

Adaptive Intent-Preference Arbitration for Conversational Recommendation

Dhayanidhi Rajasekaran^{1*}, Rajalakshmi N.R.²

¹Research Scholar, ²Professor, Department of Computer Science and Engineering, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Chennai, India.

E-mail: ^{1*} dhayanidhi.r@gmail.com, ² rajalakshmi234@gmail.com

Orcid ID: ^{1*} 0009-0008-6368-1495, ² 0000-0001-9705-0917

Abstract

Conversational recommender systems integrate long-term preferences (LTP) from previous sessions with short-term intent (STI) expressed in the current conversation. Fixed-rule fusion may produce suboptimal recommendations when these signals conflict. This paper proposes Adaptive Intent–Preference Arbitration (AIPA), which classifies their relationship into five categories: complement, consistent, conflict, override, and uncertain. Based on this classification, the model selects one of four actions: signal fusion, LTP prioritisation, STI prioritisation, or clarification. Candidates are re-ranked using learned, action-conditioned fusion weights. Additional components track recurring intent shifts across sessions and use counterfactual ablation analysis to identify the contribution of each signal. The framework is evaluated on ReDial, an English-language human–human movie recommendation corpus. Cross-session profiles are constructed from recurring seeker identities, while MovieLens genres provide item metadata. Since ReDial lacks relationship annotations, labels are generated using transparent weak-labeling rules and controlled synthetic injection covering all five categories. Results are reported separately for both sources. Training emphasises conflicting signals through conflict-weighted losses, label smoothing, and confidence-filtered self-training. All model-selection decisions use the validation split. Baselines are in-house reimplementations trained under a common protocol, limiting comparisons to this study. Across five random seeds using a pretrained sentence encoder, AIPA substantially outperforms profile-only, GRU, and SASRec baselines. However, its Hit@10 remains approximately one point below the intent-only and KBRD-style baselines across all turns and conflict subsets. Ablation results show negligible variation. The arbitration head achieves 0.99 clarification precision and a 1% wrong-override rate. Negative ranking results, an order-permutation control, MovieLens join coverage, and deployment measurements of peak memory and end-to-end latency are reported. Code, configuration files, and the executed notebook are released.

Keywords: Conversational Recommender Systems, User Intent, Long-Term Preference, Preference Conflict, Clarification, Counterfactual Analysis, Reproducibility.

1. Introduction

Conversational recommender systems allow users to express their immediate preferences, but determining how these requests should be balanced against established long-term preferences remains challenging. Prioritising long-term preferences (LTP) may

* Corresponding Author

Journal of Trends in Computer Science and Smart Technology, December 2026, Volume 8, Issue 4, Pages 1046-1074
DOI: <https://doi.org/10.36548/jtcsst.2026.4.007>

Received: 17.08.2026, received in revised form: 18.09.2026, accepted: 10.10.2026, published: 27.10.2026

© 2026 Inventive Research Organization. This is an open access article under the Creative Commons Attribution-NonCommercial International (CC BY-NC 4.0) License

disregard current requests, whereas prioritising short-term intent (STI) may neglect historical preferences. The appropriate balance depends on the context.

Most research on conversational recommendation systems (CRS) focuses on preference elicitation during dialogue [1]–[4] or grounding conversations in item knowledge [5], [6]. However, reconciling current intent with previously learned preferences has received comparatively less attention. Existing approaches often combine these signals using fixed rules or gating mechanisms that provide limited interpretability. Evaluation also rarely examines cases in which the signals conflict.

This paper investigates preference–intent reconciliation as an explicit arbitration problem. First, the model classifies the relationship between LTP and STI into five categories: complement, consistent, conflict, override, and uncertain. Second, the identified relationship determines an action: fuse the signals, prioritise one, or request clarification. The selected action determines the fusion weights. Third, counterfactual ablation measures ranking changes when each signal is removed, identifying its contribution to the recommendation without making causal claims about user behaviour.

The study uses ReDial [7], an English-language human–human movie dialogue corpus. Chronological seeker histories are constructed from recurring crowd-worker identities under strict no-look-ahead constraints. Since ReDial lacks relationship annotations, weak-rule labels and controlled synthetic conflicts are evaluated separately. Evaluation includes bootstrap confidence intervals and paired statistical tests with multiple-comparison correction.

Results show that AIPA outperforms profile-only, GRU, and SASRec baselines and matches a KBRD-style baseline on conflict turns. However, an intent-only recommender performs significantly better on the natural test set. Although arbitration behaviour is consistent with the intended design on controlled conflicts, none of the ablations achieves statistically significant improvements. Section 8 discusses the requirements for conclusive validation.

2. Related Work

2.1 Conversational Recommendation

Early conversational recommender system (CRS) research formulated dialogue as a preference elicitation process in which the system asks questions about item attributes and users provide responses [1], [2]. Estimation–Action–Reflection [3] modelled the choice between asking questions and making recommendations as a policy optimisation problem. Zhang et al. [4] also incorporated clarification questions as actions within a system–user interaction framework. The clarification mechanism in this study follows this approach but differs in its triggering condition: it considers the relationship between two evidence sources rather than attribute uncertainty alone.

The introduction of ReDial [7] facilitated research on open-ended, natural-language conversational recommendation. KBRD [5] and KGSF [6] subsequently incorporated knowledge graphs to ground conversations in item-related knowledge. Other datasets include INSPIRED [8] and the Chinese TG-ReDial [9]. Although TG-ReDial provides user-level histories, its dialogues are in Chinese. ReDial was selected because its English-language dialogues support the analysis of data, labels, case studies, and recommendation errors, while cross-session histories can be reconstructed from recurring seeker identities. Surveys by

Jannach et al. [10] and Gao et al. [11] review developments in the field and identify the reconciliation of conversational intent with historical user preferences as an open research problem.

2.2 Long- and Short-Term Signals

Session-based and sequence-aware recommendation models [12]–[14] implicitly balance recent and historical behaviour through recurrent architectures or attention mechanisms. The sequential baseline in this study follows this approach. In contrast, AIPA explicitly models the weighting process through discrete, weakly supervised, and interpretable relationship classes and arbitration actions. The fusion weights are conditioned on these classes and actions.

2.3 Counterfactual Reasoning in Recommendation

The driver diagnostic adopts the terminology of interventions [15]. However, the intervention is applied to the trained model rather than the user. Removing the LTP encoding and recalculating recommendation scores measures the model’s dependence on that pathway. This procedure constitutes ablation-based attribution. It does not establish causal effects on user behaviour; therefore, the term effect is avoided when reporting these measurements.

3. Problem Formulation

This section fixes notation and states the prediction problem as the released code constructs it. Table A1 collects the symbols. Vectors are bold lower case (h), matrices bold upper case (W), sets calligraphic ($\mathcal{J}$), and $\mathbb{1}[\cdot]$ is the indicator. When v has non-negative entries, $nrm(v) = v / \mathbb{1}^\top v$ if $\mathbb{1}^\top v > 0$ and $nrm(0) = 0$. Genre vectors thus live in $\bar{\Delta}^{|\mathcal{G}|} = \Delta^{|\mathcal{G}|} \cup \{0\}$, the simplex plus a “no evidence” point.

3.1 Seekers, sessions and the Temporal Boundary

ReDial is a set of dialogues. Each is a message sequence $D = (m_1, \dots, m_T)$. Every message carries a role (seeker or recommender), English text and the set $M(m) \subseteq \mathcal{J}$ of movies it mentions. Every dialogue has a seeker identifier u and an integer conversation identifier c . The seeker marks each mentioned movie as liked, seen and suggested. Let $\mathcal{L}(D) \subseteq \mathcal{J}$ denote the movies the seeker marked as liked or seen in D .

Conversation identifiers were assigned in sequence during collection. They are treated as the chronological order of a seeker’s sessions. This is an assumption (see Section 8). The history available at dialogue c of seeker u is

$$\mathcal{H}_u(c) = \{D_{c'}^u : c' < c\}, \quad (1)$$

the set of u ’s dialogues with a smaller identifier. Nothing from D_c^u or any later dialogue enters the long-term side.

3.2 Instances

A recommendation instance is a triple (c, t, y) . At turn $t > 1$ of dialogue c the recommender mentions a movie $y \in \mathcal{J}$ not seen earlier in the dialogue,

$$y \in M(m_t) \setminus \bigcup_{t' < t} M(m_{t'}), \quad \text{role}(m_t) = \text{recommender}. \quad (2)$$

A recommender message with several new movies yields one instance per movie. The task is to rank the catalogue $\mathcal{J}$ so that y is placed as high as possible, using only the two evidence

bundles below.

