Recommendation-as-Experience: A Framework for Context-Sensitive Adaptation in Conversational Recommender Systems

RAJ MAHMUD, University of Technology Sydney, Australia
SHLOMO BERKOVSKY, Macquarie University, Australia
MUKESH PRASAD, University of Technology Sydney, Australia
A. BAKI KOCABALLI, University of Technology Sydney, Australia

Modern conversational recommender systems excel at technical optimisation but often fail to operationalise the latent experiential dimensions that govern human decision-making. Current architectures predominantly prioritise recommendation accuracy, frequently neglecting the nuanced, context-sensitive behaviours required for meaningful interaction. We introduce Recommendation-as-Experience (RAE), a framework designed to bridge this gap by encoding experiential objectives as adaptive state variables. Through a multi-domain study ($N = 168$), we employ Bayesian hierarchical ordinal regression to quantify the interplay between three primary interactional pillars: educative (justification and transparency), explorative (discovery and serendipity), and affective (social and emotional resonance). Our findings demonstrate that domain-specific profiles and perceived item value function as systematic modulators of these priorities. Notably, we identify a strong preference for user autonomy within the tested apparel e-commerce context, suggesting that interactional agency is an important component of the conversational experience, though further cross-domain validation is required for higher-stakes scenarios. Drawing on these empirical insights, the RAE framework formalises the mapping of contextual and individual signals onto structured state representations. These representations drive experience-aligned dialogue policies through retrieval diversification, heuristic constraints, or controllable Large Language Model generation. RAE framework provides a principled approach for balancing predictive precision with high-fidelity experiential quality.

CCS Concepts: • **Information systems** → **Recommender systems**; • **Computing methodologies** → *Natural language processing*; • **Human-centered computing**;

Additional Key Words and Phrases: Adaptive conversational recommender system, dialogue management, personalised interaction

ACM Reference Format:
Raj Mahmud, Shlomo Berkovsky, Mukesh Prasad, and A. Baki Kocaballi. 2027. Recommendation-as-Experience: A Framework for Context-Sensitive Adaptation in Conversational Recommender Systems. *ACM Trans. Recomm. Syst.* 5, 2, Article 29 (March 2027), 32 pages. https://doi.org/10.1145/3838803

Authors' Contact Information: Raj Mahmud (corresponding author), School of Computer Science, University of Technology Sydney, Sydney, New South Wales, Australia; e-mail: raj.pgm@outlook.com; Shlomo Berkovsky, Australian Institute of Health Innovation, Macquarie University, Sydney, New South Wales, Australia; e-mail: shlomo.berkovsky@mq.edu.au; Mukesh Prasad, School of Computer Science, University of Technology Sydney, Sydney, New South Wales, Australia; e-mail: mukesh.prasad@uts.edu.au; A. Baki Kocaballi, School of Computer Science, University of Technology Sydney, Sydney, New South Wales, Australia; e-mail: baki.kocaballi@uts.edu.au.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
© 2027 Copyright held by the owner/author(s).
ACM 2770-6699/2027/03-ART29
https://doi.org/10.1145/3838803

1 Introduction

Conversational Recommender Systems (**CRSs**) have advanced considerably, combining natural language understanding and generation with retrieval, ranking, and adaptive dialogue policies to support multi-turn, context-aware recommendation [26]. Recent pipelines can elicit preferences, justify suggestions, and present relevant alternatives across diverse domains such as e-commerce, education, entertainment, and finance [17, 56]. The emergence of **Large Language Models** (**LLMs**) has significantly extended the functional repertoire of these systems; however, the core principles of adaptive recommendation, established well before the current generative AI paradigm, continue to dictate the efficacy of rule-based, statistical, reinforcement-learning, and hybrid approaches [11, 86]. Accordingly, our focus transcends specific modelling paradigms to address the core question of how adaptive CRS architectures can be systematically aligned with multi-faceted user experience (UX) requirements.

Despite these technical strides, much CRS research remains architecture-centric, primarily emphasising ranking accuracy, intent parsing, or surface-level language fluency [27]. While these dimensions are essential for performance, they often under-represent the fine-grained conversational behaviours and their longitudinal effects on user experience, particularly in systems capable of generating open-ended, non-deterministic content [12]. Recent user-centred evaluations have consequently underscored the necessity of moving beyond coarse outcome measures such as monolithic satisfaction or trust scores toward constructs more specific to the nuances of conversational agency, including perceived understanding, response quality, humanness, and user control [29, 47]. This paradigm shift reflects a broader consensus within both the recommender systems and **Human–Computer Interaction** (**HCI**) communities: system effectiveness cannot be conceptually or empirically decoupled from experiential quality.

Building on these observations, we propose the **Recommendation-as-Experience** (**RAE**) framework, which conceptualises recommendation as an unfolding, situated interaction rather than solely a prediction task. The RAE perspective draws upon foundations in experience-centred HCI [21], exploratory search [49], and information-seeking theories that foreground the pivotal roles of affect and uncertainty [38]. This framing is strategically aligned with adaptive dialogue management approaches that dynamically adjust strategies in response to latent user goals, emotional cues, and situational contexts [3, 25, 46]. Within this framework, user experience is operationalised through three complementary *interaction aims*: *educative aims*, which inform and justify to support transparency and decision confidence [66, 84]; *explorative aims*, which inspire and diversify to encourage novelty and serendipity [6, 34, 72]; and *affective aims*, which align emotionally and socially to foster empathy and rapport during the interaction [53, 59, 83]. While each of these aims has been examined in isolation, there remains a substantial gap in our understanding of how users evaluate these dimensions *jointly*, or how their relative importance is modulated by domain context, perceived item value, and user-level characteristics. For instance, educative support may dominate in cognitively complex, high-stakes domains such as finance or housing, whereas affective tone may become the primary driver of UX in wellness or mental-health contexts [64]. Moreover, individual and situational factors including prior CRS experience and specific preferences for dialogue control are known to significantly moderate UX and thus warrant explicit integration into adaptive policy design [28, 31, 57, 78].

To investigate these intersections, we conducted a ten-domain study involving $N = 168$ participants, measuring the perceived importance of educative, explorative, and affective aims under varying domain complexities and situational stakes. These predictors correspond to signals that a CRS can feasibly capture at runtime, such as domain metadata, item-level attributes, and persistent user traits. Our analysis provides the empirical grounding necessary to transition from static interfaces to adaptive policies that prioritise specific interaction aims in a context-sensitive manner.

The contributions of this article are four-fold. First, we provide a multi-domain empirical analysis that quantifies the joint prioritisation of interaction aims across 10 distinct CRS contexts. Second, we present empirical evidence that contextual factors including domain-specific cognitive complexity, novelty orientation, and emotional salience systematically moderate user expectations. Third, we identify key user-level influences, establishing prior CRS experience and autonomy preferences as significant modulators while demonstrating that broader demographic effects are largely contingent on specific domain contexts. Finally, we introduce the *RAE adaptation framework*, which provides a formal mapping between structured state signals (domain, value, and user traits) and adaptive interaction policies. By grounding these policies in validated user preferences, the framework advances both the selection of interaction aims (*what* to foreground) and the technical mechanisms of control (*how* to adapt them) for personalised, context-sensitive recommendation.

2 Background

Prior research on CRS spans multiple domains, including conversational agents, recommender systems, and HCI. To situate our contribution, we review four complementary strands: (1) user experience constructs in recommender systems, (2) interaction aims conceptualised as experiential goals, (3) user-level differences and dialogue initiative, and (4) adaptive dialogue policies and orchestration. These research streams reveal both the advances and limitations of existing work, motivating our study and the RAE adaptation framework. Table 1 synthesises representative contributions and delineates research gaps that our study addresses.

User experience constructs in CRS. Although CRS research increasingly acknowledges the importance of user experience (UX), most systems remain architecture-centric, optimising ranking accuracy, intent recognition, or language fluency. A growing body of work has expanded evaluation beyond these technical metrics to include constructs such as satisfaction, trust, engagement, and perceived usefulness [26, 48, 59]. Instruments such as ResQue [57] and CRS-Que [29] formalise multiple UX dimensions, including informativeness, novelty, transparency, trust, and engagement. These frameworks demonstrate that UX qualities matter for system acceptance and continued use, but they are typically employed as *static outcome criteria* in post-hoc evaluations. By contrast, our work treats UX constructs as *dynamic signals* that can be modelled at runtime and incorporated into adaptive state representations. In this perspective, qualities such as transparency or engagement are not only end goals for evaluation but also contextual features that can guide dialogue strategies, retrieval choices, and explanation depth. This reframing moves UX constructs from being purely diagnostic to being operational drivers of adaptive CRS policies. A recent systematic review of 23 empirical studies across classical and LLM-powered systems found that CRS evaluation remains fragmented, with a significant gap in linking adaptive behaviours to specific UX outcomes. This lack of turn-level affective assessment and the reliance on static post-hoc surveys necessitates a more dynamic approach to experience modelling. The RAE framework addresses this by operationalising these evaluation gaps into active adaptation drivers [47].

Interaction aims as experiential goals. Building on this perspective, recent work has identified specific experiential goals that map to concrete system behaviours. Educative aims emphasise explanations, contextual guidance, and contrastive comparisons, which improve transparency and decision confidence [35, 70]. Explorative aims foreground novelty and serendipity, encouraging curiosity-driven discovery and preference formation [10, 34, 49]. Affective aims target rapport, empathy, and sentiment-aware alignment, shown to increase user trust, engagement, and relational satisfaction in domains such as wellness and beauty [53, 59, 63]. While these aims have each been explored in CRS research, they are typically examined in isolation, either as one-off system features

Table 1. Taxonomy of Research Streams in Conversational Recommender Systems (CRS) and Identified Gaps

Research Stream	Typical Emphasis and Identified Research Gap
User Experience Constructs	Focus on measurement instruments (e.g., ResQue [57], CRS-Que [29]) and outcomes such as trust, engagement, and perceived usefulness [26, 48]. *Gap: Constructs are typically treated as static outcomes rather than dynamic state features for real-time adaptation.*
Interaction Aims	Educative (explanations, trade-off comparisons) [35, 70]; Explorative (diversity, novelty) [10, 34]; and Affective (tone, empathy) support [53, 59]. *Gap: The joint prioritisation and trade-offs between these three aims across different decision domains remains largely unstudied.*
User Traits and Initiative	Impact of demographics (age, gender) [24, 60] and prior CRS experience [20, 52] on dialogue initiative preferences [45, 62]. *Gap: Insufficient evidence regarding how individual user traits and initiative preferences should modulate specific interactional aims.*
Adaptive Policies	Classical RL (POMDP) [80], hierarchical efficiency [79], and LLM orchestration for controllable generation [22, 44]. *Gap: Current models prioritise transactional efficiency or retrieval accuracy but rarely encode multi-dimensional experiential aims into state–policy mappings.*

The present study addresses these limitations by modelling the joint prioritisation of interactional aims and formalising a state–policy–behaviour framework for adaptive, experience-centric CRS design.

or as independent evaluation dimensions. This fragmentation makes it unclear how aims interact or compete when users face complex decision contexts. Beyond CRS, educational tutors such as ArgueTutor demonstrate that educative, explorative, and affective behaviours can be orchestrated jointly and tailored dynamically to learning goals [69]. However, CRS research has yet to establish how users prioritise these aims when they co-occur, or how contextual and individual factors (e.g., domain, task value, user traits) shape their relative salience. Our study addresses this gap by quantifying how users jointly value educative, explorative, and affective aims across multiple domains, and by identifying moderators that systematically influence their prioritisation. This contributes empirical grounding for adaptive policies that move beyond single-aim optimisation toward holistic experiential alignment. Prior empirical work has demonstrated that users equate 'usefulness' and 'effectiveness' in CRS with informational depth and conversational richness, rather than just transactional speed. Furthermore, clustering analyses identified five distinct user profiles with varying preferences for exploratory vs. task-oriented styles. RAE extends this binary style contrast into a tripartite aim structure (Educative, Explorative, Affective) to better capture the nuanced needs of these profiles across diverse domains [46].

User-level differences and dialogue initiative. A further gap concerns how user traits and initiative preferences shape experiential priorities in CRS. Demographic moderators such as age and gender have been shown to affect preferences for explanation, transparency, or affective support [24, 54, 60, 85]. Prior CRS experience also conditions expectations: experienced users often demand greater explanation depth, autonomy, and efficiency, while novice users may require more guidance and reassurance [20, 33, 52]. Despite this, few CRS studies stratify samples by experience or test how familiarity systematically modulates the value assigned to different interaction aims [40]. Dialogue initiative introduces an additional dimension of variability. Mismatches between a user's preferred

level of initiative (system-led vs. user-led) and the initiative enacted by the CRS reduce trust, usability, and perceived alignment [36, 45, 62]. To address this, prior literature has explored *mixed-initiative* interaction, wherein control fluidly shifts between the user and the system depending on task ambiguity and user expertise [30]. Reviews of dialogue management further identify initiative handling as a persistent challenge for adaptive systems [4]. These findings indicate that adaptive CRS must not only respond to contextual signals such as domain and task value, but also incorporate user-level traits and initiative preferences into their interaction policies. Our study contributes by testing how demographic factors, prior CRS experience, and autonomy preferences influence the prioritisation of educative, explorative, and affective aims across domains.

