize potential privacy risks.

REPRODUCIBILITY STATEMENT

The code and implementation details for this work are available in our anonymous repository at `https://anonymous.4open.science/r/ShopMind-0E94`. The main paper describes the model architecture, training procedure, datasets, baselines, and experimental settings, while additional implementation details and hyperparameter configurations are provided in the appendix.

REFERENCES

Gati V. Aher, Rosa I. Arriaga, and Adam Tauman Kalai. Using large language models to simulate multiple humans and replicate human subject studies. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 337–371. PMLR, 2023.

Anthropic. Claude haiku 4.5 system card. Anthropic, October 2025. URL `https://www-cdn.anthropic.com/7aad69bf12627d42234e01ee7c36305dc2f6a970.pdf`.

Lisa P. Argyle, Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351, 2023. doi: 10.1017/pan.2023.2.

James R. Bettman, Mary Frances Luce, and John W. Payne. Constructive consumer choice processes. *Journal of Consumer Research*, 25(3):187–217, 1998. doi: 10.1086/209535.

Marcel Binz and Eric Schulz. Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6):e2218523120, 2023. doi: 10.1073/pnas.2218523120.

Nicolas Bougie and Narimawa Watanabe. SimUSER: Simulating user behavior with large language models for recommender system evaluation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)*, pp. 43–60, Vienna, Austria, July 2025. Association for Computational Linguistics. doi: 10.18653/v1/2025.acl-industry.5. URL `https://aclanthology.org/2025.acl-industry.5/`.

Nicolas Bougie, Gian Maria Marconi, Xiaotong Ye, and Narimasa Watanabe. Beyond offline A/B testing: Context-aware agent simulation for recommender system evaluation. *arXiv preprint arXiv:2604.09549*, 2026a. URL `https://arxiv.org/abs/2604.09549`.

Nicolas Bougie, Gian Maria Marconi, Xiaotong Ye, and Narimawa Watanabe. AlignUSER: Human-aligned LLM agents via world models for recommender system evaluation. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 16433–16451, San Diego, California, United States, July 2026b. Association for Computational Linguistics. doi: 10.18653/v1/2026.acl-long.747. URL `https://aclanthology.org/2026.acl-long.747/`.

Mengyu Chen, Feiyu Lu, Chun-Fu Chen, Lucas Vinh Tran, and Jay Katukuri. Inverse theory of mind modeling for content recommendation: From web browsing to dynamic intelligent interfaces. *arXiv preprint arXiv:2608.11354*, 2026. URL `https://arxiv.org/abs/2608.11354`.

DeepSeek-AI. DeepSeek-V4: Towards highly efficient million-token context intelligence. *arXiv preprint arXiv:2606.19348*, 2026. URL https://arxiv.org/abs/2606.19348.

DeepSeek-AI, Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, et al. Deepseek-v3.2: Pushing the frontier of open large language models. *arXiv preprint arXiv:2512.02556*, 2025. URL https://arxiv.org/abs/2512.02556.

Danica Dillion, Niket Tandon, Yuling Gu, and Kurt Gray. Can AI language models replace human participants? *Trends in Cognitive Sciences*, 27(7):597–600, 2023. doi: 10.1016/j.tics.2023.04.008.

Hao Fei and Yiran Zhao. Mental world modeling. *arXiv preprint arXiv:2607.27201*, 2026. URL https://arxiv.org/abs/2607.27201.

Gemini Team, Google. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025. URL https://arxiv.org/abs/2507.06261.

GLM-5 Team, Aohan Zeng, Xin Lv, Zhenyu Hou, Zhengxiao Du, Qinkai Zheng, Bin Chen, Da Yin, Chendi Ge, Chengxing Xie, et al. GLM-5: From vibe coding to agentic engineering. *arXiv preprint arXiv:2602.15763*, 2026. URL https://arxiv.org/abs/2602.15763.

David Ha and Jürgen Schmidhuber. Recurrent world models facilitate policy evolution. In *Advances in Neural Information Processing Systems*, volume 31, pp. 2451–2463. Curran Associates, Inc., 2018.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2555–2565. PMLR, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

John J. Horton, Apostolos Filippas, and Benjamin S. Manning. Large language models as simulated economic agents: What can we learn from homo silicus? In *Proceedings of the 25th ACM Conference on Economics and Computation*, pp. 614–615. Association for Computing Machinery, 2024. doi: 10.1145/3670865.3673513.

Peixuan Hou, Bin Chen, Li He, Jian Xu, Bo Zheng, Xiuli Ma, and Guojie Song. Behavior2Value: Benchmarking and empowering LLMs for consumer value measurement from e-commerce behaviors. *arXiv preprint arXiv:2609.18203*, 2026. URL https://arxiv.org/abs/2609.18203.

Xu Huang, Jianxun Lian, Yuxuan Lei, Jing Yao, Defu Lian, and Xing Xie. Recommender AI agent: Integrating large language models for interactive recommendations. *ACM Transactions on Information Systems*, 43(4):96:1–96:33, 2025. doi: 10.1145/3731446.

Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. doi: 10.1109/18.61115.

Zhicheng Lin. Large language models as psychological simulators: A methodological guide. *Advances in Methods and Practices in Psychological Science*, 9(1), 2026. doi: 10.1177/25152459251410153.

Yuan Ling, Chunqing Yuan, Shujing Dong, Yongjian Yang, Nataraj Mocherla, and Ayush Goyal. ShopperBench: A benchmark for personalized shopping with persona-guided simulation. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 5: Industry Track)*, pp. 837–846, Rabat, Morocco, March 2026. Association for Computational Linguistics. doi: 10.18653/v1/2026.eacl-industry.62. URL https://aclanthology.org/2026.eacl-industry.62/.

Biyuan Liu, Daigang Xu, Lei Jiang, Wenjun Guo, and Ping Chen. Modeling the mental world for embodied ai: A comprehensive review. *arXiv preprint arXiv:2601.02378*, 2026a. URL https://arxiv.org/abs/2601.02378.

Hongyang Liu, Zhu Sun, Tianjun Wei, Yan Wang, Jiajie Zhu, and Xinghua Qu. Diagnostic-guided dynamic profile optimization for LLM-based user simulators in sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 15306–15314, 2026b. doi: 10.1609/aaai.v40i18.38556.

Yuxuan Lu, Jing Huang, Yan Han, Bingsheng Yao, Sisong Bei, Yaochen Xie, Yisi Sang, Qi He, and Dakuo Wang. Can LLM agents simulate multi-turn human behavior? evidence from real online customer behavior data. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 43961–43977, San Diego, California, United States, July 2026. Association for Computational Linguistics. doi: 10.18653/v1/2026.acl-long.2034. URL https://aclanthology.org/2026.acl-long.2034/.

Saab Mansour, Leonardo Perelli, Lorenzo Mainetti, George Davidson, and Stefano D'Amato. PAARS: Persona aligned agentic retail shoppers. In *Proceedings of the 1st Workshop for Research on Agent Language Models (REALM 2025)*, pp. 143–159, Vienna, Austria, July 2025. Association for Computational Linguistics. doi: 10.18653/v1/2025.realm-1.11. URL https://aclanthology.org/2025.realm-1.11/.

OpenAI. GPT-4o system card. OpenAI, August 2024. URL https://openai.com/index/gpt-4o-system-card/.

OpenAI. GPT-5 system card. OpenAI, August 2025. URL https://openai.com/index/gpt-5-system-card/.

Joon Sung Park, Joseph C. O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–22. Association for Computing Machinery, 2023. doi: 10.1145/3586183.3606763.

David Premack and Guy Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4):515–526, 1978. doi: 10.1017/S0140525X00076512.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, Timothy Lillicrap, and David Silver. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588 (7839):604–609, 2020. doi: 10.1038/s41586-020-03051-4.

Ziyi Wang, Yuxuan Lu, Yimeng Zhang, Jing Huang, and Dakuo Wang. Customer-R1: Personalized simulation of human behaviors via RL-based LLM agent in online shopping. *arXiv preprint arXiv:2510.07230*, 2025. URL https://arxiv.org/abs/2510.07230.

Ziyi Wang, Yuxuan Lu, Wenbo Li, Amirali Amini, Bo Sun, Yakov Bart, Weimin Lyu, Jiri Gesi, Tian Wang, Jing Huang, Yu Su, Upol Ehsan, Malihe Alikhani, Toby Jia-Jun Li, Lydia Chilton, and Dakuo Wang. OPeRA: A dataset of observation, persona, rationale, and action for evaluating LLMs on human online shopping behavior simulation. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 43942–43960, San Diego, California, United States, July 2026. Association for Computational Linguistics. doi: 10.18653/v1/2026.acl-long.2033. URL https://aclanthology.org/2026.acl-long.2033/.

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

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, et al. Qwen2.5 Technical Report. *arXiv preprint arXiv:2412.15115*, 2025b.

Yimeng Zhang, Tian Wang, Jiri Gesi, Ziyi Wang, Yuxuan Lu, Jiacheng Lin, Sinong Zhan, Vianne Gao, Ruochen Jiao, Junze Liu, Kun Qian, Yuxin Tang, Ran Xue, Houyu Zhang, Qingjun Cui, Yufan Guo, and Dakuo Wang. Shop-R1: Rewarding LLMs to simulate human behavior in online shopping via reinforcement learning. In *International Conference on Learning Representations*, 2026.

Zijian Zhang, Shuchang Liu, Ziru Liu, Rui Zhong, Qingpeng Cai, Xiangyu Zhao, Chunxu Zhang, Qidong Liu, and Peng Jiang. LLM-powered user simulator for recommender system. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 13339–13347, 2025. doi: 10.1609/aaai.v39i12.33456.

Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, and Zheyan Luo. Llamafactory: Unified efficient fine-tuning of 100+ language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pp. 400–410, Bangkok, Thailand, 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-demos.38.