3.2.1 Long-Term Preference (LTP)

The liked/seen items are taken from $\mathcal{H}_u(c)$ in chronological order, keeping the most recent $L_{\max} = 50$. Call the result $h = (h_1, \dots, h_L)$, with h_L the most recent. Also kept are the last ten seeker utterances in $\mathcal{H}_u(c)$ that match a preference pattern (“I usually like...”, “my favourite...”) and are shorter than 200 characters. Their concatenation is the profile text $\mathcal{P}$. The LTP genre vector is a recency-decayed mixture of the items’ genre indicators,

$$\pi_L = nrm \left(\sum_{j=1}^L \gamma^{L-j} \frac{g(h_j)}{|G(h_j)|} \right), \quad g(i)_g = \mathbb{1}[g \in G(i)], \quad (3)$$

with $\gamma = 0.9$. The most recent item has weight one. Each older item is discounted by a further factor γ . Items with no MovieLens match ($G(i) = \emptyset$) contribute nothing. A seeker with $L < L_{\min} = 3$ history items is *cold*.

3.2.2 Short-Term Intent (STI)

3.2.0.2 Short-Term Intent (STI)

Let $S_{<t}$ be the seeker messages before turn t , R_t the concatenated text of the last 6 of them, and l_t the text of the last one. Let $\mathcal{M}_{<t}$ be the movies mentioned in $S_{<t}$ and $\mathcal{M}_{<t}^+ \subseteq \mathcal{M}_{<t}$ those the seeker liked. Title tokens are replaced by their surface form, so “@12345” becomes the movie name. The STI genre vector combines lexicon counts with the genres of liked in-dialogue movies,

$$\pi_S = nrm \left(n(R_t) + nrm \left(\sum_{i \in \mathcal{M}_{<t}^+} \frac{g(i)}{|G(i)|} \right) \right), \quad (4)$$

where $n(R_t)_g$ counts occurrences in R_t of the keywords of a small English lexicon for genre g (“scary”, “slasher”, “zombie” for Horror, and so on). Eight scalar flags summarise lexical and length evidence,

$$f = \left(f_{sti}, f_{ltp}, f_{neg}, f_{req}, \frac{|S_{<t}|}{10}, \frac{|\mathcal{M}_{<t}|}{5}, f_{cold}, \frac{L}{L_{\max}} \right), \quad (5)$$

Here f_{sti} marks an explicit departure cue (“for a change”, “tonight”, “with my kids”) and f_{ltp} an explicit habit cue (“as usual”, “stick with”) in R_t . f_{neg} marks a negated genre keyword (“not into horror”). f_{req} marks a request pattern in l_t , and $f_{cold} = \mathbb{1}[L < L_{\min}]$.

The STI side uses $S_{<t}$ only. The LTP side uses $\mathcal{H}_u(c)$ only. The model sees neither the target turn nor any later data.

3.3 Relationships, Actions and Shifts

There are five relationship classes $\mathcal{R} = \{\text{Complement, Consistent, Conflict, Override, Uncertain}\}$. In *Consistent* turns STI points where LTP already points. In *Complement* turns STI adds a genre LTP does not cover, without contradicting it. In *Conflict* turns STI opposes LTP and the user shows no sign of knowing it. In *Override* turns STI opposes LTP and the user flags the departure. In *Uncertain* turns one side carries too little evidence. The action set is $\mathcal{A} = \{\text{Fuse, Prioritise_LTP, Prioritise_STI, Ask_Clarification}\}$. A fixed default map $A_0: \mathcal{R} \rightarrow \mathcal{A}$ sends Consistent and Complement to Fuse, Conflict and Override to Prioritise_STI, and

Uncertain to Ask_Clarification. A learned policy may depart from A_0 .

A departure from habit in one session is a *temporary override*. If the same departure wins arbitration in at least k sessions of the same seeker, it is a *persistent shift* and is folded into LTP (Section 4.8).

4. Method

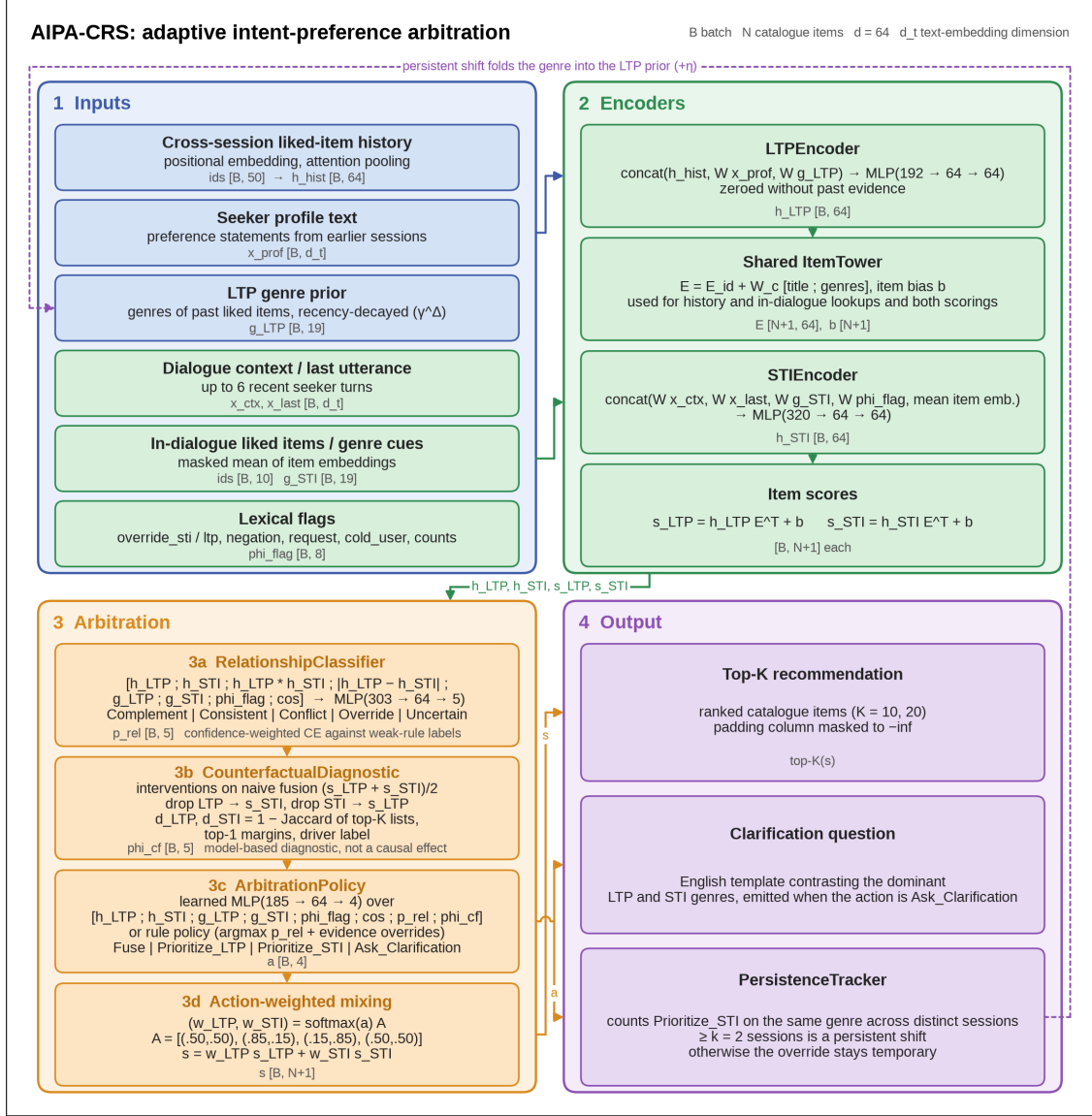


Figure 1. AIPA architecture

Figure 1 illustrates the proposed framework. Panel 1 presents the inputs. The long-term pathway uses cross-session liked-item history, profile text, and a recency-decayed genre prior. The short-term pathway uses dialogue context, in-dialogue liked items, and lexical flags. Panel 2 contains the LTP encoder, STI encoder, and shared item tower, which scores catalogue items for either query. Panel 3 presents the arbitration stage, comprising the relationship classifier (3a), counterfactual diagnostic (3b), arbitration policy (3c), and action-conditioned fusion weights (3d). Panel 4 presents the outputs, comprising a top- K recommendation list or a clarifying question, and the persistence tracker, which incorporates recurring intent shifts into the LTP genre prior through the dotted path. The joint objective in Eq. (30) trains all

stages simultaneously. Tensor dimensions represent batches of B turns across N catalogue items, with $d = 64$. All encoders are lightweight, emphasising arbitration logic rather than representational capacity. The item tower, embedding dimension ($d = 64$), and training protocol are shared across all compared systems to isolate the contribution of the arbitration design. Throughout, $MLP(z) = W_2 GELU(W_1 z + b_1) + b_2$ with hidden width d and dropout 0.1 after the activation. Each occurrence has its own parameters.