Adaptive dialogue policies and orchestration. Adaptive dialogue policies provide the computational machinery that maps contextual and user signals into behaviour selection. Classical approaches employed POMDPs and policy gradient methods to optimise dialogue under uncertainty [65, 80], and CRS-specific pipelines interleaved preference elicitation with item retrieval [11]. Reinforcement learning has been used to optimise long-term engagement, efficiency, and reward-driven outcomes [17, 79], with recent advances introducing multi-round feedback adaptation [76], graph-based policy unification [7], and knowledge-grounded CRS with policy learning [13]. More recently, LLM-assisted orchestration integrates retrieval, reasoning, and controllable generation for adaptive recommendation [44, 77], with surveys emphasising the importance of explicit orchestration layers and state representations [74]. These approaches illustrate the growing sophistication of adaptive dialogue policies, but most continue to optimise for efficiency or accuracy, rarely encoding experiential aims as first-class state variables. Our work extends this trajectory by formalising a *state–policy–behaviour* mapping in which educative, explorative, and affective aims are explicitly represented as adaptive features. This shift reframes dialogue management from a performance-centric optimisation task to one that also foregrounds experiential quality, bridging system-level orchestration with validated user preferences.

3 Method

We conducted a cross-sectional online survey to examine how users prioritise *educative*, *explorative*, and *affective* interaction aims across multiple CRS domains, and how these priorities vary with perceived item value, user characteristics, and dialogue-control preferences. The measured variables align with features that can be operationalised in a CRS *state representation*, enabling rule-based or learned policies in LLM-powered and other adaptive CRS architectures to adjust behaviours in real time.

3.1 Study Design

To evaluate user expectations and autonomy preferences, we conducted a multi-domain online experiment, utilising a within-subjects vignette-based survey design (Figure 1) for item value manipulation, to capture how interactional aims shift across varying contexts. This approach was selected over a functional system deployment to ensure comparative generalisability across distinct cognitive, emotional, and social contexts while avoiding the technical confounds associated with specific system architectures. Our design addresses three primary research questions operationalised through six hypotheses as described in Section 3.2.

Domain Selection and Categorisation. Ten domains were selected based on their prevalence in current CRS deployments and their representation in recent state-of-the-art reviews [27, 29, 56]. To ensure a broad coverage of the recommendation landscape, we categorised these domains into five functional groups:

- High-stakes and Complex: Finance, Housing, and Education.
- Hedonic and Leisure: Travel, Apparel, and Entertainment.

Survey Instrument: Educative Construct

Note: Prior to this section, participants were primed with side-by-side conversation transcripts between a user and a Recommender Chatbot to ground their mental models (Survey Section 3).

[Phase 1: Construct Definition & Baseline Anchoring]
Q4.1. Considering the 'Educative' feature of a Recommender Chatbot—which includes providing detailed information about products, explaining the reasons behind specific recommendations, and offering guidance to help you make informed decisions—how would you rate its importance in the context of online apparel shopping?

○ Not important at all ○ Slightly ○ Moderately ○ Very ○ Extremely important

[Phase 2: Dynamic Value Sensitivity (Within-Subjects Delta)]
Q4.3. Would you change your rating on the previous question (Q4.1) if it was a **higher value item** you were seeking recommendation on?

○ Significantly decrease my rating
○ Slightly decrease my rating
○ No change to my rating
○ Slightly increase my rating
○ Significantly increase my rating

[Phase 3: Dialogue Initiative & Autonomy Allocation]
Q4.4. How do you prefer to receive additional, educational information from the Recommender Chatbot: should the chatbot proactively offer this information, or would you rather request it yourself when needed?

○ Mostly system initiated ○ Balanced ○ Mostly user initiated

Fig. 1. The architectural flow of the vignette-based survey instrument (excerpted from the Educative aim module). To mitigate respondent fatigue while maintaining construct validity, the survey explicitly defines the interactional aim (Phase 1), captures relative shifts in user expectations based on situational stakes (Phase 2), and extracts specific autonomy preferences to govern dialogue control (Phase 3).

- Affect-rich and Identity-based: Wellness and Beauty.
- Functional and Technical: Technology (consumer electronics).
- Service and Hospitality: Dining.

Sample Size and Power Analysis. An *a priori* power analysis was conducted using G*Power 3.1 [15]. To detect a medium effect size ($f = 0.25$) in an omnibus test with a significance level of $\alpha = .05$ and power of $1 - \beta = .80$, a minimum sample of $N = 159$ was required. While our primary analyses utilise non-parametric and Bayesian ordinal models, we adopted this target as a conservative baseline for ensuring sufficient sensitivity to detect cross-domain differences.

Methodological Rationale. By adopting a vignette-based elicitation approach, the study identifies framework-level expectations that are architecture-agnostic. This design allows for the isolation of the RAE state variables (domain, user traits, and item value) without interference from dialogue-system latency, natural language understanding errors, or specific interface aesthetics, thereby grounding the framework in high-level user requirements.

3.2 Research Hypotheses

Based on the proposed RAE framework, we formulated six hypotheses linking contextual and individual factors to the perceived importance of interactional aims. These hypotheses focus on signals that a CRS can dynamically capture as state variables to drive adaptive policies.

RQ1: Domain effects on interactional aims. The first set of hypotheses examines how the inherent characteristics of a decision domain shape baseline user expectations:

H1: Perceived importance of the *educative* aim is significantly moderated by domain-specific task complexity.

H2: Perceived importance of the *explorative* aim is significantly moderated by domain-specific novelty orientation.

H3: Perceived importance of the *affective* aim is significantly moderated by domain-specific emotional relevance.

RQ2: Moderation by item value and user traits. The second set of hypotheses investigates how situational stakes and individual differences modulate the baseline domain effects:

H4: Situational *item value* significantly impacts user prioritisation of educative, explorative, and affective aims, with higher stakes increasing demand across all dimensions.

H5: User characteristics–specifically *age*, *gender*, and *CRS experience*–significantly moderate the perceived importance of interactional aims.

RQ3: Dialogue-control preferences. The final set of hypotheses explores the interactional agency users desire during these exchanges:

H6a: Users exhibit a significant preference for *user-initiated control* during both educative and explorative interaction phases (rated significantly above the neutral midpoint).

H6b: Demographic traits and prior experience are significantly associated with the degree of control-preference for these interactional aims.

3.3 Survey Instrument and Measures

The research instrument was developed as a structured digital questionnaire designed to capture multi-dimensional user expectations across diverse recommendation contexts. The instrument was grounded in established user-centric evaluation frameworks for recommender systems [31, 57] and specific metrics for conversational interaction [29, 47], ensuring compatibility with adaptive design parameters and multi-dimensional UX requirements [10, 84].

Elicitation of interactional aims. Central to the instrument was the elicitation of latent user preferences across three interactional aims: *Educative* (the requirement for transparent rationale), *Explorative* (the demand for discovery and serendipity), and *Affective* (the expectation for emotional and social alignment). These constructs were operationalised as single-item measures to ensure high completion quality and mitigate respondent fatigue across the intensive 10-domain block. For example, the educative aim was operationalised through the item: *'How important is it that the system explains why these options were chosen?'* (measured on a 5-point Likert scale). To neutralise potential order effects, both the sequence of domains and the presentation of interactional aims were fully randomised for each participant.

Situational item-value manipulation. To examine the within-subjects effect of situational stakes (H4), the instrument featured contrasted vignettes. Participants were presented with situated scenarios that juxtaposed low-stakes contexts (e.g., a casual $10 purchase) against high-stakes conditions (e.g., a significant $1,000 investment). This within-subjects design ensured that 'perceived value' was anchored in realistic, domain-specific consumer behaviours rather than abstract estimation, allowing for the isolation of value-sensitivity as a dynamic modulator.

User traits and autonomy preferences. To study demographic covariates and dialogue-control expectations, we recorded age group, gender identity, and prior CRS experience, the latter measured on a 5-point frequency scale. Furthermore, dialogue-control preferences for *educative* and *explorative* aims were elicited using a 5-point bipolar scale (ranging from 1 = *fully system-initiated* to 5 = *fully user-initiated*). These items were framed within a familiar apparel e-commerce context to provide

a stable reference point for assessing autonomy. Consistent with our theoretical framework, no autonomy item was collected for the affective aim, as affective support is conceptually defined as a system-led response to implicit user cues. The complete instrument underwent a pre-test phase ($n = 8$) to refine linguistic clarity and cognitive load. Items were strategically adapted from *CRS-Que* [29] and validated UX metrics [10, 59, 84] to maintain high construct validity.

Participant recruitment and data refinement. We recruited a total of $N = 191$ participants through a multi-channel strategy including Prolific ($n = 149$), Survey Swap ($n = 33$), and targeted social media recruitment ($n = 9$). Eligibility was restricted to individuals aged 18 or older with verified English proficiency. Prolific participants were further screened to include only those with a high historical performance record ($\geq 95\%$ approval rating and ≥ 100 completed studies). The survey was hosted on the Qualtrics platform, with an average completion time of 9 min; Prolific participants were compensated $2.40, a rate consistent with institutional and platform ethical guidelines. To ensure data integrity, we applied rigorous cleansing procedures to identify Insufficient Effort Responding [50]. A total of 23 cases were excluded based on the following pre-defined criteria: failing the embedded attention checks (9 cases), exhibiting straight-lining behaviour with a standard deviation across Likert blocks of $SD < 0.2$ (7 cases), or presenting implausibly rapid completion times of less than 3 min (7 cases). The final analytic sample comprised $n = 168$ valid respondents. This sample size retained sufficient statistical power for the planned frequentist and Bayesian analyses and provided a diverse distribution of age, gender, and CRS experience necessary for robust moderator analysis.

3.4 Data Analysis

Analyses were conducted in Python 3.11 (`pandas 2.2`, `scipy 1.12`, `statsmodels 0.15`); Bayesian models were fitted in `Stan 2.33` via `cmdstanpy`. Ordinal responses (1–5) were analysed using non-parametric tests or ordinal regression as appropriate. All tests were two-sided with $\alpha = .05$. Missing data were handled listwise *per analysis* (i.e., pairwise inclusion), and sample sizes (n) are reported with each result. For H6 regressions, the effective model sample was $N = 164$ after listwise inclusion of covariates.

Domain effects (H1–H3). Kruskal–Wallis tests [37] were followed, where appropriate, by Bonferroni-adjusted Mann–Whitney U comparisons. We also fitted hierarchical Bayesian cumulative-logit models [19] with domain as a fixed factor and participant-level random intercepts to estimate domain-level uncertainty. Pairwise effect sizes are reported as $r = |z|/\sqrt{n}$.

Item value (H4). Within-subject differences between low- and high-value frames were tested with Wilcoxon signed-rank tests [73]. For transparency, we report Wilcoxon W, the non-tied sample size n' (ratings equal to the midpoint are dropped by the test when applicable), p-values, and effect sizes ($r = |z|/\sqrt{n'}$). Rank-biserial and **common-language effect size** (**CLES**) are provided.

Demographics (H5). Zero-order associations were assessed with Spearman rank correlations; the Benjamini–Hochberg procedure [2] controlled the false discovery rate at $q = .05$ *per predictor family* (30 correlations per demographic predictor). Bayesian cumulative-logit models with random domain intercepts provided pooled estimates across domains (weakly informative, zero-centred priors; diagnostics reported in the supplement).

Autonomy preferences (H6a–H6b). One-sample Wilcoxon signed-rank tests against the neutral midpoint (= 3) assessed preferences for user initiation in educative and explorative aims; ties at the midpoint were dropped by the test (n' reported). Separate cumulative-logit models [23] examined demographic predictors, with Gender coded Male= 1, Female= 0, Age treated as categorical

(reference = 18–24), and CRS experience modelled as ordinal (1–5). Proportional-odds assumptions and model fit were assessed via surrogate diagnostics and posterior predictive checks.

Inference criteria and diagnostics. Frequentist significance was defined as $p \leq .05$. For Bayesian models, effects were deemed credible when 95% posterior intervals excluded zero. Convergence and adequacy were assessed using $\hat{R}$, **effective sample sizes** (**ESS**), divergence checks, and posterior predictive fit. Given the use of single-item measures, we prioritised Bayesian credible intervals (95% HDI) over internal consistency metrics (e.g., Cronbach's α). The Bayesian approach provides a robust estimate of parameter uncertainty and effect stability, which serves as a statistically rigorous proxy for reliability in multi-domain elicitation studies.

Ethics. The study was approved by the authors' institutional Human Research Ethics Committee. All participants provided informed consent. No personally identifiable data were collected; data were anonymised at source and stored in compliance with GDPR and institutional data-governance policies.

Data and code availability. To ensure reproducibility, the anonymised dataset ($N = 168$), data processing scripts, and evaluation notebooks–encompassing all frequentist and Bayesian models–have been made publicly available in our GitHub repository: https://github.com/staraj1/RAE-Framework-Study.

4 Results

This section presents the empirical findings from our vignette-based elicitation study ($N = 168$). We begin by detailing the participant profile, followed by a systematic evaluation of our six hypotheses (H1–H6) using a combination of frequentist non-parametric tests and Bayesian hierarchical models. Collectively, the data validate the core premise of the RAE framework: interactional aims are significantly modulated by the decision domain, situational item value, and prior system experience. A consolidated summary of all hypothesis outcomes is provided in Table 10 at the end of this section.