APPENDIX

A SHOPMIND IMPLEMENTATION DETAILS

This section provides additional implementation details of mental world rollout, personalized consistency evaluation, Behavior Critic training, and human-aligned behavior selection.

A.1 MENTAL WORLD ROLLOUT

At each interaction step, ShopMind first constructs the current decision context $J_t = (P_u, T_u, O_t, H_t, M_t)$ and generates K candidate actions $\mathcal{C}_t = \{a_t^1, \ldots, a_t^K\}$. We use $K = 4$ in the default configuration. Candidate generation is constrained by the valid action space of the current environment. Actions requiring an explicit target must refer to an available product, option, or interface element. The candidate set is encouraged to cover different plausible directions of the current decision, such as acquiring additional information, comparing alternatives, progressing toward purchase, or terminating the session when appropriate.

For each candidate a_t^k, ShopMind creates an independent copy of the current environment and mental state and performs a counterfactual rollout of L steps. We use $L = 2$ by default. The resulting imagined trajectory is

$$\hat{\tau}_t^k = \left(a_t^k, \hat{O}_{t+1}^k, \hat{M}_{t+1}^k, \ldots, \hat{O}_{t+L}^k, \hat{M}_{t+L}^k\right).$$

At every rollout step, the imagined action first changes the copied environment. ShopMind then applies the same mental-state update mechanism to determine whether the new observation should change the imagined mental state. Subsequent behavior is generated conditioned on the resulting environment and mental state, allowing the initial candidate to be evaluated through the future trajectory it may induce. Candidate rollouts are completely isolated from one another. Environment changes or imagined mental states from one branch cannot modify another branch or the real interaction state. A rollout terminates when the maximum horizon is reached, the simulated environment terminates, or no valid continuation is available. Algorithm 1 summarizes the rollout procedure.

All main experiments use $K = 4$ candidate actions and a rollout horizon of $L = 2$ unless otherwise specified. The sensitivity to K and L is analyzed separately in Appendix B.

A.2 PERSONALIZED CONSISTENCY EVALUATION

For each imagined trajectory, ShopMind evaluates whether the resulting behavior remains consistent with the characteristics and current decision state of the specific user. Given

$$(P_u, T_u, M_t, H_t, \hat{\tau}_t^k),$$

an LLM-based evaluator assesses three complementary dimensions: profile alignment, mental-state coherence, and intention consistency.

Algorithm 1 Mental World Rollout

Require: Decision context J_t, candidate number K, rollout horizon L

Generate valid candidates $\mathcal{C}_t = \{a_t^1, \ldots, a_t^K\}$
for each candidate $a_t^k \in \mathcal{C}_t$ **do**
  Copy the current environment and mental state
  Set the first imagined action to a_t^k
  for $l = 0, \ldots, L-1$ **do**
    Execute the imagined action in the copied environment
    Obtain the imagined observation $\hat{O}_{t+l+1}^k$
    Apply the mental-state update gate
    if a mental-state update is triggered **then**
      Generate $\hat{M}_{t+l+1}^k$
    else
      Retain the preceding imagined mental state
    end if
    if the environment terminates or $l = L-1$ **then**
      break
    end if
    Generate the next valid rollout action
  end for
end for
return $\{\hat{\tau}_t^k\}_{k=1}^K$

Profile alignment. S_{profile}^k measures whether the imagined trajectory is consistent with the user's relatively stable characteristics, preferences, and behavioral tendencies. Explicit shopping constraints, such as budget and product requirements, are also considered when judging whether a trajectory is compatible with the user.

Mental-state coherence. S_{mental}^k measures whether the behaviors and mental-state transitions throughout the rollout remain coherent with the user's current beliefs and unresolved uncertainties. The evaluator considers whether each simulated transition can be reasonably supported by the information available along the imagined trajectory and whether the resulting mental-state evolution remains internally consistent.

Intention consistency. $S_{\text{intention}}^k$ measures whether the imagined behaviors are consistent with the user's current high-level intention and whether they meaningfully advance that intention. Since one intention may govern multiple low-level actions, exact action–intention matching is not required. For example, an intention to compare two shortlisted products may be realized through several information-seeking and navigation actions.

The three scores are aggregated as

$$s_t^{k,\text{pc}} = \lambda_p S_{\text{profile}}^k + \lambda_m S_{\text{mental}}^k + \lambda_i S_{\text{intention}}^k,$$

where we set

$$\lambda_p = 0.35, \qquad \lambda_m = 0.30, \qquad \lambda_i = 0.35.$$

A.3 Behavior Critic

Behavioral regularities can differ substantially across e-commerce environments. Different datasets may have different action spaces, interface structures, product exposure mechanisms, session patterns, and population-level action distributions. We therefore train a separate Behavior Critic for each evaluation dataset using its corresponding training interactions.