4.1 Text and Item Representations

Text is encoded by a frozen map ϕ fixed before training. In the reported run ϕ is the sentence encoder all-MiniLM-L6-v2 [16]. Its mean-pooled output is l_2 -normalised, so $d_t = 384$. The code also offers a lighter option fitted on the training split. It uses word uni- and bigram TF-IDF with sublinear term frequency, then truncated SVD,

$$\phi(x) = \frac{\psi(x)}{\|\psi(x)\|_2 + 10^{-8}}, \quad \phi(\text{empty}) = 0, \quad (6)$$

with ψ either the MiniLM output or $V^\top tfidf(x)$. Embeddings are cached on disk by content hash. A re-run encodes only new strings. Item i has a content vector $x_i = [\phi(\text{title}_i); g(i)] \in \mathbb{R}^{d_t + |G|}$ and the item tower produces

$$e_i = E_i + W_c x_i + b_c, \quad b_i \in \mathbb{R}, \quad (7)$$

with a free embedding table $E \in \mathbb{R}^{|J| \times d}$, a shared projection W_c and a free per-item bias b_i . For any query $h \in \mathbb{R}^d$ the score of item i is the biased inner product

$$s_i(h) = h^\top e_i + b_i, \quad (8)$$

and $s(h) \in \mathbb{R}^{|J|}$ collects the scores of the whole catalogue. The padding index is masked to $-\infty$.

4.2 LTP Encoder

The history h is right-aligned in a window of length $L_{\max}$. Each position j receives a learned positional vector u_j . The encoder pools the item vectors with one learned attention query $a \in \mathbb{R}^d$,

$$\tilde{e}_j = e_{h_j} + u_j, \quad \alpha_j = \frac{\exp(a^\top \tilde{e}_j)}{\sum_{j'=1}^L \exp(a^\top \tilde{e}_{j'})} \quad (9)$$

$$h_{hist} = \sum_{j=1}^L \alpha_j \tilde{e}_j \quad (10)$$

The softmax runs over non-padding positions only, and $h_{hist} = 0$ when $L = 0$. The positional vectors let the attention prefer recent or old items. The recency discount itself acts on the genre vector through γ in Eq. (3), not on the attention weights. The encoding is

$$h_L = \mathbb{1}[\text{LTP evidence}] \cdot MLP([h_{hist}; W_p \phi(\mathcal{P}); W_g \pi_L]), \quad (11)$$

The indicator is one if at least one of h , π_L or $\phi(\mathcal{P})$ is non-zero. A seeker with no history has $h_L = 0$. By Eq. (8) its LTP score vector equals the item biases.

4.3 STI Encoder

The intent side has no sequence to attend over. It concatenates five projected views of the current dialogue. One is a masked mean of the liked in-dialogue items (at most the last ten),

$$\bar{e}_{cur} = \frac{1}{\max(1, |\mathcal{M}_{<t}^+|)} \sum_{i \in \mathcal{M}_{<t}^+} e_i \quad (12)$$

$$h_S = MLP([W_r \phi(R_t); W_l \phi(l_t); W'_g \pi_S; W_f f; \bar{e}_{cur}]) \quad (13)$$

Genre lexicon. The intent genre vector π_S counts matches to a manually constructed English lexicon in recent seeker turns and incorporates the genres of liked in-dialogue items. The lexicon maps the 18 MovieLens genres to 132 distinct keywords and phrases, including 36 multiword expressions. Each genre includes its name, common inflections (e.g., “comedies” and “thrillers”), colloquial synonyms (e.g., “scary,” “slasher,” “funny,” and “tearjerker”), franchise cues (e.g., “marvel” and “star wars”), and audience-related expressions (e.g., “for my kids” and “date night”). Entries were compiled by examining training dialogues and identifying phrases that clearly indicate specific genres. Phrases associated with multiple genres were assigned to each relevant category, such as “rom com” for both Comedy and Romance. Matching is performed on whole words after converting text to lowercase. Two additional lists contain 38 departure markers (e.g., “for a change” and “something different”) and 20 habit markers (e.g., “as always” and “like I usually do”). The genre lexicon matches 59% of natural test turns, while incorporating liked-item genres provides genre evidence for 81% of turns. The lexicon serves exclusively as a feature extractor, not as a source of labels. Its inability to identify genre evidence in some turns contributes to the large proportion assigned to the Uncertain class.

4.4 Scoring and the Fusion Family

Both encodings are scored against the catalogue with Eq. (8), giving $s_L = s(h_L)$ and $s_S = s(h_S)$. Every compared system ranks by a convex combination

$$s = w_L s_L + w_S s_S, \quad w_L + w_S = 1, \quad w_L, w_S \geq 0. \quad (14)$$

The score is linear in the query, so Eq. (14) equals $s(w_L h_L + w_S h_S)$. Mixing scores and mixing encodings are the same operation, and the item bias enters once. What separates one system from another is simply the rule that picks (w_L, w_S) .

$$(w_L, w_S) = \begin{cases} (1, 0) & \text{LTP-only,} \\ (0, 1) & \text{STI-only,} \\ (\alpha, 1 - \alpha) & \text{naive fusion, } \alpha = 0.5, \\ (\sigma(z), 1 - \sigma(z)) & \text{adaptive fusion,} \\ q^\top W_A & \text{AIPA.} \end{cases} \quad (15)$$

The naive weight is also swept over $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$ at test time (Table 7). The sweep uses naive fusion on the natural test set unless stated. Adaptive fusion learns a scalar gate

$$z = MLP([h_L; h_S; \pi_L; \pi_S; f]) \in \mathbb{R}, \quad (16)$$

It has no notion of relationship or action. AIPA replaces the gate by a distribution $q \in \Delta^4$ over actions (Section 4.7) and a table that states what each action means for the weights. Each row of the table is the softmax of two free logits $l_a \in \mathbb{R}^2$,

$$[W_A]_a = softmax(l_a), \quad W_A^{(0)} = \begin{pmatrix} 0.50 & 0.50 \\ 0.85 & 0.15 \\ 0.15 & 0.85 \\ 0.50 & 0.50 \end{pmatrix} \begin{matrix} \text{Fuse} \\ \text{Prioritise_LTP} \\ \text{Prioritise_STI} \\ \text{Ask_Clarification} \end{matrix} \quad (17)$$

so $w_L = \sum_a q_a [W_A]_{a1}$ and the constraint in Eq. (14) holds by construction. The logits start at the hand-set table $W_A^{(0)}$ and are trained end-to-end through the recommendation loss (learned_action_weights). The fixed table remains a configuration option. Each action keeps one inspectable weight pair, but the pair is fitted. In the reported run the Prioritise_STI row moved to about 0.22/0.78 and the Fuse row stayed near 0.5/0.5 (Section 7.9). An optional instance-level residual $g \in \mathbb{R}$ from the policy, $w_L \leftarrow \sigma(\text{logit}(w_L) + g)$, is implemented but off in every reported run. A clarifying turn still ranks the catalogue with the Ask_Clarification row, so an offline ranking metric exists for every instance.

4.5 Relationship Classifier

The classifier sees both encodings, their element-wise interactions, the raw genre vectors and flags, and the cosine of the encodings,

$$\varphi_{rel} = [h_L; h_S; h_L \odot h_S; |h_L - h_S|; \pi_L; \pi_S; f; \cos(h_L, h_S)], \quad (18)$$

$$r = MLP(\varphi_{rel}) \in \mathbb{R}^5, \quad p = softmax(r), \quad (19)$$

with $\cos(x, y) = x^\top y / \max(\|x\| \|y\|, 10^{-6})$. The posterior p reaches the policy through a stop-gradient. The action loss cannot reshape the relationship head. That head learns only from its own (weak) labels.

4.6 Counterfactual Driver Diagnostic

The diagnostic asks how much a ranking depends on each pathway. It answers by intervening on the model, not the data. The reference ranking is the naive fusion of the two component scores. Removing a pathway means scoring with the other pathway alone,

$$s_\emptyset = \frac{1}{2}(s_L + s_S), \quad s_\emptyset|_{do(L=\emptyset)} = s_S, \quad s_\emptyset|_{do(S=\emptyset)} = s_L. \quad (20)$$

This is a convention, not a literal $h_L \leftarrow 0$. The literal version would give $\frac{1}{2}h_S^\top e_i + b_i$ and double the weight of the item bias. The convention keeps the interventional list equal to what the STI or LTP tower would recommend alone. Disruption is one minus the Jaccard overlap of the factual and interventional top- K lists ($K = 10$),

$$J_K(s, s') = \frac{|Top_K(s) \cap Top_K(s')|}{|Top_K(s) \cup Top_K(s')|}, \quad (21)$$

$$\delta_L = 1 - J_K(s_\emptyset, s_S), \quad \delta_S = 1 - J_K(s_\emptyset, s_L), \quad (22)$$

so δ_L measures how far the list moves when LTP is removed, and δ_S when STI is removed. With the top-1 margins $m(s) = s_{(1)} - s_{(2)}$ of the three score vectors they form the feature vector handed to the policy,

$$c = (\delta_L, \delta_S, m(s_\emptyset), m(s_S), m(s_L)) \in \mathbb{R}^5, \quad (23)$$

computed without gradient. The per-turn driver label uses a threshold $\tau = 0.1$ and a dominance ratio $\rho = 1.5$.

$$driver = \begin{cases} \text{Neither} & \delta_L < \tau \wedge \delta_S < \tau, \\ \text{STI} & \delta_S \geq \tau \wedge \delta_S \geq \rho \delta_L, \\ \text{LTP} & \delta_L \geq \tau \wedge \delta_L \geq \rho \delta_S, \\ \text{Jointly} & \text{otherwise,} \end{cases} \quad (24)$$

The cases are tested in the order written. Each label says which input the fitted model leans on, nothing more. Nothing here identifies an effect on users. A second, post-training probe re-scores the *fused* model of Eq. (14) under $h_L \leftarrow 0$ and $h_S \leftarrow 0$. This also changes p , q and hence the weights. It reports rank and NDCG changes and the plain overlap $|Top_{10} \cap Top'_{10}|/10$. That is the quantity in Table 6.