4.1 Participant Demographics

Table 2 provides a breakdown of the study population. The sample primarily represents younger to middle-aged adults, with 73.2 % of participants aged 18–34 years. In terms of gender identity, the sample comprised 57.1 % female, 40.5 % male, and 1.8 % non-binary or other. Notably, the cohort was nearly evenly split regarding system familiarity: 52.4 % reported at least some prior experience with CRSs, while 47.6 % were novices. Within the RAE framework, these attributes serve as persistent user-state features that inform the baseline dialogue policy.

4.2 Domain Effects on Interaction Preferences (H1–H3)

All three interaction aims varied significantly across domains, as shown in Table 3, with Education, Travel, and Wellness emerging as the top-ranked domains for educative, explorative, and affective aims, respectively. Furthermore, we conducted Kruskal–Wallis H tests to evaluate domain-level differences for all three interaction aims. All three yielded significant omnibus effects: Educative ($H(9) = 93.15$, $p < .001$), Explorative ($H(9) = 33.34$, $p < .001$), and Affective ($H(9) = 25.10$, $p = .003$), indicating that users' prioritisation of each aim differed significantly by domain. Table 4 shows the test statistics and ranks the top and bottom three domains for each aim by mean rank. Follow-up pairwise comparisons using Bonferroni-adjusted Mann–Whitney U tests identified several domain pairs with meaningful differences ($r \geq 0.20$). As summarised in Table 5, the largest effect emerged between Education and Dining for the Educative aim ($r = .36$), supporting the hypothesis that

Table 2. Demographic Profile and Baseline System Experience of the Study Population ($N = 168$)

Characteristic	Category	n	%
Age Group	18–24 years	58	34.5
	25–34 years	65	38.7
	35–44 years	31	18.5
	45–54 years	8	4.8
	55–64 years	6	3.6
	65+ years	0	0.0
Gender Identity	Female	96	57.1
	Male	68	40.5
	Other (incl. non-binary)	3	1.8
	Prefer not to say	1	0.6
CRS Experience	None	80	47.6
	Some / Frequent	88	52.4

The sample represents a broad distribution of young and middle-aged adults, with a balanced split between users with and without prior Conversational Recommender System (CRS) experience.
Note: CRS Experience was self-reported based on prior interaction with AI-driven travel or shopping assistants. Percentages are calculated row-wise and may not sum to 100% due to rounding. CRS Experience was self-reported based on prior interaction with AI-driven travel or shopping assistants.

Table 3. Descriptive Statistics for Perceived Interactional Aims Across Ten Decision Domains ($N = 168$)

Domain	Educative			Explorative			Affective		
	Med.	Mode	IQR	Med.	Mode	IQR	Med.	Mode	IQR
Apparel	3.00	4	3–4	4.00	4	3–4	3.00	3	2–4
Beauty	4.00	4	3–4	3.00	4	3–4	3.00	4	2–4
Entertainment	3.00	4	3–4	4.00	4	3–4	3.00	3	3–4
Tech	4.00	4	3–5	4.00	4	3–5	3.00	3	2–4
Dining	3.00	4	3–4	4.00	4	3–4	3.00	3	3–4
Wellness	4.00	4	3–4	3.00	3	3–4	4.00	4	3–4
Travel	4.00	5	3–5	4.00	4	3–5	4.00	5	3–5
Education	4.00	4	3–5	4.00	4	3–5	3.00	4	2–4
Finance	4.00	4	3–5	4.00	4	3–4	3.00	4	2–4
Housing	4.00	4	3–4	4.00	4	3–4	3.00	3	2–4

Values represent the Median (Med.), Mode, and Interquartile Range (IQR). Notably, high-stakes domains such as Travel, Finance, and Housing consistently elicit higher central tendency scores for Educative support, while hedonic domains like Apparel and Entertainment show a stronger median preference for Explorative interaction.
Note: Ratings measured on a 5-point Likert scale (1 = Low priority, 5 = High priority). IQR reported as 25th–75th percentiles.

educative features are more valued in complex, high-stakes domains. For the Explorative aim, Travel was rated significantly higher than Wellness ($r = .25$), consistent with novelty-seeking preferences in exploratory domains. For the Affective aim, Wellness was rated higher than Tech ($r = .21$), reflecting heightened affective expectations in emotionally salient domains.

We further validated these findings using Bayesian hierarchical ordinal regression models, which incorporated domain-level random intercepts and individual-level fixed effects. Posterior estimates confirmed the frequentist results. For example, Education domain had the highest estimated intercept for educative interaction ratings ($\beta_{\text{Education}} = 1.25$, 94% HDI [0.47, 1.96]), while Travel had the highest for explorative interaction ratings ($\beta_{\text{Travel}} = 1.30$, 94% HDI [0.52, 1.99]), and Wellness for

Table 4. Kruskal–Wallis H-Test Results and Domain-Level Mean Ranks for Interactional Aims ($N = 168$ Users Across 10 Domains)

Interaction Aim	H	df	p	Top 3 Domains (Mean Rank)	Bottom 3 Domains (Mean Rank)
Educative	93.15	9	$< .001$	Education (1014.9) Tech (981.8) Travel (960.9)	Dining (681.0) Apparel (730.4) Entert. (732.5)
Explorative	33.34	9	$< .001$	Travel (948.9) Education (916.4) Tech (898.3)	Wellness (730.4) Beauty (770.5) Dining (792.9)
Affective	25.10	9	.003	Wellness (922.4) Travel (918.0) Entert. (885.2)	Tech (720.6) Housing (779.5) Finance (822.4)

Significant H-statistics across all three aims ($p \leq .003$) confirm that user prioritisation of Educative, Explorative, and Affective support is context-dependent. Mean ranks reveal the divergence between high-stakes/functional domains (e.g., Education, Tech) and hedonic/wellness domains.

Note: H follows a χ^2 distribution with 9 degrees of freedom (df). Mean ranks represent the relative prioritisation after pooling all user responses ($n = 1680$ observations). Higher ranks indicate a stronger demand for that specific interactional quality.

Table 5. Summary of Significant Pairwise Differences from Bonferroni-Adjusted Mann–Whitney U Tests

Interaction Aim	Significant Domain Pairs (Directional)	Effect Size (r)
Educative	Education > Dining; Education > Apparel; Tech > Dining; Tech > Apparel; Travel > Dining	0.25–0.36
Explorative	Travel > Wellness; Education > Wellness	0.23–0.25
Affective	Tech < Wellness; Tech < Travel	0.21–0.22

Only pairs with a notable effect size ($r \geq 0.20$) are reported. The results indicate that Educative support requirements exhibit the strongest differentiation between domains (e.g., Education vs. Dining), while Affective requirements are uniquely prioritised in Wellness and Travel contexts.

Note: p-values were Bonferroni-corrected. Effect sizes r are interpreted via Cohen's guidelines: ≥ 0.10 (small), ≥ 0.30 (medium). A > symbol indicates the first domain in the pair was rated significantly higher for that specific interaction aim.

affective interaction ratings ($\beta_{\text{Wellness}} = 1.12$, 94% HDI [0.38, 1.83]). These results provide robust support for H1, H2, and H3. User preferences for educative, explorative, and affective interaction aims are not uniformly distributed but vary systematically across application domains. While our analysis does not directly measure latent factors such as domain complexity, novelty orientation, or emotional salience, the observed patterns are consistent with these theoretical characteristics. Although interaction demands may differ across specific items within a domain, the present results indicate domain-level user expectations that are robust enough to inform adaptive CRS design. These domain effects can serve as priors for domain-conditioned policies, enabling systems to initialise aim weightings and explanation styles according to task context.

4.3 Impact of Item Value on Interaction Aims (H4)

To evaluate H4, participants rated the perceived importance of educative, explorative, and affective aims across two distinct situational contexts: a low-value and a high-value item scenario. This yielded $n = 168$ paired observations per interactional aim. Figure 2 illustrates the distribution of

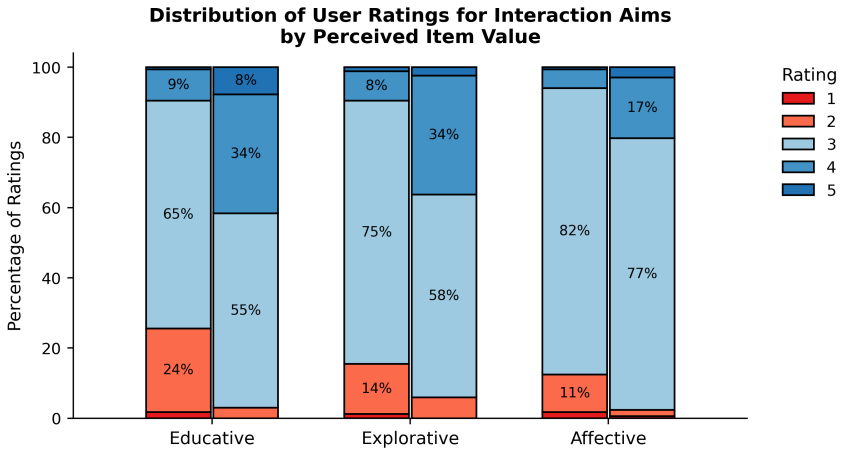


Fig. 2. Stacked-bar distribution (%) of user ratings for educative, explorative, and affective interaction aims, comparing low-value and high-value item scenarios. Bars on the *left* represent low-value items; bars on the *right* represent high-value items. Colour gradients from red (rating = 1) to dark blue (rating = 5) indicate increasing perceived importance.

these ratings. Across all three aims, the transition to a high-value scenario produced a pronounced upward shift in user expectations. The proportion of low-importance ratings (1–2) decreased substantially, while high-importance ratings (4–5) saw a corresponding surge. For the educative aim, high-importance ratings rose from 9 % to 42 % (a 33 percentage point increase), while the affective aim saw a shift from 1 % to 18 %. Two-sided Wilcoxon signed-rank tests (Table 6) confirmed that these increases were statistically significant across all dimensions: educative ($W = 368.5$, $z = -10.66$, $p < .001$, $r = .82$), explorative ($W = 473.5$, $z = -10.49$, $p < .001$, $r = .81$), and affective ($W = 68.5$, $z = -11.13$, $p < .001$, $r = .86$). All observed effects were categorised as large according to conventional benchmarks. While median scores remained at 3.0, this measure of central tendency masked the substantial distributional shifts evident in both the stacked-bar visualisation and the large effect sizes. These findings provide robust support for H4, suggesting that perceived decision stakes significantly elevate the importance users place on all three interactional aims. Within the RAE framework, item value functions as a high-priority *state variable* that should trigger real-time policy adjustments. Specifically, under high-value conditions, the dialogue policy should scale behavioural intensity across three fronts: *educative behaviours* should transition toward deeper, more contextualised explanations and comparative justifications; *explorative behaviours* should expand to include broader, more diverse retrieval sets; and *affective behaviours* should adopt a more empathetic, alignment-oriented tone. This mapping ensures that system responses, whether implemented via prompt conditioning or controllable NLG, are scaled proportionally to the situational stakes inferred from task metadata.

4.4 Moderation by User Characteristics (H5)

Hypothesis 5 posited that individual differences, specifically *age*, *gender*, and prior *CRS experience*, would predict user prioritisation of interactional aims. To test this, we first conducted an exploratory domain-level correlation screen with Benjamini–Hochberg (FDR) correction, followed by confirmatory Bayesian hierarchical models to estimate global effects across domains.

Table 6. Wilcoxon Signed-Rank Test Results Comparing User Prioritisation of Interactional Aims between Low-Value and High-Value Item Conditions ($n = 168$ Pairs)

Interaction Aim	W	z	p	r
Educative	368.5	−10.66	$< .001$	.822
Explorative	473.5	−10.49	$< .001$	.809
Affective	68.5	−11.13	$< .001$	.859

The extremely large effect sizes across all dimensions ($r > .80$) indicate that the perceived value of the recommendation item is a primary driver of user demand for Educative, Explorative, and Affective support, justifying the adaptive state-logic of the RAE framework.
Note: All tests are paired and two-tailed. Effect size r is computed as $|z|/\sqrt{n}$. According to Cohen's guidelines, $r > .5$ represents a large effect.