Specifically, experiments on OPeRA use a Critic trained on OPeRA behavioral trajectories, while ShopCART and Taobao use Critics trained on their respective training data. All dataset-specific Critics share the same backbone, fine-tuning strategy, pairwise training objective, and optimization configuration. No interactions from an evaluation test split are used to train its corresponding Critic.

Training data construction. For each real behavioral context h_t, the observed next action is treated as the positive example a^+. We construct alternative actions a^- under the same interaction prefix so that the Critic learns to distinguish the behavior actually observed in real trajectories from contextually mismatched alternatives. Four types of negative samples are used. The first keeps the observed action type but replaces its target with an incorrect product or SKU. The second uses a different plausible action on the same target. The third involves another product from the same category. The fourth uses a more broadly mismatched random action. Together, these negatives expose the Critic to both hard context-dependent alternatives and easier mismatched behaviors.

For each dataset-specific Critic, we sample $30{,}000$ training contexts, $5{,}000$ validation contexts, and $5{,}000$ test contexts from the corresponding dataset. Each context is paired with one observed positive action and four negative alternatives, yielding $120{,}000$, $20{,}000$, and $20{,}000$ positive–negative pairs for training, validation, and testing, respectively.

Model and training. We instantiate the Behavior Critic with `Qwen3-4B-Instruct-2507` and train it as a reward model with a learned scalar value head. LoRA is applied to all target modules with rank 16, scaling factor 32, and dropout 0.05. Each dataset-specific Critic is trained for three epochs with a learning rate of 1×10^{-4}. We use cosine learning-rate scheduling with 100 warmup steps, weight decay of 0.01, and a maximum gradient norm of 1.0. The per-device training batch size is 1 with 16 gradient accumulation steps. Training uses BF16 precision and gradient checkpointing, and the random seed is fixed to 42. Validation is performed after every training epoch.

A.4 Human-Aligned Behavior Selection

ShopMind combines Personalized Consistency and Behavioral Realism through a gated selection strategy. Candidate trajectories are first ranked by their personalized consistency scores $s_t^{k,\mathrm{pc}}$. Let the two highest scores be $s_t^{(1),\mathrm{pc}}$ and $s_t^{(2),\mathrm{pc}}$, with margin $\Delta_t = s_t^{(1),\mathrm{pc}} - s_t^{(2),\mathrm{pc}}$. We set $\tau_{\mathrm{margin}} = 0.15$, $\tau_{\mathrm{compat}} = 0.60$, and $\tau_{\mathrm{critic}} = 0.20$. When $\Delta_t \geq \tau_{\mathrm{margin}}$, the top-ranked candidate indicates a sufficiently clear user-specific preference and is selected unless its Behavior Critic score falls below τ_{critic}. When $\Delta_t < \tau_{\mathrm{margin}}$, we retain candidates whose personalized consistency scores satisfy $s_t^{k,\mathrm{pc}} \geq \tau_{\mathrm{compat}}$, forming

$$\mathcal{C}_t^{\mathrm{pc}} = \left\{ a_t^k : s_t^{k,\mathrm{pc}} \geq \tau_{\mathrm{compat}} \right\}.$$

The Behavior Critic then resolves the ambiguity by selecting

$$A_t = \arg\max_{a_t^k \in \mathcal{C}_t^{\mathrm{pc}}} C_\phi(H_t, O_t, a_t^k).$$

If no candidate satisfies the compatibility threshold, ShopMind falls back to the candidate with the highest Personalized Consistency score. Purchase and exit actions receive no special preference and follow the same grounding and selection rules as other candidate actions. Algorithm 2 summarizes the complete selection procedure.

This selection mechanism gives priority to clear user-specific tendencies when the Personalized Consistency signal is decisive, while using behavioral patterns learned from real consumer trajectories to resolve ambiguous choices. Personalized Consistency and the Behavior Critic therefore provide complementary signals for preserving personalization and behavioral realism.

B Hyperparameter Sensitivity

We further examine the sensitivity of ShopMind to two key hyperparameters: the number of candidate behaviors K and the rollout horizon L. These hyperparameters control the diversity of candidate actions considered at each decision step and the amount of future evolution used to evaluate each candidate, respectively. Unless otherwise specified, all other settings are fixed to the default configuration used in the main experiments. We report results on the Taobao dataset using the same metrics as in the main evaluation.

Algorithm 2 Human-Aligned Behavior Selection

Require: Candidate trajectories $\mathcal{T}_t$

Evaluate $s_t^{k,\mathrm{pc}}$ for all $\hat{\tau}_t^k \in \mathcal{T}_t$
Rank candidates by $s_t^{k,\mathrm{pc}}$
Let k_1 and k_2 denote the top two candidates
$\Delta_t \leftarrow s_t^{k_1,\mathrm{pc}} - s_t^{k_2,\mathrm{pc}}$
Compute $C_\phi(H_t, O_t, a_t^k)$ for all candidates
if $\Delta_t \geq \tau_{\mathrm{margin}}$ **and** $C_\phi(H_t, O_t, a_t^{k_1}) \geq \tau_{\mathrm{critic}}$ **then**
    return $a_t^{k_1}$
end if
$\mathcal{C}_t^{\mathrm{pc}} \leftarrow \left\{ a_t^k : s_t^{k,\mathrm{pc}} \geq \tau_{\mathrm{compat}} \right\}$
if $\mathcal{C}_t^{\mathrm{pc}} = \emptyset$ **then**
    return $a_t^{k_1}$
else
    return $\arg\max_{a_t^k \in \mathcal{C}_t^{\mathrm{pc}}} C_\phi(H_t, O_t, a_t^k)$
end if

Table 6: Sensitivity of ShopMind to the number of candidate behaviors K on Taobao logs. The rollout horizon is fixed at $L = 2$.