4.7 Arbitration Policy and Clarification Trigger

4.7.1 Learned Policy

The action logits are a function of the encodings, the evidence, the detached relationship posterior and the counterfactual features,

$$\varphi_{act} = [h_L; h_S; \pi_L; \pi_S; f; \cos(h_L, h_S); sg(p); c], \quad (25)$$

$$o = MLP(\varphi_{act}) \in \mathbb{R}^4, \quad q = softmax(o), \quad (26)$$

where sg is the stop-gradient. The ablations drop $sg(p)$ or c from Eq. (25). The “without clarification” ablation sets $o_{Ask} = -\infty$ before the softmax. The weights follow from Eq. (17), and the emitted action is $\hat{a} = \arg \max_a q_a$.

4.7.2 Rule Policy

The rule variant replaces Eq. (26) by a deterministic function of the relationship posterior and two evidence tests. Let $\hat{r} = \arg \max_r p_r$, $\kappa = \max_r p_r$, $\theta = 0.5$, $cold = \mathbb{1}[f_{cold} > \frac{1}{2}]$ and $noSTI = \mathbb{1}[\pi_S = 0]$. Then

$$\hat{a} = \begin{cases} \text{Prioritise_STI} & cold \wedge \neg noSTI, \\ \text{Ask_Clarification} & \kappa < \theta, \\ A_0(\hat{r}) & \text{otherwise,} \end{cases} \quad (27)$$

The first case takes precedence. When clarification is disabled, the second case is dropped, any Ask_Clarification from A_0 becomes Fuse, and a non-cold seeker with $noSTI$ goes to Prioritise_LTP. The rule output becomes logits $o = 20 \cdot \mathbb{1}_{\hat{a}}$. Its softmax is near one-hot, so Eq. (17) returns the weight row of $\hat{a}$ almost exactly.

4.7.3 Clarification

A clarifying question is emitted whenever $\hat{a} = \text{Ask_Clarification}$. For the learned policy this is the arg-max of Eq. (26). For the rule it is the second case of Eq. (27). The question is filled from a small English template set using the arg-max genres of π_L and π_S . A typical instance reads “*Earlier you enjoyed drama movies, but right now you seem interested in horror. Should I focus on horror for tonight, or find something that fits both?*” Offline the answer cannot be observed. The question is judged by whether the reference label called for it (Section 6.2).

4.8 Persistence Tracker

At inference the natural test instances are visited in order of (seeker, conversation id, turn). For each seeker u and genre g the tracker keeps the set $\mathcal{C}_{u,g}$ of sessions whose first instance was resolved by Prioritise_STI with g the arg-max of π_S ,

$$\mathcal{C}_{u,g} \leftarrow \mathcal{C}_{u,g} \cup \{c\} \quad \text{if } \hat{a} = \text{Prioritise_STI} \wedge g = \arg \max_{g'} \pi_{S,g'}. \quad (28)$$

Before an instance of seeker u is scored, its LTP genre vector is replaced by

$$\pi'_L = \text{nrm} \left(\pi_L + \eta \sum_{g: |\mathcal{C}_{u,g}| \geq k} 1_g \right), \quad (29)$$

with $k = 2$ and $\eta = 0.3$, where 1_g is the unit vector of genre g . Only sessions already visited count, so the adjustment never uses the future. Each time $|\mathcal{C}_{u,g}|$ first reaches k , a persistent shift (u, g, c) is logged. The model is then re-run with π'_L in place of π_L . Encoders and policy stay as trained. The tracker can therefore only reach the output through the genre prior, that is through Eqs. (11), (18), (16) and (25). Synthetic instances are excluded.

4.9 Training Objective

Each instance n has a target y_n . For AIPA variants, it also has a relationship label $r_n \in \mathcal{R}$, an action label $a_n \in \mathcal{A}$, and a confidence score $\omega_n \in [0, 1]$ assigned by the labelling procedure described in Section 5. Synthetic instances have $\omega_n = 1$. Let $CE(\mathbf{z}, y) = -\log \text{softmax}(\mathbf{z})_y$, and let CE_e denote its label-smoothed version. The mini-batch loss is defined as

$$\mathcal{L} = \frac{1}{B} \sum_{n=1}^B [\nu_n \mathcal{L}_{\text{rec}}(\mathbf{s}_n, y_n) + \omega_n (\lambda_{\text{rel}} CE_e(\mathbf{r}_n, r_n) + \lambda_{\text{act}} CE(\mathbf{o}_n, a_n))]. \quad (30)$$

Three components address turns involving disagreement between LTP and STI. Conflict weighting. The weight ν_n is set to 2.0 for instances labelled Conflict or Override, whether weak-rule or synthetic, and to 1 otherwise. The weights are normalised to have a batch mean of one, preserving the overall loss scale. This option is controlled by `conflict_loss_weight`, with a value of 1 corresponding to the unweighted loss.

Recommendation loss. The recommendation loss $\mathcal{L}_{\text{rec}}$ can use the full catalogue softmax, sampled softmax with in-batch positive items and 256 uniformly sampled negative items, or Bayesian personalised ranking (BPR) [17]. These alternatives were evaluated using the validation split of the development configuration. AIPA achieved validation Hit@10 scores of 0.123 with the full softmax, 0.072 with sampled softmax, and 0.046 with BPR. Accordingly, all reported experiments use the full softmax.

Self-training. From epoch 3 onwards, weak-rule relationship labels with $\omega_n < 0.6$ are replaced by the model's own predictions when the maximum softmax probability is at least 0.9. Relabelling decisions are always evaluated against the original weak labels to prevent cumulative drift. Synthetic labels remain unchanged. The action label is updated according to the default mapping A_0 for the revised relationship. In the reported experiments, 150-1,400 of the 36,883 training instances per epoch were relabelled, corresponding to 0.4-3.8%. This number increased as the classifier became more confident.

The relationship and action auxiliary losses are included only in AIPA variants, and

the action loss is used only with the learned policy. Each auxiliary term is weighted by the label confidence, reducing the influence of low-confidence weak labels. The values of λ_{rel} , λ_{act} , the conflict weight, the label-smoothing parameter, and the self-training setting were selected using the validation split. Natural validation Hit@10 was the primary selection criterion, with conflict-subset Hit@10 and macro-F1 also reported. The tested combinations, $\lambda_{\text{rel}} \in \{0.5, 1\}$ and $\lambda_{\text{act}} \in \{0.1, 0.6\}$, produced Hit@10 variations within 0.002. The test split was not used for model selection.

For the without clarification ablation, instances labelled Ask_Clarification are remapped to Fuse in Eq. (30), and the corresponding action is masked in Eq. (26). Thus, this ablation modifies the training objective as well as the action-decoding process.

All models use AdamW [18] with a learning rate of 0.003, weight decay of 10^{-5} , batch size 256, and gradient-norm clipping at 5. Training proceeds for a maximum of 20 epochs, with early stopping based on validation Hit@10 using a patience of three epochs. The parameters from the best epoch are restored. Baselines that do not use fusion, including GRU and the conversation-aware model, share the item tower defined in Eq. (8) and the recommendation-loss component of Eq. (30).

5. Data and Labels

5.1 ReDial and Cross-Session Histories

ReDial [7] contains 11,348 English-language dialogues between seekers and recommenders, comprising 10,006 training dialogues and 1,342 test dialogues. Each mentioned movie includes seeker annotations indicating whether it was liked, seen, or suggested. A further 1,000 training dialogues are allocated to validation. The corpus contains 764 distinct seeker identifiers, and many seekers participate in multiple dialogues (Fig. 2). Figure 2 also presents dialogue counts per split, genre coverage following the MovieLens join, and the distribution of turns per dialogue. Dialogues for each seeker are ordered by conversation identifier, which is assumed to represent chronological order. This assumption could not be verified against timestamps and is discussed as a limitation. For dialogue D_{m+1} , LTP is constructed exclusively from $D_1, \dots, D_m$. A seeker with fewer than three preceding sessions is classified as cold.

Movie metadata, including genres and release years, are obtained from MovieLens ml-latest [19] and joined using normalised titles and release years. The join covers 86.7% of the 6,924 titles in the ReDial catalogue and 87.2% of titles appearing in dialogues. Coverage is higher when weighted by occurrence: 93.7% of all movie mentions and 93.6% of recommender mentions, which constitute the recommendation targets, have genre metadata. Unmatched titles are primarily obscure or ambiguous entries. They retain empty genre vectors and remain in the candidate catalogue without imputation. Section 7.7 evaluates the chronological-order assumption using an order-permutation control.