Table 7. Correlation Analysis of Demographic Covariates and Interaction Aims

Domain	**Interaction Aim**	**Age** $\rho\,(p)$	**Gender** $\rho\,(p)$	**CRS Exp.** $\rho\,(p)$
Apparel	Explorative	−0.04 (.99)	−0.01 (.93)	0.27** (.00)
Education	Educative	−0.15 (.78)	−0.00 (.99)	0.18* (.05)
Education	Explorative	−0.07 (.99)	0.14 (.26)	0.19* (.04)
Entertain.	Affective	−0.01 (.99)	−0.02 (.92)	0.24* (.01)
Finance	Affective	0.02 (.99)	0.14 (.26)	0.21* (.03)
Finance	Educative	0.00 (.99)	0.04 (.77)	0.19* (.04)
Finance	Explorative	0.10 (.99)	0.24* (.04)	0.20* (.04)
Housing	Affective	−0.00 (.99)	0.24* (.04)	0.16 (.07)
Housing	Explorative	−0.06 (.99)	0.19 (.15)	0.18* (.05)
Tech	Affective	0.04 (.99)	0.07 (.70)	0.32** (.00)
Tech	Explorative	−0.02 (.99)	0.04 (.77)	0.20* (.04)
Wellness	Affective	−0.16 (.78)	0.14 (.26)	0.23* (.01)
Wellness	Educative	−0.00 (.99)	0.04 (.77)	0.24* (.01)
Wellness	Explorative	−0.02 (.99)	0.12 (.33)	0.33** (.00)

The table presents Spearman's rank correlation coefficients (ρ) and Benjamini–Hochberg (BH) adjusted p-values. Notably, CRS Experience shows a consistent positive association, whereas Age and Gender exhibit minimal significant effects.
Note: Gender effect-coded (+1 Female, −1 Male). CRS Exp. measured 1–5. Significance: $^{*}p < .05$,$^{**}p < .01$ (BH-adjusted). Only rows with at least one significant result shown.

Domain-level associations (Spearman). For this analysis, gender was effect-coded (female = +1, male = −1) and CRS experience was treated as an ordinal 5-point scale. Table 7 reports the resulting Spearman's ρ coefficients and FDR-corrected p-values. *CRS experience* emerged as the most robust predictor, with 13 of 30 domain-interaction combinations reaching significance after correction. This included educative importance in *Dining* and *Wellness*; explorative importance across almost all domains except *Tech* and *Entertainment*; and affective importance in *Apparel*, *Beauty*, *Finance*, and *Tech*. In contrast, *gender* effects were highly localized, appearing only for explorative aims in *Finance* and affective aims in *Housing*. Notably, *age* yielded no significant associations across any domain or aim after multiple-comparison correction.

Bayesian hierarchical ordinal models. To confirm these trends, we fitted cumulative-logit mixed models using `Stan` (4 chains, 2000 post-warmup draws). These models incorporated domain-level random intercepts ($\alpha_j \sim \mathcal{N}(0, \sigma_\alpha)$) and weakly informative $\mathcal{N}(0, 1)$ priors. Convergence was excellent, with $\hat{R} < 1.01$ and ESS exceeding 1500 for all parameters. Posterior predictive checks (see

Table 8. Descriptive Statistics and One-Sample Wilcoxon Signed-Rank Tests for Dialogue Control Preferences ($N = 168$)

Interaction Aim	Mean (SD)	Median	W	z	r
Educative	3.47 (1.02)	4.00	1345.0	−5.12	.39
Explorative	3.45 (0.98)	3.50	1566.0	−4.88	.38

Participants exhibited a significant preference for user-initiated control across both aims ($p < .001$), supporting H6a. Effect sizes (r) suggest users value autonomy in directing the conversation during both educative and explorative phases.
Note: Scale: 1 = System-initiated, 3 = Neutral, 5 = User-initiated. All tests are compared against the neutral midpoint (3). Spearman's ρ(Educative, Explorative) = 0.47, $p < .001$.

Appendix Figures B.1–B.3) show strong alignment between predicted and observed frequencies, indicating good model calibration and fit. The estimated global effects (94% HDI) across all domains were:
Educative: $\beta_{exp} = 0.21\ [0.12, 0.29]$; $\beta_{gender} = -0.25\ [-0.43, -0.06]$; $\beta_{age} = -0.08\ [-0.17, 0.00]$.
Explorative: $\beta_{exp} = 0.36\ [0.27, 0.44]$; $\beta_{gender} = -0.42\ [-0.60, -0.25]$.
Affective: $\beta_{exp} = 0.40\ [0.32, 0.49]$; $\beta_{gender} = -0.51\ [-0.69, -0.34]$.

Summary and implications. Across both analytical lenses, CRS experience emerged as the most consistent and positive predictor of interaction-aim importance, supporting its role as a stable input for adaptive user modelling. In contrast, gender effects were notably sparse and domain-specific (e.g., Explorative in *Finance* and Affective in *Housing*), while age showed no reliable influence after multiple-comparison correction. These results imply that only CRS experience warrants global policy adaptation, while gender-based adjustments should remain localised and cautious, and age can be excluded as a policy variable. This finding is consistent with the RAE framework's implementation-agnostic guidance for lean, ethically grounded adaptation in CRS. By deprioritising blunt demographic categories in favour of behavioural signals like system experience, the framework achieves a more robust and personalized user-state representation.

4.5 User Autonomy and Dialogue Control (H6)

While the preceding analyses established the situational and individual drivers of interactional aims, the practical implementation of the RAE framework requires an understanding of how dialogue initiative should be distributed between the agent and the user. Consequently, we analysed user preferences for dialogue control within the *educative* and *explorative* CRS interaction phases.[1]

Descriptive statistics and preference tests. Table 8 reports the descriptive statistics and Wilcoxon signed-rank tests for the two 5-point control items (where 1 = *fully system-initiated* and 5 = *fully user-initiated*). User ratings for both educative and explorative control were significantly above the neutral midpoint (3.0), indicating a clear general preference for user-initiated interaction and providing strong support for H6a. The observed effect sizes were medium–large for both educative ($W = 1345, z = -5.12, p < .001, r = 0.39$) and explorative ($W = 1566, z = -4.88, p < .001, r = 0.38$) interaction types. Furthermore, a moderate positive correlation between the two items (Spearman's $\rho = 0.47, p < .001$) indicated high internal consistency in autonomy preferences across distinct interactional aims.

[1]We did not include a control item for *affective* interactions as affective support is inherently system-led, arising from implicit user cues and contextual inference rather than explicit user direction. Including control ratings for affective behaviour would risk conflating preference with feasibility.

Table 9. Cumulative Logit (Ordinal) Models Predicting User-Initiated Dialogue-Control Preferences ($N = 164$)

Predictor	Educative		Explorative	
	β	p	β	p
Gender (Male)	−0.79	**.009**	−0.27	.349
CRS Experience	−0.03	.815	0.14	.264
Age Group[a]				
25–34 years	0.15	.650	0.27	.420
35–44 years	−0.29	.489	0.09	.819
45–54 years	−0.16	.849	−0.03	.966
55–64 years	−0.13	.852	0.05	.947

Notably, Gender emerges as a significant predictor for the *Educative* aim ($p = .009$), where females indicate a significantly higher preference for user-led control compared to males.
Note: Gender coded as Male=1, Female=0. Negative β for Gender indicates higher odds for females to prefer user-initiated control. Reference group for Age is 18–24 years. Bold indicates statistical significance ($p < .05$).

Ordinal logistic regression models. To evaluate H6b, we fitted separate ordinal logistic regression models predicting user preferences for control. Predictors included gender (Male=1, Female=0), CRS experience (ordinal), and age (categorical, with 18–24 as the reference group). As shown in Table 9, gender significantly predicted autonomy preferences within the *educative* model. Specifically, *female* participants exhibited significantly higher odds of preferring user-initiated control (Male coefficient $\beta = -0.79$, $p = .009$; Female vs. Male OR $= 2.19$, 95% $CI\,[1.22, 3.95]$). In contrast, no demographic predictors reached significance in the *explorative* control model. These findings suggest that while overall user preferences favour high autonomy (H6a), demographic modulation is limited to a modest gender-specific effect in educative dialogue (H6b).

Summary and implications for adaptive CRS design. Our findings demonstrate a statistically significant preference for user-initiated dialogue control during both educative and explorative exchanges. This autonomy orientation was consistent across interactional aims, supporting H6a. Regarding H6b, gender emerged as the sole demographic predictor, and specifically only within the *educative* context, where females were more likely than males to prefer user-led control. Within the RAE framework, these autonomy preferences act as a vital user-state signal for dialogue management. While the results do not dictate a specific technical stack, they suggest that adaptive systems can improve user alignment by dynamically adjusting initiative-taking strategies. Given the observed stability of these preferences, implementing a lightweight personalisation rule or model feature can sufficiently support experiential consistency without introducing unnecessary architectural complexity.

4.6 Summary of Hypothesis Tests

Table 10 consolidates all six hypotheses, summarising methods, key findings, and outcomes. Patterns across results indicate that *domain context*, *perceived item value*, *prior CRS experience*, and *autonomy preferences* are the most consistent and discriminative user-state signals for tailoring educative, explorative, and affective interaction aims. H1–H3 confirmed systematic variation across domains, with high-stakes and cognitively complex contexts elevating educative aims, and affect-rich contexts elevating affective aims. H4 showed that high-value items consistently increased ratings across all aims, underscoring sensitivity to perceived decision stakes. H5 revealed that CRS experience

Table 10. Summary of Hypothesis Testing Outcomes (H1–H6)

Hypothesis	Method(s)	Key Findings	Status
H1: Educative / complexity	KW, MWU, Bayes	Omnibus $H = 93.15$, $p < .001$; Education > Dining ($r = .36$), Tech > Apparel ($r = .32$). Bayesian $\beta_{\text{Educ.}} = 1.25$, 94% HDI [0.47, 1.96].	Supported
H2: Explorative / novelty	KW, MWU, Bayes	Omnibus $H = 33.34$, $p < .001$; Travel > Wellness ($r = .25$). Bayesian $\beta_{\text{Trav.}} = 1.30$, 94% HDI [0.52, 1.99].	Supported
H3: Affective / emotion	KW, MWU, Bayes	Omnibus $H = 25.10$, $p = .003$; Wellness > Tech ($r = .21$). Bayesian $\beta_{\text{Well.}} = 1.12$, 94% HDI [0.38, 1.83].	Supported
H4: Item value moderation	Wilcoxon (paired)	High-value items rated significantly higher for all aims ($p < .001$): Educative ($r = .82$); Explorative ($r = .81$); Affective ($r = .86$).	Supported
H5: User traits	Spearman, Bayes	CRS experience positive in 13/30 pairs ($\beta = .21$–$.40$, HDIs $\not\ni 0$). Gender/Age effects largely negligible.	Partially Supported
H6: Autonomy / Mode	Wilcoxon, Regr.	Preference for user-initiated control ($r \approx 0.50$, $p < .001$). Females prefer user-led educative control (OR = 2.19, $p = .009$).	Partially Supported

Findings synthesise frequentist tests (Kruskal–Wallis, Wilcoxon) and Bayesian hierarchical modelling. The results confirm that interactional aims are significantly modulated by decision domain (H1–H3) and item value (H4), while user traits (H5) and autonomy preferences (H6) exhibit more nuanced, partial support.
Note: KW: Kruskal–Wallis; MWU: Mann–Whitney *U*; Bayes: Bayesian hierarchical ordinal regression; HDI: Highest Density Interval. $n = 168$.
Supported: $p < .05$ and Bayesian HDI excludes zero. Partially Supported: Significant effects found in specific subsets only.

reliably predicted higher ratings, whereas gender effects were rare and domain-specific, and age effects negligible. H6 indicated a general preference for user-initiated dialogue, with only minor demographic modulation. From an RAE framework perspective, these variables can be modelled as features in an *adaptive CRS state representation*, regardless of underlying implementation. This enables dialogue policies, whether learned or heuristic, to adjust aim weighting, explanation depth, exploratory breadth, and conversational initiative in ways that are context- and user-aligned. The findings thus provide direct, evidence-based parameters for adaptive policy design, extending beyond static interface guidelines toward experience-sensitive conversational strategies.

5 Discussion

Prior work on CRS has typically examined educative, explorative, or affective aims in isolation, or treated user experience as a static evaluation outcome rather than an adaptive driver of dialogue policy [7, 70, 80, 81]. Our findings extend this literature by showing how these aims can be modelled jointly and how their relative importance shifts with domain, item value, and user characteristics. This reframing positions experiential aims not solely as post-hoc evaluation constructs but as actionable state features, providing a bridge between user-centred evaluation frameworks (e.g., ResQue [57], CRS-Que [29]) and adaptive dialogue management approaches [4, 8]. By demonstrating that educative, explorative, and affective aims operate as dynamic state variables shaped by context and user signals, our study establishes an empirical basis for context-sensitive CRS dialogue management.

5.1 From Empirical Evidence to Adaptive Logic

The empirical findings presented in Section 4 provide more than a statistical validation of our hypotheses; they offer the necessary grounding for a multi-layered approach to conversational

adaptation. By quantifying the systematic shifts in user expectations across 10 domains, varying stakes, and diverse individual backgrounds, we can move beyond "one-size-fits-all" dialogue policies. The following sections synthesise these insights into the RAE framework, establishing how contextual and user-state signals can be mapped to experience-aligned system behaviours.

Domain-specific prioritisation of interaction aims. Results for RQ1 and H1–H3 confirm that the importance users assign to educative, explorative, and affective aims varies systematically across domains. From an adaptive CRS perspective, these patterns can be encoded as *domain-level priors* in the system's state representation, providing baseline weightings that guide the dialogue policy before finer-grained situational or user-specific signals are incorporated [17, 27]. Such priors allow a CRS to initialise explanation depth, novelty breadth, and affective tone in ways that match the cognitive and emotional demands characteristic of each application area, thereby supporting context-sensitive interaction from the outset.