	Local Fidelity			Sequential Realism				Individual Consistency		
Candidates K	Exact Acc. ↑	Macro F1 ↑	Top-3 Acc. ↑	Trans. JS ↓	Len. Err. ↓	Buy Err. ↓	Action JS ↓	Indiv. JS ↓	Spearman ρ ↑	Δ JS ↑
$K = 2$	51.02	56.74	69.5	15.74	4.23	4.12	11.53	19.74	0.404	+3.91
$K = 4$	**53.58**	**59.02**	**72.4**	**12.18**	**3.62**	**2.84**	**9.41**	**18.63**	**0.421**	**+4.36**
$K = 6$	53.49	58.93	72.3	12.24	3.65	2.89	9.46	18.69	0.420	+4.33
$K = 8$	53.31	58.79	72.1	12.31	3.69	2.95	9.53	18.77	0.418	+4.28

B.1 Number of Candidate Behaviors

We first vary the number of candidate behaviors K while fixing the rollout horizon at $L = 2$. As shown in Table 6, increasing K from 2 to 4 consistently improves performance across the three evaluation dimensions. Increasing the candidate set beyond $K = 4$ provides only marginal changes and slightly degrades several metrics, suggesting that four candidates are sufficient to cover meaningful behavioral alternatives without introducing excessive redundancy. We therefore use $K = 4$ as the default setting.

B.2 Rollout Horizon

We next study the effect of the rollout horizon L while fixing the number of candidate behaviors at $K = 4$. As shown in Table 7, introducing future rollout substantially improves performance, particularly on trajectory-level metrics. Increasing the horizon from $L = 0$ to $L = 2$ steadily improves local fidelity, sequential realism, and individual consistency. Without rollout ($L = 0$), candidate selection relies only on the current decision context and performs substantially worse on sequential metrics. Extending the horizon beyond $L = 2$ provides little additional benefit and slightly degrades several metrics. This suggests that short-horizon rollout is sufficient to capture the near-term consequences of candidate behaviors, while longer rollouts may introduce accumulated simulation errors. We therefore use $L = 2$ in all main experiments.

C Mental State Reconstruction Details

ShopMind jointly generates behavioral trajectories and the latent mental-state evolution underlying these behaviors. These aligned trajectories provide supervision for Mental State Reconstruction (MSR), which learns to recover consumers' latent beliefs, uncertainties, and intentions from ob-

Table 7: Sensitivity of ShopMind to the rollout horizon L on Taobao logs. The number of candidate behaviors is fixed at $K = 4$.

	Local Fidelity			Sequential Realism				Individual Consistency		
Horizon L	**Exact Acc. ↑**	**Macro F1 ↑**	**Top-3 Acc. ↑**	**Trans. JS ↓**	**Len. Err. ↓**	**Buy Err. ↓**	**Action JS ↓**	**Indiv. JS ↓**	**Spearman ρ ↑**	**Δ JS ↑**
$L = 0$	50.84	56.47	68.9	22.63	5.41	7.38	15.27	20.12	0.397	+3.72
$L = 1$	52.31	57.78	70.7	16.37	4.28	4.63	11.76	19.37	0.409	+4.08
$\mathbf{L = 2}$	**53.58**	**59.02**	**72.4**	**12.18**	**3.62**	**2.84**	**9.41**	**18.63**	**0.421**	**+4.36**
$L = 3$	53.44	58.89	72.2	12.36	3.68	2.93	9.54	18.72	0.419	+4.31
$L = 5$	53.12	58.61	71.9	12.81	3.81	3.18	9.87	18.94	0.415	+4.21

served shopping behavior. This section describes the construction and quality control of the MSR training data, the supervision format, and the model training configuration.

C.1 Training Data Construction

Each MSR training instance contains a task instruction, a complete behavioral trajectory, and its aligned mental-state sequence. The instruction specifies the reconstruction objective, valid shopping stages, and structured output format. Given a trajectory $H_{1:T} = (A_1, O_1, \ldots, A_T, O_T)$, MSR learns to reconstruct the corresponding mental-state sequence $\mathcal{M}_{1:K} = (M_1, \ldots, M_K)$. Importantly, K need not equal T. Multiple consecutive low-level actions may share the same latent mental state, while a new state is introduced only when the trajectory exhibits a meaningful cognitive transition.