Each recommendation instance is represented by a dialogue, turn, and target-movie triple in which the recommender introduces a movie that has not previously been mentioned. Only the preceding six turns and up to 50 historical items are used. All dialogues in each split are included using a fixed random seed, yielding 36,883 training instances and 4,472 natural test instances.

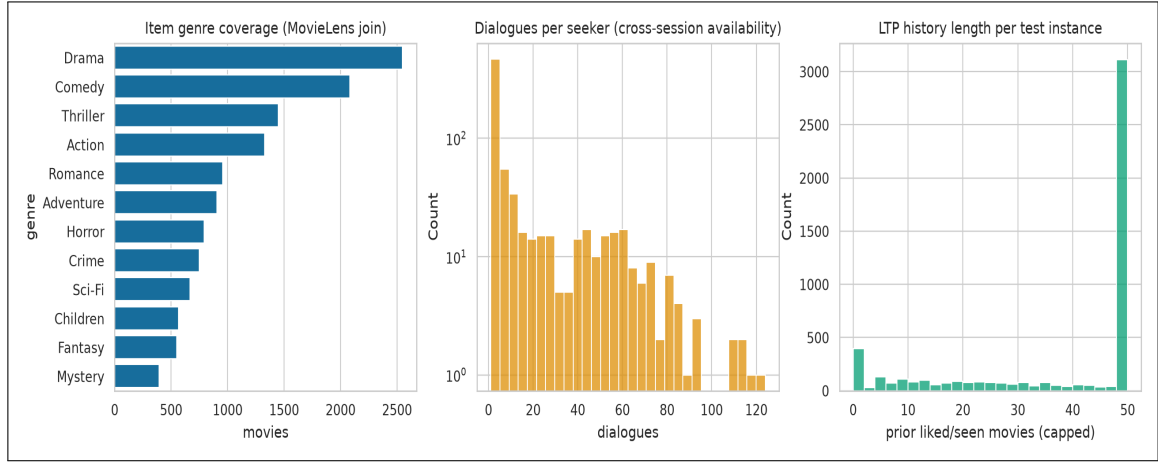


Figure 2. ReDial overview

5.2 Relationship Labels

ReDial has no relationship annotations. Two label sources are used and kept apart.

5.2.1 Weak-rule Labels

A deterministic function of (π_L, π_S, R_t, f) assigns a relationship r_n , an action a_n and a confidence ω_n . Let $sim = \cos(\pi_L, \pi_S)$. Let $T(\pi)$ be the (at most two) largest genres of π with mass at least 0.15. Let $hasL = \mathbb{1}[\pi_L \neq 0 \wedge \neg f_{cold}]$. Three lexical tests read R_t . They are a habit marker (f_{ltp}), a departure marker (f_{sti}), and $negH = \mathbb{1}[\exists g \text{ negated in } R_t: \pi_{L,g} \geq 0.2]$. Two structural tests compare the genre vectors. $tens = \mathbb{1}[\exists s \in T(\pi_S), g \in T(\pi_L): \{s, g\} \in \mathcal{T} \wedge \pi_{L,s} < 0.1]$ uses a fixed list $\mathcal{T}$ of genre pairs that rarely share a request (Horror–Children, Comedy–War, ...). $new = \mathbb{1}[\exists s \in T(\pi_S): \pi_{L,s} < 0.05]$. Write C/O for Override if $f_{sti} = 1$ and Conflict otherwise. The rule is read top to bottom.

$$(r_n, a_n, \omega_n) = \begin{cases} (\text{Uncertain}, \text{Ask}, 0.9) & \pi_S = 0 \wedge \neg hasL, \\ (\text{Uncertain}, \text{Ask}, 0.6) & \pi_S = 0, \\ (\text{Uncertain}, \text{Prio_STI}, 0.6) & \neg hasL, \\ (\text{Consistent}, \text{Fuse}, 0.8) & f_{ltp} = 1, \\ (C/O, \text{Prio_STI}, 0.8) & negH, \\ (C/O, \text{Prio_STI}, \omega_t) & tens \wedge sim < 0.25, \\ (\text{Consistent}, \text{Fuse}, \omega_s) & sim \geq 0.5, \\ (\text{Complement}, \text{Fuse}, \omega_n) & new, \\ (\text{Consistent}, \text{Fuse}, 0.5) & sim \geq 0.25, \\ (\text{Uncertain}, \text{Ask}, 0.5) & \text{otherwise,} \end{cases} \quad (31)$$

with $\omega_s = 0.5 + 0.5 sim$, $\omega_n = 0.5 + 0.3(1 - sim)$, $\omega_t = 0.75$ if $f_{sti} = 1$ and $\omega_t = 0.55 + 0.2(0.25 - sim)/0.25$ otherwise. The action label follows A_0 except in the third case. There a cold seeker with a stated intent is labelled Prioritise_STI, not Ask. These labels are noisy by construction. ω_n down-weights them in Eq. (30) and defines the “necessary clarification” event in Eq. (37).

5.2.2 Synthetic Controlled Labels

Synthetic controlled labels. For 15% of natural test instances whose seekers have usable LTP profiles, an English seeker utterance is appended and the intended relationship is specified. For Conflict, Override, and Consistent, the utterance requests a genre that conflicts

with, explicitly overrides, or agrees with the seeker’s dominant historical genre. Three levels of request intensity are used, ranging from tentative expressions to explicit requests. The target is replaced with a movie belonging to the requested genre. For Complement, the requested genre co-occurs with the historical genres in the catalogue and is absent from the tension list $\mathcal{T}$, thereby adding a genre dimension without contradiction. For Uncertain, the utterance is vague, such as “anything is fine, surprise me”; intent cues are removed, and the target remains unchanged. Each synthetic instance records its source instance, injected utterance, intended relationship, and intensity, together with a synthetic-data flag. Synthetic instances are also included in training so that the relationship classifier receives clean examples of the less frequent classes. Because the injected utterances use a limited set of explicit genre expressions, they are easier to classify than natural requests involving indirect language or negation. Synthetic results therefore assess whether the policy follows its intended rules under controlled conditions, rather than demonstrating recommendation quality on natural dialogue. The two sources are never pooled when reporting results. Table 1 presents the relationship-label distributions. Override is rare in the natural data, consistent with the limited frequency of explicit statements indicating a departure from established preferences. Although the pipeline supports human-verified labels, none were available for this experiment; consequently, results requiring such labels are reported as “not run.”

Table 1. Relationship label counts by source and split

Source	Relationship	Train	Test
weak_rule	Complement	5530	744
weak_rule	Conflict	3211	292
weak_rule	Consistent	16193	2070
weak_rule	Override	237	39
weak_rule	Uncertain	7178	1327
synthetic_controlled	Complement	996	134
synthetic_controlled	Conflict	742	112
synthetic_controlled	Consistent	1038	123
synthetic_controlled	Override	776	114
synthetic_controlled	Uncertain	982	141

6. Experimental Setup

6.1 Compared Systems

LTP-only and STI-only rank items using their respective encoders. Naive fusion combines the two encodings using equal weights, $\alpha = 0.5$. Adaptive fusion learns a scalar gate from both encodings but does not include relationship classification or action selection.

The Sequential (GRU) baseline applies a GRU [12] to historical items followed by current-dialogue mentions. SASRec [13] uses two causal self-attention blocks over the same item sequence and scores items from the final valid position. It does not process dialogue text and serves as the strongest history-only comparison.

The Conversation-aware baseline attends to the current dialogue using the historical preference vector as its query. The KBRD-style baseline follows the entity-centric design of KBRD [5]. It embeds and pools items and genre entities mentioned in the dialogue using self-attention, combines them with the dialogue-text encoding through a learned gate, and

incorporates a genre-aware bias into item scoring. The original external knowledge graph and response generator are replaced by the shared item tower and MovieLens genres.

All four baselines are in-house reimplementations using the same parameter budget and item tower. They are not reproductions of the original published systems. Their results are therefore compared only within this controlled experimental setting and are not compared directly with published benchmark results. All models share the same data, text encoder, item tower, loss functions, random seeds, and parameter budget.

AIPA ablations remove the relationship classifier, counterfactual features, clarification mechanism, or persistence tracker, or replace the learned policy with the rule-based policy.

6.2 Metrics and Statistics

6.2.1 Ranking

For an instance with target y and fused scores $\mathbf{s}$, the rank of the target is the number of items strictly ahead of it plus one. The three ranking metrics depend on that rank alone,

$$r = 1 + |\{i \in \mathcal{I} : s_i > s_y\}|, \quad (32)$$

$$\begin{aligned} \text{Hit}@K &= \mathbb{1}[r \leq K], & \text{NDCG}@K &= \frac{\mathbb{1}[r \leq K]}{\log_2(r+1)}, \\ \text{MRR}@K &= \frac{\mathbb{1}[r \leq K]}{r}, \end{aligned} \quad (33)$$

for $K \in \{10, 20\}$. With one target per instance, $\text{Hit}@K$ equals $\text{Recall}@K$. Reported values are means over test instances and over the 5 seeds.