Educative aims: cognitive complexity and contextual nuance. As hypothesised in H1, educative aims were highest in complex domains (*Education*, *Technology*), confirmed by a robust Kruskal–Wallis test ($H(9) = 93.15$, $p < .001$). Pairwise effects (e.g., Education vs. Dining, $r = 0.36$) show strong differentiation from less cognitively demanding domains. In an adaptive policy model, such results justify higher initial weights for educative behaviours when `domain_complexity` is high, triggering richer rationales and scaffolded guidance [84]. Interestingly, *Travel* also ranked highly for educative aims alongside explorative and affective priorities, suggesting a blended aim configuration. Conversely, *Finance* showed heterogeneous responses, implying sub-domain profiling or user-specific signals are necessary.

Explorative aims: hedonic vs. stability-driven contexts. Consistent with H2, explorative aims peaked in novelty-oriented domains (*Travel*, *Entertainment*), $H(9) = 33.34$, $p < .001$. These domains can be tagged with a high `novelty_orientation` feature, prompting broader candidate diversity and serendipitous recommendations. Lower explorative prioritisation in *Wellness* illustrates the need for policies that suppress novelty when affective stability is valued, aligning exploration parameters with domain-specific risk profiles [26].

Affective aims: emotional and social salience. Aligned with H3, affective aims dominated in *Wellness* and *Beauty* ($H(9) = 25.10$, $p = .003$), supporting high default weights for empathetic tone and rapport-building behaviours when `emotional_salience` is high [59]. These features can serve as triggers in an adaptive orchestration layer, activating affective templates or tone modulation modules.

These domain-level patterns map directly to policy initialisation parameters (*domain priors*) that can be overridden or reweighted in real time as situational (Section 5.1) and user-specific (Section 5.1) signals become available.

Item value sensitivity as a dynamic modulator. Addressing H4, perceived item value emerged as a strong within-subject modulator of aim prioritisation. Wilcoxon signed-rank tests confirmed significant increases under high-value scenarios for all three aims: Educative ($r = .82$), Explorative ($r = .81$), and Affective ($r = .86$), with all $p < .001$. These large effects demonstrate that as decision stakes rise, users recalibrate their expectations toward deeper explanation, broader exploration, and richer emotional engagement. This finding resonates with prior research suggesting that decision value amplifies user demands for justification, transparency, and support in recommender interactions [10, 29]. From an adaptive dialogue management perspective, `item_value` serves as a high-sensitivity variable within the state vector, capable of triggering substantial reweighting of aims within a single session. This aligns with the RAE proposition that situational variables can

override or amplify baseline policy parameters in real time. *Operational requirements:* Incorporating these insights into functional CRS deployments necessitates that the system treats `item_value` as a tracked, dynamic state feature. The architecture must support the mapping of high-value conditions to specific aim-weight adjustments, whether through heuristic thresholds or learned gradients, while ensuring a graceful return to baseline aim levels as situational cues diminish. This mechanism of *layered adaptivity* prevents over-optimisation and user fatigue, ensuring the system remains efficient in low-stakes cases while providing high-fidelity support when stakes are elevated.

Moderation by user characteristics. Addressing H5, *prior CRS experience* emerged as a consistent predictor of higher ratings across multiple domains. This supports the notion that familiarity with conversational systems fosters higher expectations for depth, novelty, and socially expressive behaviours [5, 27]. In contrast, *gender* effects were narrower and context-contingent, appearing primarily in the *Finance* domain for explorative aims and *Housing* for affective aims. For instance, females assigned significantly higher affective ratings in the *Housing* context (Bayesian $\beta = -0.51$, 94% HDI excluding zero), suggesting a context-linked sensitivity to socially supportive behaviours [58]. Conversely, *age* effects were negligible after correction for multiple comparisons, aligning with contemporary findings that user expectations for conversational assistants are more uniformly distributed across age groups than previously theorised [1].

User-centric policy implications. Translating these insights into the RAE framework necessitates a tiered approach to demographic features. Given its predictive robustness, `crs_experience` should be treated as a high-confidence behavioural feature for real-time policy adjustment. For expert users, the system should reduce conversational redundancy and increase the injection of novelty; for novices, the policy should shift to prioritise clarity, scaffolding, and interactional reassurance. Furthermore, the results suggest that demographic features like `gender` must be treated as conditional modulators, applied only in specific domains where reproducible, context-specific effects have been validated. Importantly, while our analysis found age to have a negligible predictive influence on interactional aims, this finding must be interpreted within the constraints of our sample demographic. With 73.2 % of participants aged 18–34 and zero representation of adults aged 65 and older (Table 2), the current dataset lacks the variance necessary to definitively conclude that age does not influence CRS expectations. Consequently, while age may be deprioritised as a primary policy variable for younger cohorts, future research specifically targeting older demographics is required before drawing conclusions regarding age-based adaptation. For a detailed discussion on the ethical considerations of this approach, including safeguards against stereotyping, see Section 5.2.2.

User autonomy and dialogue control preferences. Addressing H6a and H6b, participants showed a clear preference for *user-initiated* dialogue for both the *educative* ($M = 3.47$; median = 4) and *explorative* (M = 3.45; median = 3.5) aims. One-sample Wilcoxon signed-rank tests against the neutral midpoint (3 = equal user/system initiative) indicated ratings above neutral for both aims (Educative: $W = 1345$, $n' = 110$, $p < .001$, $r = 0.51$, rank-biserial = 0.56, CLES = 0.78; Explorative: $W = 1566$, $n' = 116$, $p < .001$, $r = 0.49$, rank-biserial = 0.54, CLES = 0.77; Table 8). A moderate positive association between the two aims (Spearman's $\rho = 0.471$, $p < .001$) suggests that autonomy preferences are broadly consistent across educative and explorative contexts.[2] These findings echo broader work in conversational systems, where users frequently prefer to retain initiative in information-rich interactions [32, 41, 67].

Demographic moderation. Ordinal logistic regression identified *gender* as the only reliable demographic predictor, and only for the *educative* aim: *female* participants had higher odds of preferring

[2] n' denotes the number of non-tied observations (ratings $\neq 3$) used by the Wilcoxon test.

user-initiated control (Male coefficient $\beta = -0.785$, $p = .009$; Female vs. Male OR = 2.19, 95% CI [1.22, 3.95]; see Table 9). Age and CRS experience were non-significant. Notably, domain-specific gender effects observed for aim *importance* (e.g., explorative–Finance; affective–Housing) did not map onto corresponding differences in *control* preference; the detected control effect appears specific to educative dialogue.

Contextual scope of autonomy preferences. While the observed preference for user-initiated dialogue provides a strong baseline for the RAE framework's `autonomy_pref` parameter, it is crucial to acknowledge the contextual bounding of this finding. The dialogue-control items in this study were framed explicitly within an apparel e-commerce context—a domain characterised by high hedonic value but relatively low cognitive risk. User autonomy preferences when browsing for clothing may differ vastly from the initiative structures desired when applying for a mortgage (finance) or seeking medical advice (wellness). Furthermore, our framework intentionally excluded a control metric for the affective aim, operating on the theoretical assumption that genuine affective support is inherently system-led, arising from implicit user cues rather than explicit user commands. While this assumption aligns with current affective computing paradigms, it remains an empirical question whether users desire explicit control over emotional interaction styles. Future multi-domain autonomy studies are necessary to validate whether these initiative preferences hold across higher-stakes contexts and across all experiential dimensions.

Policy implications. Within the RAE framework,
`autonomy_pref` can be modelled as a state feature that allocates conversational initiative:
High `autonomy_pref` → system defers to user-initiated turns; unsolicited prompts are minimised except under high-stakes conditions or when clarification is urgent.
Low `autonomy_pref` → system adopts more initiative, proactively scaffolding the dialogue while maintaining easy opt-out.

For affective behaviours, often more effective when timely and unsolicited, system-initiated turns can remain the default, triggered by sentiment cues and always overrideable by the user [43, 78].

Interaction with domain and value. Autonomy preferences interact with domain profile and perceived value: in educative-dominant, higher-value contexts (e.g., *Education*, *Technology*), user-initiated turns can enhance trust and comprehension; in affect-rich contexts (e.g., *Wellness*, *Beauty*), gentle system-initiated affective support may better sustain rapport. Thus, `autonomy_pref` complements domain profile, item value, and user traits to yield an adaptive initiative strategy within the RAE state–policy mapping.

5.2 RAE Perspectives and Framework

The empirical validation of interactional aims as dynamic variables–shaped by domain, item value, and user experience–enables a direct transition to architectural design. Consequently, the RAE framework translates these layered findings into a formal control model for CRS.

The RAE adaptation framework. The dialogue manager maintains a *state vector* with four feature groups: (1) `domain_profile` (task complexity, novelty orientation, emotional salience), (2) `item_value` (stakes), (3) `user_traits` (behavioural history, conditional demographics), and (4) `autonomy_pref` (initiative pattern). A policy function maps this state to an *aim-weight vector* over educative, explorative, and affective aims. Formally, at any dialogue turn t, the conversational state is represented as the vector:

$$\mathbf{s}_t = \langle \mathcal{D}, \mathcal{V}_t, \mathcal{U}_{exp}, \mathcal{U}_{aut}, \mathbf{H}_t \rangle$$

where $\mathcal{D}$ represents the domain priors, $\mathcal{V}_t$ the dynamic item value, $\mathcal{U}_{exp}$ and $\mathcal{U}_{aut}$ the persistent user traits and autonomy preferences, and $\mathbf{H}_t$ the latent dialogue history. The adaptation policy π

Table 11. Context-Sensitive Modulation of Interactional Aims within the Recommendation-as-Experience (RAE) Framework

Domain cluster	Edu.	Exp.	Aff.	Preferred mode	Indicative triggers
High-stakes / Complex	✓✓	–	–	System → User	High item value; extreme decision complexity
Cross-cutting (e.g., Travel)	✓	✓	✓	Mixed-initiative	Multi-criteria trade-offs; novelty seeking
Hedonic / Leisure	–	✓✓	✓	User-led or Mixed	Visual novelty cues; trend-based browsing
Affect-rich / Identity	✓	✓	✓✓	Gentle System Init.	Emotional salience; expressed sentiment
Social / Contextual	✓	✓✓	✓	User-led / Nudges	Occasion framing; group decision context
Functional / Pragmatic	✓✓	–	–	System → User	Specification comparison; expertise signals

The table illustrates how the system dynamically shifts its emphasis across Educative (Edu.), Explorative (Exp.), and Affective (Aff.) dimensions based on domain-specific triggers and situational complexity. By mapping these high-level clusters to preferred dialogue modes (e.g., system-led vs. user-led), the framework facilitates an adaptive initiative-taking policy that balances epistemic decision support with affective alignment.
Note: ✓✓: Primary/Strong emphasis; ✓: Secondary/Moderate emphasis; –: Low/De-emphasized.
Preferred dialogue mode specifies the allocation of conversational initiative based on the predicted user information need.

then maps this state to the weight vector $\mathbf{a}_t$:

$$\pi(\mathbf{s}_t) \rightarrow \mathbf{a}_t, \quad \text{where } \mathbf{a}_t = \{w_{edu}, w_{exp}, w_{aff}\}$$

The transition from empirical user expectations to computational system parameters is illustrated in Figure 4, which maps the significant effect sizes derived from our multi-domain study to the primary weights of the RAE state vector. This mapping is *implementation-agnostic* and can be realised as rules, learned policies, or LLM-based orchestration layers [27, 39, 77, 77, 80]. Table 11 encodes rules consistent with our post-hoc analyses; these serve as defaults that a learned or hybrid policy can refine. Figure 3 presents the RAE adaptation framework in a triangular view. The three interaction aims define a ternary space, with contextual signals feeding into a mixer that produces a blended behavioural orientation.

5.2.1 Illustrative Design Vignettes and Rule-set. Building directly on the synthesis presented in Section 5.2, we illustrate how the layered predictors of domain, item value, user traits, and autonomy preferences are operationalised through the RAE framework. To demonstrate the framework's flexibility, we present two contrasting interaction scenarios.