The input behavior history is a natural-language representation of the complete shopping trajectory. It preserves decision-relevant interactions such as search queries, product views, review inspection, specification checking, option selection, cart operations, and purchase-related behaviors, together with relevant product, price, and attribute information. Low-level logging fields that do not contribute to reconstructing the consumer's decision process are removed. This representation allows MSR to operate on behavioral information that can also be obtained from real e-commerce logs without depending on platform-specific event identifiers.

C.2 Quality Control

Synthetic mental states may contain inconsistencies even when the generated behavior itself appears plausible. We therefore apply a multi-stage quality control pipeline before using a trajectory for MSR training. The procedure combines deterministic validation, LLM-based semantic evaluation, and human review.

We first apply rule-based checks to verify the structural validity of each trajectory along three dimensions. First, the terminal outcome must be internally consistent: the episode must contain a nonempty behavior sequence and terminate with a valid purchase or exit action. For purchase episodes, the purchased product must also be identifiable. Second, when candidate planning records are available, the selected candidate must match the action actually executed in the environment. Third, mental-state transitions must remain consistent with the update gate: recorded transitions must align with the behavior trajectory, and the mental state cannot change when the update gate is closed. Any trajectory violating these deterministic constraints is rejected before semantic evaluation. Trajectories passing deterministic validation are further evaluated by an LLM-based quality judge along four dimensions: *task fit*, *decision consistency*, *state–action consistency*, and *evidence quality*. Task fit evaluates whether the final decision is compatible with the shopping task. Decision consistency assesses whether the purchased product is coherent with the beliefs, uncertainties, and preferences formed before purchase. State–action consistency examines whether the evolving intentions and mental states can reasonably explain subsequent user actions. Evidence quality evaluates whether mental-state judgments are grounded in information that has actually been observed rather than unsupported, future, or counterfactual evidence. The four dimensions contribute 30, 30, 25, and 15 points, respectively, giving a maximum quality score of 100:

$$S_{\text{quality}} = S_{\text{task}} + S_{\text{decision}} + S_{\text{state-action}} + S_{\text{evidence}}.$$

Table 8: Quality-control statistics for MSR training data. Acceptance at each stage is computed relative to its input.

Stage	Input	Accepted	Acceptance
Generated purchase trajectories	7,200	7,200	100.0%
Deterministic checks	7,200	6,550	91.0%
LLM quality evaluation	6,550	5,100	77.9%
Human review and final filtering	5,100	5,000	98.0%
Final MSR training set	–	**5,000**	**69.4%**[†]

[†] Overall retention relative to the initially generated purchase trajectories.

A trajectory passes the score-based criterion when

$$S_{\text{quality}} \geq 70.$$

In addition to the numerical score, several severe semantic inconsistencies are treated as hard failures. These include an unexplained conflict between the final purchase and the preceding mental state, a purchase immediately following an incompatible intention, unsupported information appearing in the mental trajectory, and leakage of evidence from unselected counterfactual rollouts. A trajectory with any such hard failure is rejected regardless of its aggregate score.

We further conduct human review on all trajectories that pass the automatic filtering stages. Human annotators examine whether the reconstructed beliefs are supported by the observed trajectory, whether unresolved uncertainties evolve plausibly, whether intentions explain subsequent behaviors, and whether mental-state transitions occur at meaningful decision boundaries.

Table 8 summarizes the quality-control statistics. Starting from approximately $7{,}200$ successfully generated purchase trajectories, deterministic validation retains $6{,}550$, and LLM-based evaluation further retains $5{,}100$. All $5{,}100$ LLM-accepted trajectories are then manually reviewed. After human validation, $5{,}000$ trajectories are retained as the final high-quality MSR training corpus.

C.3 Model Training

We initialize MSR from `Qwen3-8B` and fine-tune it using LoRA on the $5{,}000$ quality-checked behavior–mental-state trajectories. Training is distributed across 16 GPUs for $2{,}000$ optimization steps. The per-device batch size is 2, and gradients are accumulated over 8 steps. We use BF16 precision and a learning rate of 5×10^{-5}.

D Comparison Methods and Experimental Setup

This section describes the implementation of the comparison methods and the experimental protocols used to evaluate behavioral realism and individual consistency. All methods are evaluated under the same environment interface, data preprocessing pipeline, and structured action space whenever applicable.

D.1 Datasets and Evaluation Splits

We evaluate behavior simulation on OPeRA, ShopCART, and an in-house Taobao interaction dataset. For local behavioral fidelity and sequential realism, we follow the official train–test splits of OPeRA and ShopCART and construct a user-disjoint split for Taobao. The first interaction of each test session is provided as observed context, and evaluation starts from the second action. Dataset statistics are summarized in Table 9.

OPeRA contains 527 filtered sessions, including 437 training sessions and 90 test sessions. The test split contains 992 recorded actions, resulting in 902 evaluated next-action decisions after excluding the first action of each session. ShopCART contains $31{,}865$ sessions, including $31{,}508$ training sessions and 357 test sessions. Its test split contains $1{,}506$ actions and yields $1{,}149$ evaluated decisions.