6.2.2 Conflict Subsets

Two natural conflict subsets are reported. The *strict* subset holds the natural test instances whose weak-rule label is Conflict or Override ($n=331$). The *broad* subset adds instances whose weak label has confidence $\omega_n \geq 0.6$ and whose genre vectors disagree, $JS(\pi_L, \pi_S) \geq 0.5$ (Jensen–Shannon divergence, base 2), giving $n=1272$. Synthetic Conflict/Override injections form a third subset ($n=226$). It is never pooled with the natural ones.

6.2.3 Relationship Classification

With $\hat{r}_n = \arg \max_r p_{n,r}$ the reported quantities are accuracy, per-class F1 and

$$\text{macro-F1} = \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} F1_r, \quad \text{weighted-F1} = \sum_{r \in \mathcal{R}} \frac{N_r}{N} F1_r, \quad (34)$$

where $F1_r$ is zero for a class that is never predicted and never occurs. Calibration uses the confidence $\kappa_n = \max_r p_{n,r}$ and ten equal-width bins $B_b = \{n : \kappa_n \in ((b-1)/10, b/10]\}$,

$$\text{ECE} = \sum_{b=1}^{10} \frac{|B_b|}{N} \left| \frac{1}{|B_b|} \sum_{n \in B_b} \mathbb{1}[\hat{r}_n = r_n] - \frac{1}{|B_b|} \sum_{n \in B_b} \kappa_n \right|, \quad (35)$$

$$Brier = \frac{1}{|\mathcal{R}^+|} \sum_{r \in \mathcal{R}^+} \frac{1}{N} \sum_{n=1}^N (p_{n,r} - \mathbb{1}[r_n = r])^2, \quad (36)$$

Empty bins are skipped. Here $\mathcal{R}^+$ collects the classes that actually appear in the subset being scored. The Brier score is one-versus-rest, averaged over those classes [20].

6.2.4 Arbitration and Clarification

Let $\hat{a}_n$ be the emitted action, a_n the reference action, and define the per-instance events

$$\begin{aligned} asked_n &= \mathbb{1}[\hat{a}_n = \text{Ask_Clarification}], \\ conf_n &= \mathbb{1}[r_n \in \{\text{Conflict}, \text{Override}\}], \\ nec_n &= \mathbb{1}[r_n = \text{Uncertain}] \vee (conf_n \wedge \omega_n < \theta_c), \end{aligned} \quad (37)$$

with $\theta_c = 0.6$. A question counts as necessary when the reference relationship is Uncertain, or when it is a conflict the labelling rule itself was unsure about. Then

$$\begin{aligned} \text{clar. rate} &= \frac{1}{N} \sum_n asked_n, \\ \text{clar. precision} &= \frac{\sum_n asked_n nec_n}{\sum_n asked_n}, \\ \text{clar. efficiency} &= \frac{\sum_n asked_n nec_n}{\sum_n nec_n}, \\ \text{unnecessary rate} &= \frac{1}{N} \sum_n asked_n (1 - nec_n), \end{aligned} \quad (38)$$

The wrong-override rate is $\frac{1}{N} \sum_n \mathbb{1}[\hat{a}_n = \text{Prioritise_STI} \wedge r_n = \text{Consistent}]$. Arbitration accuracy is $\frac{1}{N} \sum_n \mathbb{1}[\hat{a}_n = a_n]$. Conflict-resolution accuracy is the same quantity restricted to $conf_n = 1$. Override success is the mean Hit@10 over instances with $r_n = \text{Override}$ and $\hat{a}_n = \text{Prioritise_STI}$. Precision is undefined when no question is asked.

6.2.5 Uncertainty and Paired Comparisons

Each metric x_n is first averaged over seeds for every test instance n . A 95% interval for a mean is the percentile bootstrap over instances with $B = 1000$ resamples,

$$\bar{x}^{(b)} = \frac{1}{N} \sum_{n=1}^N x_{\xi_n^{(b)}}, \quad \xi_n^{(b)} \stackrel{\text{iid}}{\sim} \text{Unif}\{1, \dots, N\}, \quad (39)$$

The interval is $[\bar{x}_{2.5\%}^{(\cdot)}, \bar{x}_{97.5\%}^{(\cdot)}]$, the empirical percentiles of the resampled means, with a fixed resampling seed. AIPA (full) is compared with every other system on the same instances through the paired differences $d_n = x_n^{\text{AIPA}} - x_n^{\text{other}}$. The tests are the paired t -test and the Wilcoxon signed-rank test (zero differences split between the signs). The m p -values of one family (one metric, one subset) are adjusted by the Holm step-down procedure [21]. With $p_{(1)} \leq \dots \leq p_{(m)}$,

$$\tilde{p}_{(j)} = \max_{l \leq j} \min(1, (m-l+1)p_{(l)}). \quad (40)$$

Effect sizes are Cohen's d on the paired differences and Cliff's δ [22] over all instance pairs,

$$d = \frac{\bar{d}}{s_d}, \quad \delta = \frac{1}{N^2} \sum_{i,j=1}^N (\mathbb{1}[x_i > x'_j] - \mathbb{1}[x_i < x'_j]), \quad (41)$$

where s_d is the sample standard deviation of the d_n , $x = x^{AIPA}$ and $x' = x^{other}$. For $N > 2000$ the sum over j runs over the first 2000 control instances. A comparison whose differences are all zero (e.g. an ablation that changed no prediction) is reported as undefined, not as $p = 1$. Natural and synthetic results are always reported separately.

6.3 Compute

Everything runs on CPU with 8 cores, PyTorch 2.14.0+cpu and Python 3.12.8. The “full” configuration uses the whole corpus (100%), 5 seeds, 1000 bootstrap resamples, at most 20 epochs with early stopping on validation Hit@10, hidden size 64 and the MiniLM text encoder. Every hyper-parameter, including $\lambda_{rel} = 1.0$, $\lambda_{act} = 0.6$, the conflict weight 2.0 and label smoothing 0.1, was chosen on the validation split of a smaller “quick” configuration before this run. The test split was scored once. The full run takes about 72 minutes. “The earlier reduced-budget run” means that development configuration (25% of the corpus, two seeds, TF-IDF/SVD text). It is quoted only to show what changed with scale and encoder.

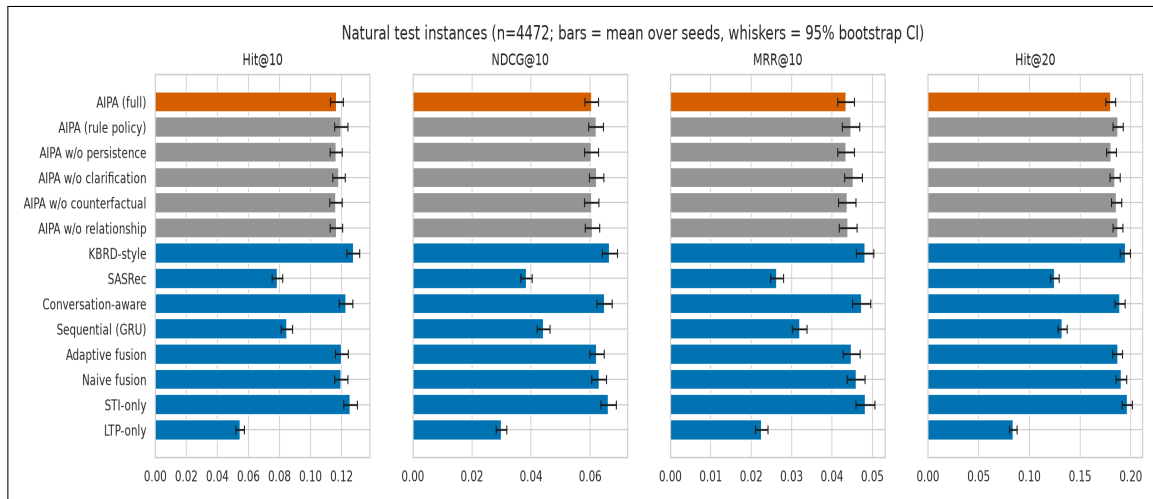
7. Results

7.1 Overall Ranking Quality

Table 2 and Fig. 3 report the natural test set. The set has $n=4472$ turns. Values are means over 5 seeds with 95% bootstrap intervals in brackets. Rows above the rule in the table are baselines. Two things stand out. First, the intent signal dominates on ReDial. STI-only (0.125) more than doubles LTP-only (0.055). The two history-only sequential baselines (GRU 0.085, SASRec 0.079) add little over the profile. Every model that reads the dialogue lands between 0.116 and 0.128 Hit@10. Second, AIPA (full) at 0.117 sits at the low end of that band. It is far better than LTP-only ($\Delta = +0.062$, Holm-corrected $p < 0.001$, $d = 0.19$), SASRec ($+0.038$, $p < 0.001$) and the GRU ($+0.032$, $p < 0.001$). It is indistinguishable from naive fusion, adaptive fusion and its own rule-policy variant (all within 0.003, $p > 0.6$). It is slightly but significantly *worse* than the conversation-aware baseline (-0.006 , $p = 0.026$), STI-only ($\Delta = -0.009$, $p < 0.001$, $d = -0.03$) and the KBRD-style baseline (-0.011 , $p < 0.001$, $d = -0.04$, Table 3). In that table Δ is the mean per-instance difference, positive in favour of AIPA, p_{Holm} the Holm-corrected paired t -test p -value and d Cohen's d . Dashes mark comparisons with zero variance. The effect sizes are tiny. With 4,472 test turns and five seeds they are real. The pre-registered criterion H1 (beat the best baseline on natural Hit@10) is not met. The best system on this corpus is the KBRD-style model, with STI-only a close second.