Scenario A: High-Stakes Travel Planning. Consider a user planning a once-in-a-lifetime honeymoon, representing a *high-value* situational context within the *Travel* domain. In this scenario, the *domain layer* is characterised as novelty-heavy and moderately complex; consequently, the RAE framework assigns balanced emphasis to explorative and educative aims (Section 5.1). Within the *user layer*, we observe a highly experienced CRS user, triggering a shift in system tone toward concise and socially engaging responses (Section 5.1). Finally, at the *situational layer*, the high `item_value` signals the need for deeper explanations, wider option diversity, and the activation of affect-sensitive phrasing (Section 5.1). The resulting interaction surfaces serendipitous travel

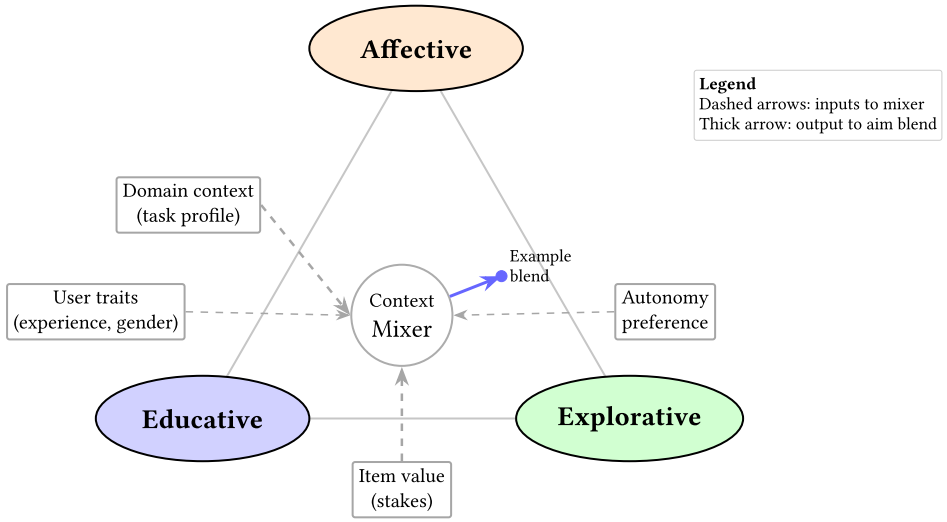


Fig. 3. RAE adaptation framework (triangle view). Educative, explorative, and affective aims form the vertices of a ternary space. Contextual inputs (domain, user traits, item value, autonomy preference) feed into a mixer, yielding an adaptive blend of aims that guides CRS behaviour.

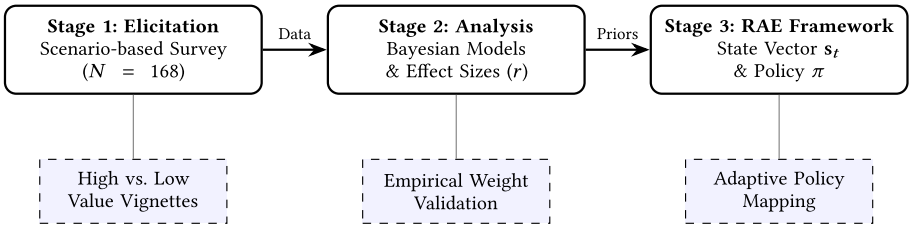


Fig. 4. The three-stage methodological workflow for operationalising the Recommendation-as-Experience (RAE) framework. Stage 1 elicits user preferences through high-stakes vignettes; Stage 2 utilises Bayesian modeling to derive robust empirical weights and effect sizes (r); and Stage 3 translates these weights into a formal computational policy π and state vector $\mathbf{s}_t$ for adaptive system behaviour. Dashed boxes represent the specific sub-components validated at each phase of the pipeline.

options paired with transparent tradeoffs, such as detailed cost breakdowns and contextualised justifications.

Scenario B: Postgraduate Education Selection. In contrast, consider a user evaluating postgraduate study options. Here, the *Education* domain profile is primarily cognitively demanding and high-stakes. At the *domain layer*, educative aims dominate to provide the structured comparisons and rationale necessary for such a significant decision. Within the *user layer*, the system identifies a novice user with limited CRS experience; the dialogue policy thus prioritises scaffolding, clarity, and step-by-step guidance. Given the high `item_value` at the *situational layer*, the system proactively surfaces additional resources such as career outcomes and alumni reviews to bolster user confidence. The system ultimately articulates degree options with transparent tradeoffs (e.g., tuition fees vs. global ranking), carefully balancing the depth of functional explanation with active reassurance.

Operationalising the RAE Framework. These narratives illustrate how domain profiles, user traits, and situational value cues jointly determine adaptive aim blends. This mapping logic can be realised

through various computational architectures, ranging from heuristic rule-sets to learned policies or LLM-based orchestration layers [80, 87]. As demonstrated in the pseudocode below, the RAE state vector serves as a principled input for these policies, ensuring that real-time system behaviour remains grounded in validated predictors.

```
# State vector derived from multi-source signals
state = {
    "domain_profile": "Travel",
    "item_value": "High",
    "user_traits": {"crs_experience": "High"},
    "autonomy_pref": "User-led"
}

# (a) Rule-based Mapping (Heuristic Policy)
if state["item_value"] == "High":
    set_aim_weights(educative=0.8, explorative=0.6, affective=0.7)

# (b) Learned Policy (Supervised/RL Mapping)
# Vectorises state and predicts weight distribution
weights = policy_model.predict(encode(state))

# (c) Orchestration (LLM adaptation)
# Adapts prompt instructions based on aim weights
prompt = adapt_prompt(base_template, weights)
response = generator(prompt)
```

This state-driven approach ensures that the RAE framework remains independent of the specific implementation stack while providing a conceptual vocabulary for designing context-sensitive recommendations.

5.2.2 Design Guidance. The RAE framework facilitates adaptation across three primary dimensions:

System initialisation. Levers include domain-aware defaults, where weights are derived from a `domain_profile`, and aim-aligned evaluation, which complements standard ranking metrics with specific experiential measures [29].

User-centric personalisation. Personalisation is achieved through behavioural calibration based on CRS experience and configurable autonomy, allowing users to manage initiative through interface elements like detail–discovery sliders.

Contextual dynamics. The system maintains relevance via real-time value triggers that scale aims during high-stakes episodes and uses transparent adaptivity to communicate these transitions to the user.

These findings establish a principled foundation for encoding experiential aims as adaptive state variables in CRS. Building on established foundations of explainability [3, 84], exploratory search [9, 49], and affective support [14, 59, 64], our study addresses a substantial gap in the literature. While prior research has significantly advanced ranking accuracy and intent modelling [68, 81], as well as conversational fluency and empathy [59, 71], these experiential strands have largely been studied in isolation or treated as static evaluation outcomes. Our study addresses this gap by providing empirical evidence for *when* and *how* educative, explorative, and affective aims should be foregrounded jointly. Building on these insights, the proposed RAE adaptation framework *conceptually outlines* how contextual signals (domain, value, user traits) are mapped to policy mechanisms and to specific behavioural realisations: *educative* (justifications, contrastive comparisons), *explorative* (serendipitous suggestions, novelty framing), and *affective* (empathetic

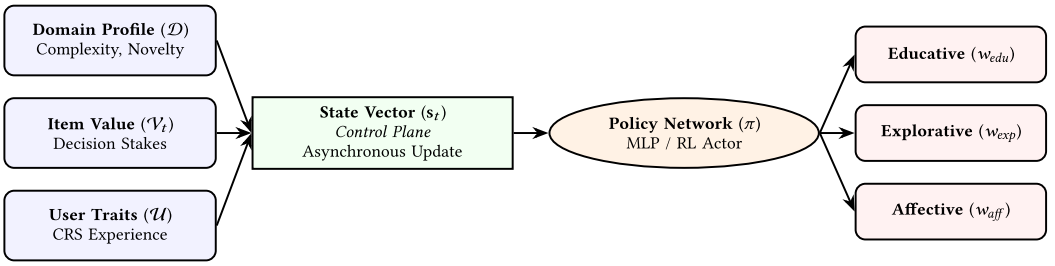


Fig. 5. The Recommendation-as-Experience architectural flow. The Control Plane continuously aggregates contextual inputs into a State Vector ($\mathbf{s}_t$), which a Policy Network maps to an aim-weight vector ($\mathbf{a}_t$). This decoupled approach isolates state calculations from the generation path.

tone, sentiment alignment). By treating these aims as adaptive state variables, the RAE perspective provides both a conceptual vocabulary and practical levers for designing personalised, transparent, and context-sensitive conversational recommendations.

Responsible adaptation and safeguards. In applying the RAE framework, we recommend a conservative stance on personalisation. *Behavioural signals* (e.g., prior CRS experience) should be prioritised as primary drivers of adaptation, while *demographic variables* are treated only as conditional features, used selectively and only where validated, context-specific effects exist. This approach prevents demographic attributes from serving as crude segmentation criteria and mitigates risks of stereotyping or unfair treatment [10, 16, 51]. In addition, policy shifts should be logged for auditability and user oversight, ensuring that adaptive behaviours remain transparent, accountable, and respectful of user autonomy.

5.2.3 Computational Feasibility and Live Inference Architecture. While the heuristic mappings presented in Table 11 establish a conceptual baseline, operating the RAE framework in a live, dynamic environment necessitates a transition from static rule-based logic to continuous real-time adaptation. Deploying such adaptive mechanisms within modern CRS introduces computational constraints, primarily concerning inference latency and memory overhead. To satisfy these requirements without compromising the architecture-agnostic nature of the framework, the RAE system is conceptualised as a decoupled, two-plane architecture (Figure 5): a lightweight state-tracking control plane and a parameter-efficient execution data plane.

The control plane: Dynamic state calculation and neural policy routing. To capture the fluid dynamics of user experience, the conversational state $\mathbf{s}_t$ must be updated asynchronously, avoiding synchronous blocking during live inference. As user utterances are processed via Natural Language Understanding pipelines, a lightweight embedding model (e.g., a quantised bi-encoder [61, 82]) extracts semantic intent to update the temporal item value ($\mathcal{V}_t$) and conversational history ($\mathbf{H}_t$). The mapping of this state $\mathbf{s}_t$ to the continuous aim-weight vector $\mathbf{a}_t = \{w_{edu}, w_{exp}, w_{aff}\}$ is executed via a shallow Neural Policy Network (e.g., a multi-layer perceptron or a reinforcement learning actor network) [18, 42, 75], as illustrated in the *Learned Policy* pathway of our framework. Because this policy network is computationally lightweight, the forward pass to calculate $\mathbf{a}_t$ incurs sub-millisecond overhead. This isolates the state-calculation cost from the critical path, ensuring policy routing completes well within the strict latency budgets required for fluid conversational agents before the computationally expensive generation phase begins [17, 55, 79].

The data plane: Orchestration, latency, and memory footprints. Once the aim-weight vector $\mathbf{a}_t$ is derived, it must be deployed to modulate system behaviour. In LLM-powered CRS environments,

the primary computational bottlenecks are **Time-To-First-Token** (**TTFT**) latency and GPU VRAM consumption associated with the KV-cache. To ensure the deployment of $\mathbf{a}_t$ does not cripple system throughput, the RAE framework relies on parameter-efficient conditioning rather than prompt-bloating heuristics. This can be achieved via dynamic system prompt orchestration—where only highly relevant contextual triggers are prepended based on the dominant weights in $\mathbf{a}_t$—or through advanced serving engines employing multi-LoRA (Low-Rank Adaptation) multiplexing [22]. Under a multi-LoRA architecture, interactional aims are trained as independent, low-rank adapters. During inference, the serving engine dynamically interpolates these adapters into a shared base LLM on a per-request basis. This allows a single deployment endpoint to simultaneously serve numerous concurrent users, each experiencing uniquely weighted RAE policies, with an $\mathcal{O}(1)$ memory footprint overhead per adapter and negligible impact on autoregressive decoding latency.

Hybrid efficiency and graceful degradation. Finally, robust live deployments must account for system resilience and computational quotas. By maintaining the heuristic rule-set as a fallback mechanism, the RAE framework ensures graceful degradation. If dynamic LLM generation exceeds latency thresholds, or if the user is operating in a highly structured, low-stakes domain where generative models offer diminishing returns, the policy smoothly defaults to deterministic, retrieval-augmented templates. This hybrid approach ensures the system maintains high availability and continuous experiential support while optimising overall computational expenditure.

5.3 Limitations and Future Work

While this study establishes an empirical foundation for the RAE adaptation framework, certain methodological constraints present clear avenues for future research. First, to mitigate respondent fatigue across the ten-domain elicitation, we operationalised complex constructs (Educative, Explorative, and Affective aims) using single-item measures. While this enabled broad cross-domain comparison and the extraction of high-level heuristic weights, it inherently limits construct validity compared to multi-item psychometric scales. Future iterations should employ robust multi-item instruments and deeper cognitive profiling to better capture user heterogeneity. Second, our methodology relies on stated preferences elicited through vignettes, which introduces the well-known gap between what users explicitly say they want and their implicit, observed interaction behaviours. While individual self-reports can contain noise, our robust sample size ($N = 168$) helps mitigate the risk of consistent, unidirectional biases skewing the overall aim weightings. Furthermore, extracting these nuanced, multi-dimensional expectations purely from implicit behavioural data would necessitate a highly complex experimental design involving multiple fully functional, domain-specific CRS deployments, which is currently infeasible. While our vignette approach successfully isolates domain variables from technical confounds, this static design cannot account for the temporal dynamics of conversational decision-making, where a user's interactional needs may evolve dynamically within a single session. Consequently, the framework's ultimate efficacy requires iterative validation through longitudinal field deployments and live A/B testing of policy rules. Such deployments are essential to confirm whether dynamic aim adaptation improves objective outcomes (e.g., task success, retention) in functioning LLM-powered or hybrid CRSs. Finally, our findings regarding dialogue-control preferences were bounded by an apparel e-commerce context. Future multi-domain studies are necessary to validate whether these user-led autonomy preferences hold in higher-stakes environments (e.g., finance or health), and to empirically test our theoretical assumption that affective support is most effective when system-led. Exploring these dimensions alongside richer, sequential interaction data will further refine the state-policy mapping required for fully adaptive, context-aware conversational recommendation.