Table 9: Dataset statistics used in the behavior simulation experiments.

Dataset	Sessions	Train	Test	Eval. Steps
OPeRA	527	437	90	902
ShopCART	31,865	31,508	357	1,149
Taobao	10,000	8,000	2,000	17,436

The Taobao dataset contains $10{,}000$ shopping sessions, with $8{,}000$ used for training and $2{,}000$ for testing. The test split contains $19{,}436$ actions, yielding $17{,}436$ evaluated decisions.

Individual consistency additionally requires multiple sessions associated with the same user. OPeRA provides stable cross-session user identities and is therefore used for this evaluation. ShopCART does not provide sufficiently stable cross-session user or persona identifiers and is excluded from the individual-consistency experiment.

D.2 Unified Input and Action Interface

The comparison methods are adapted to a common interaction interface. The primary backbone used in the comparison study is `Qwen2.5-7B-Instruct`, with a maximum context budget of $32{,}768$ tokens including the generated output. At each decision step, the model input may contain the method-specific system instruction, user profile or method-specific memory, preceding actions and associated rationales when required by the method, and the current webpage. Historical webpages are not additionally concatenated (`history_obs`=0); behavioral history is represented through the recorded action sequence and method-specific memory. Webpages are cleaned and compressed using the same preprocessing pipeline before being passed to the model. When the complete context exceeds the token budget, method-specific history is truncated while preserving the current observation. All methods produce actions in the same structured environment space. A `click` action specifies the target element name, an `input` action specifies both the target element and input text, and `terminate` ends the current session. Evaluation uses strict matching: `click` predictions must match the target element name, `input` predictions must match both the element and entered text, and termination actions are matched by action type.

D.3 Baseline Implementations

We compare ShopMind with zero-shot prompting, supervised fine-tuning, reasoning- and reinforcement-learning-based methods, and existing user simulators. For each baseline, we use the backbone model reported in the original work and preserve its core training or inference procedure whenever applicable. All methods are evaluated on the same three datasets using identical webpage preprocessing, context budgets, and a unified action interface. Unless otherwise specified, stochastic decoding uses temperature 0.7.

Zero-shot. The original backbone model directly predicts the next valid shopping action from the user context, interaction history, and current webpage without task-specific training.

SFT and SFT+Rationale. Both methods are trained separately on the training split of each evaluation dataset. SFT directly predicts the target action, while SFT+Rationale first generates a behavioral rationale and then predicts the action. We use the same optimization settings across datasets, with a global batch size of 64 and a learning rate of 2×10^{-5}. For individual-consistency evaluation, training is restricted to temporally preceding sessions to avoid future information leakage.

Shop-R1 and Customer-R1. We follow the reinforcement-learning procedures of the original methods and initialize them from the corresponding supervised checkpoint. Shop-R1 uses its hierarchical action reward with self-certainty, whereas Customer-R1 adopts a difficulty-weighted action reward. We use learning rates of 1×10^{-7} and 5×10^{-7}, respectively, with a KL coefficient of 0.001 and 64 sampled completions per update.

Table 10: CTR prediction results on real-world Taobao interaction logs. Best results within each backbone are highlighted in **bold**.

Backbone	Input	AUC ↑	GAUC ↑	PR-AUC ↑
DeepSeek-V4-Pro	Raw	0.6441	0.6456	0.1653
	Raw + Zero-shot Mental State	0.6081	0.6090	0.1597
	Raw + MSR Mental State	**0.7156**	**0.7057**	**0.1924**
Qwen2.5-72B	Raw	0.6317	0.6332	0.1628
	Raw + Zero-shot Mental State	0.5964	0.5981	0.1491
	Raw + MSR Mental State	**0.6968**	**0.6895**	**0.1857**

AlignUSER. We adapt AlignUSER following its two-stage training procedure, combining world-model learning with behavior cloning and counterfactual reflection. The method is trained separately on the available training data of each evaluation dataset, with two epochs per stage and a learning rate of 1×10^{-5}.

ContextSim. We implement the policy variant of ContextSim using LoRA with rank 8, scaling factor 16, and dropout 0.05. Models are trained separately on the three evaluation datasets for five epochs with a learning rate of 1×10^{-5}. For individual-consistency evaluation, only temporally preceding training data are used. At inference time, the reasoning-action variant uses temperature 1.0 and top-$p = 0.95$.

SimUSER, DGDPO, and IToM. We use project-adapted implementations while preserving their method-specific prompting, state modeling, and decision procedures. Each method is applied consistently across the three evaluation datasets, and its outputs are normalized to the shared action representation for fair comparison.

E Downstream Task Implementation Details

E.1 Consumption Value Measurement

We consider two variants that differ only in how the mental-state sequence is obtained.