Table 2. Natural test set

Model	Hit@10 [95% CI]	NDCG@10	MRR@10	Hit@20	NDCG@20
LTP-only	0.055 [0.052, 0.057]	0.030	0.023	0.084	0.037
STI-only	0.125 [0.121, 0.130]	0.066	0.048	0.196	0.084
Naive fusion	0.119 [0.115, 0.124]	0.063	0.046	0.191	0.081
Adaptive fusion	0.120 [0.116, 0.124]	0.062	0.045	0.187	0.079
Sequential (GRU)	0.085 [0.081, 0.088]	0.044	0.032	0.132	0.056
Conversation-aware	0.123 [0.118, 0.127]	0.065	0.047	0.189	0.081
SASRec	0.079 [0.075, 0.082]	0.038	0.026	0.125	0.050
KBRD-style	0.128 [0.123, 0.132]	0.067	0.048	0.195	0.083
AIPA w/o relationship	0.117 [0.113, 0.121]	0.061	0.044	0.187	0.078
AIPA w/o counterfactual	0.116 [0.112, 0.120]	0.060	0.044	0.186	0.078
AIPA w/o clarification	0.118 [0.114, 0.122]	0.062	0.045	0.184	0.079
AIPA w/o persistence	0.116 [0.113, 0.120]	0.060	0.044	0.181	0.076
AIPA (rule policy)	0.120 [0.115, 0.124]	0.062	0.045	0.187	0.079
AIPA (full)	0.117 [0.113, 0.121]	0.060	0.043	0.180	0.076

**Figure 3.** Hit@10 and NDCG@10, natural test set**Table 3.** Paired comparisons against AIPA (full)

Control	Natural (all, n=22360)			Natural conflict (n=1655)			Synthetic conflict (n=1130)		
	Δ	p_{Holm}	d	Δ	p_{Holm}	d	Δ	p_{Holm}	d
LTP-only	+0.062	<0.001 *	+0.19	+0.115	<0.001 *	+0.31	+0.219	<0.001 *	+0.51
STI-only	-0.009	<0.001 *	-0.03	-0.019	0.161	-0.06	+0.009	1.000	+0.02
Naive fusion	-0.003	0.771	-0.01	+0.002	1.000	+0.01	+0.010	1.000	+0.03
Adaptive fusion	-0.003	0.644	-0.01	-0.012	0.998	-0.04	-0.018	1.000	-0.04
Sequential (GRU)	+0.032	<0.001 *	+0.10	+0.065	<0.001 *	+0.18	+0.230	<0.001 *	+0.53

Conversation-aware	-0.006	0.026 *	-0.02	-0.012	1.000	-0.03	+0.019	1.000	+0.05
SASRec	+0.038	<0.001 *	+0.11	+0.079	<0.001 *	+0.22	+0.240	<0.001 *	+0.55
KBRD-style	-0.011	<0.001 *	-0.04	-0.022	0.161	-0.06	-0.012	1.000	-0.03
AIPA w/o relationship	+0.000	1.000	+0.00	-0.003	1.000	-0.01	+0.004	1.000	+0.01
AIPA w/o counterfactual	+0.001	1.000	+0.00	-0.003	1.000	-0.01	+0.004	1.000	+0.01
AIPA w/o clarification	-0.001	1.000	-0.01	+0.003	1.000	+0.01	+0.004	1.000	+0.01
AIPA w/o persistence	+0.000	1.000	+0.00	+0.005	0.232	+0.05	–	–	–
AIPA (rule policy)	-0.003	0.771	-0.01	+0.006	1.000	+0.02	+0.014	1.000	+0.04

7.2 Conflict Turns

The motivating hypothesis concerns turns in which LTP and STI disagree. We evaluate the strict natural conflict subset ($n = 331$, based on weak-rule labels), the broad conflict subset ($n = 1272$), and the synthetic conflict subset ($n = 226$). Figure 4 compares conflict and non-conflict turns. Natural conflicts are identified using weak-rule labels. The strict subset comprises Conflict/Override cases, while the broad subset additionally includes confident genre disagreements. Synthetic conflicts are generated through controlled injections. The confidence intervals for the conflict subsets are wide because of the smaller sample sizes. For strict natural conflicts, the model ranking is consistent with that of the full dataset: KBRD-style (0.172), STI-only (0.169), adaptive fusion and conversation-aware (0.162), AIPA (full) (0.150), naive fusion (0.148), and LTP-only (0.035). With 331 turns and five random seeds, the confidence intervals overlap. AIPA's performance differences relative to the two leading models (-0.019 to -0.022) are not statistically significant after multiple-comparison correction ($p_{\text{Holm}} = 0.16$). Therefore, the strict subset provides no evidence of a significant difference among the intent-reading models, although AIPA has a lower point estimate. The broad subset ($n = 1272$) provides clearer evidence of performance differences. STI-only (0.164) and KBRD-style (0.163) significantly outperform AIPA (full) (0.148), with $p_{\text{Holm}} < 0.02$. Adaptive fusion (0.154) and naive fusion (0.149) do not differ significantly from AIPA.

On synthetic conflicts, all models that rely on STI perform well, achieving scores between 0.23 and 0.27, whereas LTP-only, GRU, and SASRec fall below 0.03. AIPA (full) achieves 0.247, between adaptive fusion (0.265), the best-performing model, and STI-only (0.238), with no statistically significant difference from either. Thus, the synthetic subset confirms that AIPA responds to explicitly defined departures. However, it also demonstrates that consistently following the intent signal is an equally effective strategy on this dataset. Consequently, criteria H2 (strict and broad conflict subsets) and H3 are not satisfied.

Figure 5 presents the natural dataset results by relationship class. The Uncertain class (1,327 instances, 30% of which are cold seekers) is where the KBRD-style and conversation-aware baselines achieve their largest advantages, with scores of 0.104 and 0.103, respectively, compared with 0.096 for both AIPA (full) and STI-only. For Consistent turns, which constitute the majority, all fusion models achieve comparable scores, with differences

of only a few thousandths. When results are grouped by history length (Fig. 8), AIPA (full) performs competitively on cold seekers, scoring 0.133 compared with 0.129 for STI-only and 0.138 for KBRD-style. For histories containing 10–24 items, AIPA (full) achieves 0.124, compared with 0.128 and 0.131 for STI-only and KBRD-style, respectively. However, on the 3,273 long-history turns, AIPA (full) underperforms, scoring 0.116 compared with 0.126 for STI-only and 0.128 for KBRD-style. These results suggest that excessive reliance on profile information may introduce more errors than benefits for long interaction histories.

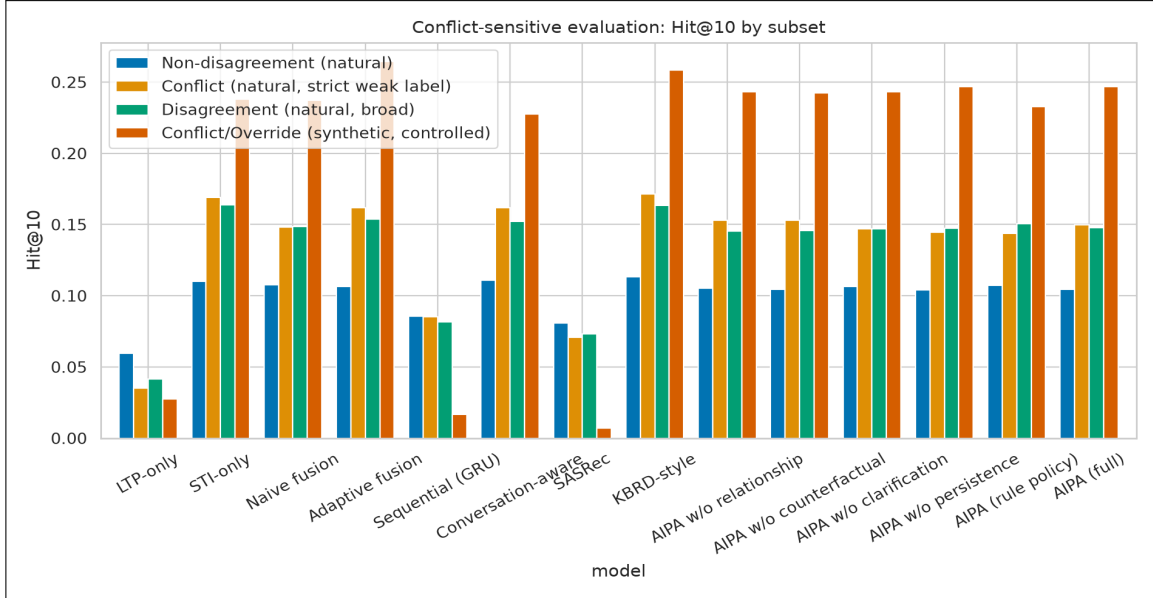


Figure 4. Conflict versus non-conflict turns



Figure 5. Hit@10 per relationship class

7.3 Relationship Classification and Arbitration

Table 4 reports the relationship head. Scores are means over seeds. W-F1 denotes weighted F1. The last five columns report per-class F1, and a dash indicates that the class is absent from the subset. On natural data, accuracy is 0.80 and macro-F1 is 0.62–0.66. AIPA