6 Conclusion

This study advances the design of user-centred CRS through the formalisation of the *RAE* adaptation framework. By conceptualising recommendation as an unfolding, situated interaction rather than solely a prediction task, we provide a blueprint for aligning system behaviours with experiential quality. Empirically grounded in a multi-domain study ($N = 168$), our results quantify how users jointly prioritise three core interaction aims: *educative*, *explorative*, and *affective*. We demonstrate that these priorities are systematically modulated by domain context, perceived item value, and user-level traits. Educative aims dominate in high-stakes, cognitively complex domains such as housing and finance, while affective aims gain prominence in emotionally salient contexts such as wellness. Furthermore, perceived item value acts as a pivotal dynamic modulator, significantly elevating expectations across all interaction aims in high-stakes scenarios. While demographic influences appeared modest, it is important to note that our sample exhibited a strong age skew (73.2% aged 18–34, with no participants aged 65+). While age was not predictive in this cohort, future research targeting older demographics is required before drawing conclusions. Prior CRS experience and stable autonomy preferences emerged as robust signals for adaptive user modelling. However, as dialogue-control preferences were tested exclusively within an apparel e-commerce context, future cross-domain validation is necessary to confirm if these user-led autonomy preferences hold in higher-stakes environments.

The RAE framework operationalises these insights through a layered *state–policy–behaviour* mapping. By encoding contextual and user signals into structured state representations, the framework enables dialogue policies to adaptively modulate explanation depth, retrieval diversification, and affective tone. This shift from performance-centric optimisation to holistic experiential alignment allows for seamless integration with rule-based, hybrid, or LLM-based controllable generation. Future work should validate RAE through longitudinal field deployments to assess the stability of these experiential priorities over time. Additionally, exploring the integration of RAE state representations with adaptive policy learning or automated prompt-orchestration mechanisms will further strengthen its ability to deliver context-aware, experientially rich conversational recommendation in practice.

Acknowledgements

During the preparation of this work the authors used Gemini 3 in order to improve readability and language of this manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

[1] Mostafa Al-Emran and Charla Griffy-Brown. 2023. The role of technology adoption in sustainable development: Overview, opportunities, challenges, and future research agendas. *Technology in Society* 73 (2023), 102240.

[2] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 1 (1995), 289–300.

[3] Shlomo Berkovsky, Maurits C. Kaptein, and Massimo Zancanaro. 2016. Adaptivity and personalization in persuasive technologies. In *Proceedings of the International Workshop on Personalization in Persuasive Technology (PPT 2016)* . Rita Orji, Marc Reisinger, Markus Busch, Arie Dijkstra, Andre Stibe, and Manfred Tscheligi (Eds.), CEUR Workshop Proceedings, Vol. 1581, CEUR-WS.org, Salzburg, Austria, 13–25. Retrieved from https://hdl.handle.net/2066/161893 Held in conjunction with Persuasive Technology 2016.

[4] Hayet Brabra, Marcos Báez, Boualem Benatallah, Walid Gaaloul, Sara Bouguelia, and Shayan Zamanirad. 2021. Dialogue management in conversational systems: A review of approaches, challenges, and opportunities. *IEEE Transactions on Cognitive and Developmental Systems* 14, 3 (2021), 783–798.

[5] Wanling Cai, Yucheng Jin, and Li Chen. 2022. Impacts of personal characteristics on user trust in conversational recommender systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–14.

[6] Pablo Castells, Neil Hurley, and Saul Vargas. 2021. Novelty and diversity in recommender systems. In *Recommender Systems Handbook*. Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 603–646.

[7] Keyu Chen and Shiliang Sun. 2021. Knowledge-based conversational recommender systems enhanced by dialogue policy learning. In *Proceedings of the 10th International Joint Conference on Knowledge Graphs*. 10–18.

[8] Lu Chen, Cheng Chang, Zhi Chen, Bowen Tan, Milica Gašić, and Kai Yu. 2018. Policy adaptation for deep reinforcement learning-based dialogue management. In *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6074–6078.

[9] Minmin Chen. 2021. Exploration in recommender systems. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 551–553.

[10] Minmin Chen, Yuyan Wang, Can Xu, Ya Le, Mohit Sharma, Lee Richardson, Su-Lin Wu, and Ed Chi. 2021. Values of user exploration in recommender systems. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 85–95.

[11] Konstantina Christakopoulou, Filip Radlinski, and Katja Hofmann. 2016. Towards conversational recommender systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 815–824.

[12] Yashar Deldjoo, Z. He, J. McAuley, A. Korikov, S. Sanner, A. Ramisa, R. Vidal, M. Sathiamoorthy, A. Kasirzadeh, and S. Milano. 2024. A review of modern recommender systems using generative models (Gen-RecSys). In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 6448–6458.

[13] Yang Deng, Yaliang Li, Fei Sun, Bolin Ding, and Wai Lam. 2021. Unified conversational recommendation policy learning via graph-based reinforcement learning. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1431–1441.

[14] Soussan Djamasbi, Diane M. Strong, and Mark Dishaw. 2010. Affect and acceptance: Examining the effects of positive mood on the technology acceptance model. *Decision Support Systems* 48, 2 (2010), 383–394.

[15] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39, 2 (2007), 175–191.

[16] Batya Friedman and Helen Nissenbaum. 2022. Excerpt from bias in computer systems. In *Ethics of Data and Analytics*. Kirsten Martin (Ed.). Auerbach Publications, 20–26.

[17] Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten De Rijke, and Tat-Seng Chua. 2021. Advances and challenges in conversational recommender systems: A survey. *AI Open* 2 (2021), 100–126.

[18] Jianfeng Gao, Chenyan Xiong, Paul Bennett, and Nick Craswell. 2023. *Neural Approaches to Conversational Information Retrieval*. Vol. 44. Springer.

[19] Andrew Gelman and Cosma Rohilla Shalizi. 2013. Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology* 66, 1 (2013), 8–38.

[20] Ella Glikson and Anita Williams Woolley. 2020. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals* 14, 2 (2020), 627–660.

[21] Marc Hassenzahl. 2010. *Experience Design: Technology for All the Right Reasons*. Vol. 8. Morgan & Claypool Publishers.

[22] Zhankui He, Zhouhang Xie, Rahul Jha, Harald Steck, Dawen Liang, Yesu Feng, Bodhisattwa Prasad Majumder, Nathan Kallus, and Julian McAuley. 2023. Large language models as zero-shot conversational recommenders. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 720–730.

[23] David W. Hosmer Jr, Stanley Lemeshow, and Rodney X. Sturdivant. 2013. *Applied Logistic Regression*. John Wiley & Sons.

[24] Xueting Huang, Zhibo Zhang, Fusen Guo, Xianghao Wang, Kun Chi, and Kexin Wu. 2024. Research on older adults' interaction with e-health interface based on explainable artificial intelligence. In *Proceedings of the International Conference on Human-Computer Interaction*. Springer, 38–52.

[25] Anthony Jameson. 2007. Adaptive interfaces and agents. In *The Human-Computer Interaction Handbook*. Andrew Sears and Julie A. Jacko (Eds.). CRC Press, 459–484.

[26] Dietmar Jannach. 2023. Evaluating conversational recommender systems: A landscape of research. *Artificial Intelligence Review* 56, 3 (2023), 2365–2400.

[27] Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)* 54, 5 (2021), 1–36.

[28] Dietmar Jannach, Sidra Naveed, and Michael Jugovac. 2017. User control in recommender systems: Overview and interaction challenges. In *Proceedings of the 17th International Conference on E-Commerce and Web Technologies, EC-Web 2016, Revised Selected Papers 17*. Springer, 21–33.

[29] Yucheng Jin, Li Chen, Wanling Cai, and Xianglin Zhao. 2024. CRS-Que: A user-centric evaluation framework for conversational recommender systems. *ACM Transactions on Recommender Systems* 2, 1 (2024), 1–34.

[30] Bart P. Knijnenburg and Martijn C. Willemsen. 2015. Evaluating recommender systems with user experiments. In *Recommender Systems Handbook*. Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 309–352.

[31] Bart P. Knijnenburg, Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4–5 (2012), 441–504.

[32] Ahmet Baki Kocaballi, Shlomo Berkovsky, Juan C. Quiroz, Liliana Laranjo, Huong Ly Tong, Dana Rezazadegan, Agustina Briatore, and Enrico Coiera. 2019. The personalization of conversational agents in health care: systematic review. *Journal of Medical Internet Research* 21, 11 (2019), e15360.

[33] Sherrie Y. X. Komiak and Izak Benbasat. 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly* 30, 4 (2006), 941–960.

[34] Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen. 2016. A survey of serendipity in recommender systems. *Knowledge-Based Systems* 111 (2016), 180–192.

[35] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2020. Generating and understanding personalized explanations in hybrid recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10, 4 (2020), 1–40.

[36] Matthias Kraus, Nicolas Wagner, and Wolfgang Minker. 2021. Modelling and predicting trust for developing proactive dialogue strategies in mixed-initiative interaction. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. 131–140.

[37] William H. Kruskal and W. Allen Wallis. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* 47, 260 (1952), 583–621.

[38] Carol C. Kuhlthau, Jannica Heinström, and Ross J. Todd. 2008. The 'information search process' revisited: Is the model still useful. *Information Research* 13, 4 (2008), 13–4.

[39] Wai-Chung Kwan, Hong-Ru Wang, Hui-Min Wang, and Kam-Fai Wong. 2023. A survey on recent advances and challenges in reinforcement learning methods for task-oriented dialogue policy learning. *Machine Intelligence Research* 20, 3 (2023), 318–334.

[40] Raina Langevin, Ross J. Lordon, Thi Avrahami, Benjamin R. Cowan, Tad Hirsch, and Gary Hsieh. 2021. Heuristic evaluation of conversational agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[41] Nuno Laranjeiro, Seyma Nur Soydemir, and Jorge Bernardino. 2015. A survey on data quality: Classifying poor data. In *Proceedings of the 2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing (PRDC)*. IEEE, 179–188.

[42] Wenqiang Lei, Gangyi Zhang, Xiangnan He, Yisong Miao, Xiang Wang, Liang Chen, and Tat-Seng Chua. 2020. Interactive path reasoning on graph for conversational recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2073–2083.

[43] Yan Li, Xinyan Zhou, and Thomas K. F. Chiu. 2025. Systematics review on artificial intelligence chatbots and ChatGPT for language learning and research from self-determination theory (SDT): What are the roles of teachers? *Interactive Learning Environments* 33, 3 (2025), 1850–1864.

[44] Jianghao Lin, Xinyi Dai, Yunjia Xi, Weiwen Liu, Bo Chen, Hao Zhang, Yong Liu, Chuhan Wu, Xiangyang Li, Chenxu Zhu, et al. 2025. How can recommender systems benefit from large language models: A survey. *ACM Transactions on Information Systems* 43, 2 (2025), 1–47.

[45] Yuan Ma and Jürgen Ziegler. 2023. Initiative transfer in conversational recommender systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 978–984.

[46] Raj Mahmud, Shlomo Berkovsky, Mukesh Prasad, and A. Baki Kocaballi. 2025. Understanding user preferences for interaction styles in conversational recommender system: The predictive role of system qualities, user experience, and traits. In *Proceedings of the 37th Australian Conference on Human-Computer Interaction (OZCHI '25)*. 68–80. DOI: https://doi.org/10.1145/3764687.3764722

[47] Raj Mahmud, Yufeng Wu, Abdullah Bin Sawad, Shlomo Berkovsky, Mukesh Prasad, and A. Baki Kocaballi. 2025. Evaluating user experience in conversational recommender systems: A systematic review across classical and LLM-powered approaches. In *Proceedings of the 37th Australian Conference on Human-Computer Interaction (OZCHI '25)*. 81–93. DOI: https://doi.org/10.1145/3764687.3764714

[48] Ahtsham Manzoor, Samuel C. Ziegler, Klaus Maria Pirker Garcia, and Dietmar Jannach. 2024. ChatGPT as a conversational recommender system: A user-centric analysis. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. 267–272.

[49] Gary Marchionini. 2006. Exploratory search: From finding to understanding. *Communications of the ACM* 49, 4 (2006), 41–46.

[50] Adam W. Meade and S. Bartholomew Craig. 2012. Identifying careless responses in survey data. *Psychological Methods* 17, 3 (2012), 437.

[51] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)* 54, 6 (2021), 1–35.

[52] Mahsan Nourani, Joanie King, and Eric Ragan. 2020. The role of domain expertise in user trust and the impact of first impressions with intelligent systems. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 8. 112–121.

[53] Ayano Okoso, Keisuke Otaki, Satoshi Koide, and Yukino Baba. 2025. Impact of tone-aware explanations in recommender systems. *ACM Transactions on Recommender Systems* 3, 4 (2025), 1–34.

[54] Ayano Okoso, Mingzhe Yang, and Yukino Baba. 2025. Do expressions change decisions? Exploring the impact of AI's explanation tone on decision-making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.

[55] Reiner Pope, Sholto Douglas, Aakanksha Chowdhery, Jacob Devlin, James Bradbury, Anselm Levskaya, Jonathan Heek, Kefan Xiao, Shivani Agrawal, and Jeff Dean. 2022. Efficiently scaling transformer inference. In *Proceedings of the 6th Conference on Machine Learning and Systems (MLSys 2023)*, Miami, FL, USA.