B2V+MSR Mental State. For this variant, the mental-state sequence is reconstructed using the trained MSR model:

$$\hat{\mathcal{M}}^{\mathrm{MSR}} = \mathrm{MSR}(H).$$

The reconstructed beliefs, uncertainties, and intentions are concatenated with the corresponding behavior trajectory and paired with the B2V-Bench value annotation. These augmented training instances are then used to fine-tune `Qwen3-8B` with LoRA for consumption value prediction. The model is trained for $1{,}500$ optimization steps on 16 GPUs. We use a per-device batch size of 1, gradient accumulation over 4 steps, BF16 precision, and a learning rate of 5×10^{-5}.

B2V+Zero-shot Mental State. To isolate the benefit of MSR from the general benefit of explicitly providing mental-state information, we construct a second variant using zero-shot mental-state reconstruction. Instead of MSR, the original `Qwen3-8B` model directly reconstructs the mental-state sequence from the behavior trajectory without task-specific MSR fine-tuning:

$$\hat{\mathcal{M}}^{\mathrm{ZS}} = \mathrm{Qwen3\text{-}8B}_{\text{zero-shot}}(H).$$

The reconstructed mental states are concatenated with the same behavioral trajectory and paired with the same value labels:

$$[H; \hat{\mathcal{M}}^{\mathrm{ZS}}] \longrightarrow V.$$

We then fine-tune another `Qwen3-8B` value prediction model using the same LoRA training setup as B2V+MSR Mental. This controlled comparison keeps the downstream value predictor, behavioral input, value supervision, and optimization configuration fixed, while changing only the source of the reconstructed mental states. It therefore directly evaluates whether mental states reconstructed by MSR provide more useful representations for consumption value measurement than those obtained from a general-purpose LLM in a zero-shot setting.

E.2 CTR PREDICTION

Dataset construction. We construct the CTR dataset from real-world Taobao interaction logs. For each shopping session, we randomly select a product-click event after the first interaction as the prediction point, ensuring that a nonempty behavioral history is available.

Let t_c denote the cutoff time of the selected click. The model observes only interactions before this point, i.e., $H_{<t_c} = (A_t, O_t) : t < t_c$. The click at t_c and all subsequent interactions are excluded, preventing future-information leakage. The product clicked at t_c serves as the positive candidate. Negative candidates are drawn from two sources: (1) other products appearing in the same session, excluding the positive item, and (2) products from the same second-level category that do not appear in the session, with at most four category-matched negatives per prediction point. All candidates for the same prediction point therefore share an identical behavioral history and differ only in candidate-product information and click labels. Each candidate is represented by its title, price, category, and shop information. A prediction instance is defined as

$$x_{u,j} = (H_{<t_c}, q_j), \qquad y_{u,j} \in \{0, 1\},$$

where q_j denotes the candidate product and $y_{u,j} = 1$ only for the product clicked at t_c. The behavioral history preserves decision-relevant information, including product names, prices, shops, and interaction types, while the click label is used only as supervision and never exposed to the model input.

Mental-state augmentation. We use the same controlled construction as in the consumption value measurement experiment. Starting from the same truncated behavior history, we construct two mental-state-augmented variants that differ only in how the mental states are reconstructed.

For the zero-shot variant, we directly prompt the original `Qwen3-8B` model, without MSR-specific fine-tuning, to reconstruct the mental-state trajectory:

$$\hat{\mathcal{M}}^{\mathrm{ZS}}_{<t_c} = \mathrm{Qwen3\text{-}8B}_{\text{zero-shot}}(H_{<t_c}).$$

For the MSR variant, we use the trained Mental State Reconstruction model:

$$\hat{\mathcal{M}}^{\mathrm{MSR}}_{<t_c} = \mathrm{MSR}(H_{<t_c}).$$

Both reconstruction methods follow the same mental-state schema, describing the evolution of shopping stage, beliefs, uncertainties, and intentions before the prediction point. The resulting mental-state sequence is concatenated with the original behavioral history.

We therefore evaluate three input conditions:

$$\begin{aligned} \text{Raw}: &\quad H_{<t_c}, \\ \text{Raw + Zero-shot Mental State}: &\quad [H_{<t_c}; \hat{\mathcal{M}}^{\mathrm{ZS}}_{<t_c}], \\ \text{Raw + MSR Mental State}: &\quad [H_{<t_c}; \hat{\mathcal{M}}^{\mathrm{MSR}}_{<t_c}]. \end{aligned}$$

The three conditions use exactly the same behavioral histories, candidate products, and click labels. Thus, the comparison isolates both the incremental value of explicit mental-state information and the quality of the mental-state reconstruction. In particular, "Zero-shot Mental State" refers to zero-shot *mental-state reconstruction*, rather than zero-shot CTR prediction.

Results. Table 10 reports the CTR prediction results. Adding zero-shot reconstructed mental states does not improve the Raw baseline, indicating that additional mental-state text alone is insufficient. In contrast, MSR-reconstructed mental states consistently improve all three ranking metrics for both prediction backbones.