[56] Dhanya Pramod and Prafulla Bafna. 2022. Conversational recommender systems techniques, tools, acceptance, and adoption: A state of the art review. *Expert Systems with Applications* 203 (2022), 117539.

[57] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the 5th ACM Conference on Recommender Systems*. 157–164.

[58] Atika Qazi, Najmul Hasan, Olusola Abayomi-Alli, Glenn Hardaker, Ronny Scherer, Yeahia Sarker, Sanjoy Kumar Paul, and Jaafar Zubairu Maitama. 2022. Gender differences in information and communication technology use & skills: A systematic review and meta-analysis. *Education and Information Technologies* 27, 3 (2022), 4225–4258.

[59] Aravind Sesagiri Raamkumar and Yinping Yang. 2022. Empathetic conversational systems: A review of current advances, gaps, and opportunities. *IEEE Transactions on Affective Computing* 14, 4 (2022), 2722–2739.

[60] Kanishk Rana, Rahul Madaan, and Jainendra Shukla. 2021. Effect of polite triggers in chatbot conversations on user experience across gender, age, and personality. In *Proceedings of the 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 813–819.

[61] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 3982–3992.

[62] Supraja Sankaran, Chao Zhang, Marisela Gutierrez Lopez, and Kaisa Väänänen. 2020. Respecting human autonomy through human-centered AI. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–3.

[63] Ashish Sharma, Inna W. Lin, Adam S. Miner, David C. Atkins, and Tim Althoff. 2021. Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach. In *Proceedings of the Web Conference 2021*. 194–205.

[64] Sneha and Shaina Raza. 2024. Affective computing for health management via recommender systems: Exploring challenges and opportunities. *Affective Computing for Social Good: Enhancing Well-being, Empathy, and Equity*. Muskan Garg and Rajesh Shardanand Prasad (Eds.). (2024), 163–182.

[65] Yueming Sun and Yi Zhang. 2018. Conversational recommender system. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 235–244.

[66] Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook*. Francesco Ricci, Lior Rokach, Bracha Shapira (Eds.). Springer, 353–382.

[67] Chun-Hua Tsai and Peter Brusilovsky. 2021. The effects of controllability and explainability in a social recommender system. *User Modeling and User-Adapted Interaction* 31, 3 (2021), 591–627.

[68] Svitlana Vakulenko, Ilya Markov, and Maarten de Rijke. 2017. Conversational exploratory search via interactive storytelling. In *Proceedings of the Search-Oriented Conversational AI Workshop (SCAI'17)*. Amsterdam, Netherlands.

[69] Thiemo Wambsganss, Tobias Kueng, Matthias Soellner, and Jan Marco Leimeister. 2021. ArgueTutor: An adaptive dialog-based learning system for argumentation skills. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[70] Nan Wang, Hongning Wang, Yiling Jia, and Yue Yin. 2018. Explainable recommendation via multi-task learning in opinionated text data. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 165–174.

[71] Kathrin Wardatzky, Oana Inel, Luca Rossetto, and Abraham Bernstein. 2025. Whom do explanations serve? A systematic literature survey of user characteristics in explainable recommender systems evaluation. *ACM Transactions on Recommender Systems* 3, 4 (2025), 1–35.

[72] Ryen W. White and Resa A. Roth. 2009. *Exploratory Search: Beyond the Query-Response Paradigm*. Number 3. Morgan & Claypool Publishers.

[73] Frank Wilcoxon. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1, 6 (1945), 80–83.

[74] Likang Wu, Zhi Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, et al. 2024. A survey on large language models for recommendation. *World Wide Web* 27, 5 (2024), 60.

[75] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. 2020. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems* 32, 1 (2020), 4–24.

[76] Kerui Xu, Jingxuan Yang, Jun Xu, Sheng Gao, Jun Guo, and Ji-Rong Wen. 2021. Adapting user preference to online feedback in multi-round conversational recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 364–372.

[77] Wei Xu, Jue Xiao, and Jianlong Chen. 2024. Leveraging large language models to enhance personalized recommendations in e-commerce. In *Proceedings of the 2024 International Conference on Electrical, Communication and Computer Engineering (ICECCE)*. IEEE, 1–6.

[78] Xi Yang and Marco Aurisicchio. 2021. Designing conversational agents: A self-determination theory approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[79] Yao-Chun Yang, Chiao-Ting Chen, Tzu-Yu Lu, and Szu-Hao Huang. 2023. Hierarchical reinforcement learning for conversational recommendation with knowledge graph reasoning and heterogeneous questions. *IEEE Transactions on Services Computing* 16, 5 (2023), 3439–3452.

[80] Steve Young, Milica Gašić, Blaise Thomson, and Jason D. Williams. 2013. POMDP-based statistical spoken dialog systems: A review. *Proceedings of the IEEE* 101, 5 (2013), 1160–1179.

[81] Zeping Yu, Jianxun Lian, Ahmad Mahmoody, Gongshen Liu, and Xing Xie. 2019. Adaptive user modeling with long and short-term preferences for personalized recommendation. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Vol. 7. 4213–4219.

[82] Ofir Zafrir, Guy Boudoukh, Peter Izsak, and Moshe Wasserblat. 2019. Q8bert: Quantized 8bit bert. In *Proceedings of the 2019 5th Workshop on Energy Efficient Machine Learning and Cognitive Computing-NeurIPS Edition (EMC2-NIPS)*. IEEE, 36–39.

[83] Xiaoyu Zhang, Ruobing Xie, Yougang Lyu, Xin Xin, Pengjie Ren, Mingfei Liang, Bo Zhang, Zhanhui Kang, Maarten de Rijke, and Zhaochun Ren. 2024. Towards empathetic conversational recommender systems. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 84–93.

[84] Yongfeng Zhang and Xu Chen. 2020. Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval* 14, 1 (2020), 1–101.

[85] Zhan Zhang, Yegin Genc, Dakuo Wang, Mehmet Eren Ahsen, and Xiangmin Fan. 2021. Effect of AI explanations on human perceptions of patient-facing AI-powered healthcare systems. *Journal of Medical Systems* 45, 6 (2021), 64.

[86] Zihuai Zhao, Wenqi Fan, Jiatong Li, Yunqing Liu, Xiaowei Mei, Yiqi Wang, Zhen Wen, Fei Wang, Xiangyu Zhao, Jiliang Tang, et al. 2024. Recommender systems in the era of large language models (LLMs). *IEEE Transactions on Knowledge and Data Engineering* 36, 11 (2024), 6889–6907.

[87] Xi Zhu, Yu Wang, Hang Gao, Wujiang Xu, Chen Wang, Zhiwei Liu, Kun Wang, Mingyu Jin, Linsey Pang, Qingsong Weng, et al. 2025. Recommender systems meet large language model agents: A survey. *Foundations and Trends® in Privacy and Security* 7, 4 (2025), 247–396.

Appendices

A Interaction Preferences Across Domains

To evaluate H1–H3, we tested whether perceived importance ratings for educative, explorative, and affective interaction varied across application domains. Participants rated each strategy on a 5-point Likert scale for each domain. Figure A.1 reports descriptive statistics (medians, modes, interquartile ranges), complementing the inferential analyses in Section 4.

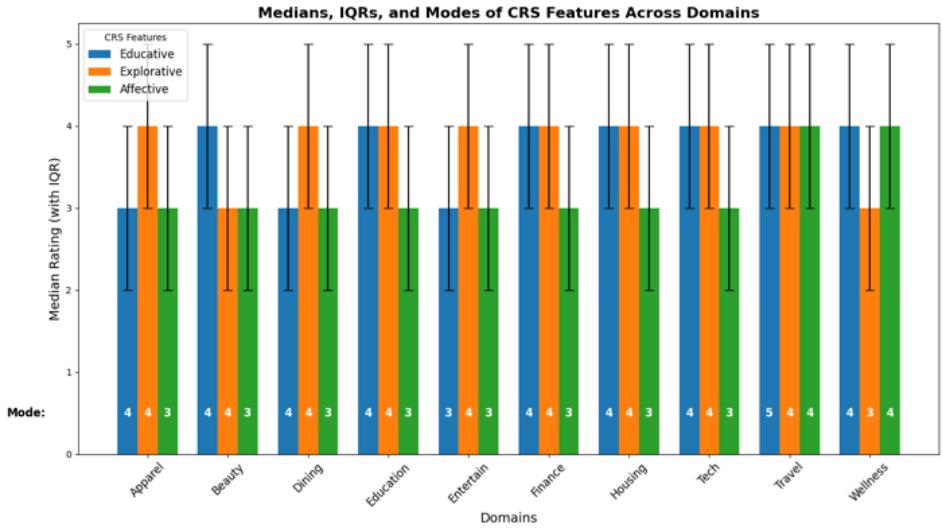


Fig. A.1. Descriptive distribution of ratings across domains. Medians (markers), modes (labels), and interquartile ranges (error bars) illustrate central tendency and variability for educative, explorative, and affective interaction. These statistics contextualise the domain-level effects reported in Section 4.

B Bayesian Posterior Predictive Checks

Posterior predictive checks (PPCs) were conducted for each model to evaluate whether simulated responses replicated the observed Likert-scale distributions (Figures B.1–B.3). In each plot, shaded densities show predicted ordinal response distributions, black lines indicate observed values, and dashed lines mark predictive means. Close alignment between predicted and observed distributions provides evidence of satisfactory model fit. Full Bayesian regression results are reported in Section 4.4.

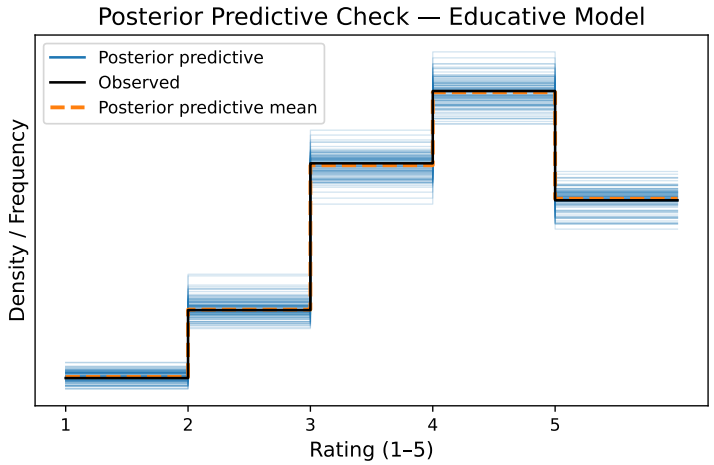


Fig. B.1. Posterior predictive check for the **Educative** model. Predicted distributions closely align with observed responses across all ordinal categories, indicating robust fit.

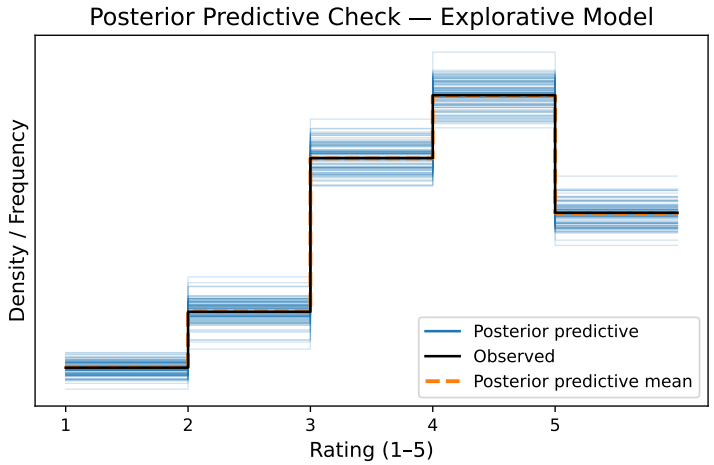


Fig. B.2. Posterior predictive check for the **Explorative** model. Predicted and observed distributions exhibit strong agreement, supporting model adequacy.

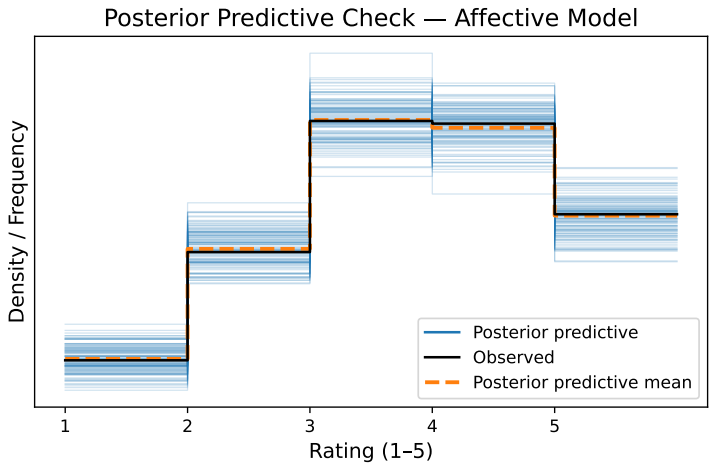


Fig. B.3. Posterior predictive check for the **Affective** model. Predicted distributions align well with observed data, providing evidence of model validity.

Received 20 January 2026; revised 1 June 2026; accepted 29 July 2026