(full) achieves 0.627, meeting the pre-set criterion of 0.5 and improving from 0.54 in the reduced-budget run. The head performs well on Consistent and Uncertain (F1-scores of 0.85 and 0.95). It performs moderately on Complement (0.55) and Conflict (0.41, up from below 0.2 but still below the 0.5 criterion). Its weakest performance is on Override (0.37), which has only 39 natural test instances. Two caveats apply. The reference labels are weak, so some errors may reflect label noise. The two classes with the lowest performance are also the two defined by a similarity threshold. On synthetic data, the head achieves near-perfect performance (macro-F1 0.96, Conflict 0.93, Override 0.99, vague Uncertain utterances 1.00, and Complement requests 0.93). It recognises the patterns when the text states them explicitly. The confusion matrix in Figure 6 shows that most natural conflicts missed by the model are classified as Consistent. The rows represent the weak-rule reference labels, and the columns represent the predictions.

Table 4. Relationship classification.

Model	Subset	Acc.	Macro-F1	W-F1	Compl.	Consist.	Confl.	Overr.	Uncert.
AIPA w/o relationship	natural	0.794	0.621	0.791	0.521	0.845	0.422	0.363	0.953
AIPA w/o relationship	synthetic	0.971	0.970	0.970	0.941	0.970	0.943	0.996	1.000
AIPA w/o counterfactual	natural	0.795	0.621	0.792	0.534	0.847	0.406	0.368	0.951
AIPA w/o counterfactual	synthetic	0.973	0.973	0.973	0.947	0.970	0.951	0.997	1.000
AIPA w/o clarification	natural	0.801	0.617	0.796	0.549	0.847	0.400	0.334	0.956
AIPA w/o clarification	synthetic	0.965	0.964	0.965	0.934	0.968	0.929	0.992	1.000
AIPA w/o persistence	natural	0.807	0.656	0.804	0.556	0.854	0.444	0.471	0.954
AIPA w/o persistence	synthetic	0.963	0.962	0.962	0.927	0.961	0.931	0.988	1.000
AIPA (rule policy)	natural	0.802	0.636	0.799	0.543	0.852	0.427	0.406	0.953
AIPA (rule policy)	synthetic	0.967	0.966	0.967	0.936	0.968	0.930	0.996	1.000
AIPA (full)	natural	0.800	0.627	0.797	0.553	0.848	0.412	0.369	0.954
AIPA (full)	synthetic	0.963	0.962	0.962	0.927	0.961	0.931	0.988	1.000

The arbitration head (Table 5) performs as intended. The table reports means over seeds, and clarification precision is undefined when no question is asked. The figure shows the predicted actions within each reference class on natural turns. The learned policy asks a clarifying question on approximately 20% of natural turns. When it asks, the reference label is Uncertain or a low-confidence conflict in 99% of cases. Unnecessary clarifications occur in 0.2% of cases, and wrong overrides occur in 1%, meeting both arbitration criteria. These rates measure agreement with the weak-rule reference. They indicate that the policy has learned the intended rule, but do not establish whether users welcome a question or are satisfied with an override. Only a user study can establish this. The rule policy asks questions more frequently (32%), with lower precision (0.64) and an unnecessary-clarification rate of 12%. Thus, the learned policy does more than reproduce the initial rule table. Conflict-resolution accuracy on

natural conflicts is low (0.35 for AIPA (full) and 0.29 for the rule) for the reasons discussed above. On synthetic conflicts, it reaches 0.94. Calibration of the relationship head is good (ECE 0.03–0.06 on natural data). The gap between synthetic macro-F1 (approximately 0.96) and natural macro-F1 (0.63) is substantial. This indicates that the head is reliable on clean, explicit requests but less reliable on natural turns, where cues are indirect and the reference labels are noisy.

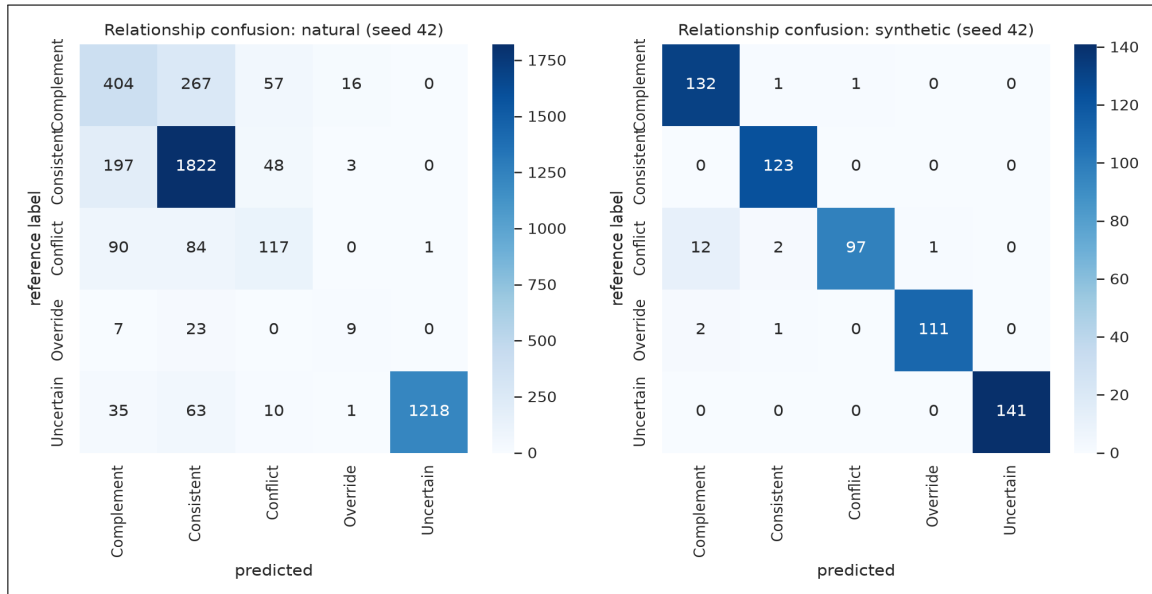


Figure 6. Relationship confusion matrix

Table 5. Arbitration and clarification

Model	Subset	Arb. acc.	Confl. res.	Overr. succ.	Clar. rate	Clar. prec.	Unnec. clar.	Wrong overr.
AIPA w/o relationship	natural	0.894	0.418	0.200	0.196	0.989	0.002	0.017
AIPA w/o relationship	synthetic	0.980	0.958	0.205	0.226	1.000	0.000	0.000
AIPA w/o counterfactual	natural	0.892	0.414	0.218	0.199	0.979	0.004	0.014
AIPA w/o counterfactual	synthetic	0.984	0.963	0.209	0.226	1.000	0.000	0.000
AIPA w/o clarification	natural	0.709	0.351	0.306	0.000	–	0.000	0.010
AIPA w/o clarification	synthetic	0.750	0.937	0.203	0.000	–	0.000	0.000
AIPA w/o persistence	natural	0.905	0.406	0.196	0.196	0.991	0.002	0.011
AIPA w/o persistence	synthetic	0.974	0.938	0.208	0.226	1.000	0.000	0.000
AIPA (rule policy)	natural	0.811	0.286	0.393	0.323	0.642	0.116	0.009
AIPA (rule policy)	synthetic	0.952	0.900	0.214	0.257	0.880	0.031	0.000
AIPA (full)	natural	0.902	0.353	0.263	0.196	0.991	0.002	0.010
AIPA (full)	synthetic	0.974	0.938	0.208	0.226	1.000	0.000	0.000

7.4 Counterfactual Driver Diagnostic

Table 6 and Figure 7 summarise the intervention probe for AIPA (full). The table reports the mean absolute change in NDCG@10, the top-10 overlap when one pathway is removed, and the proportion of turns under each driver label. Removing the STI pathway disrupts the top-10 list more than removing LTP. The overlap with the original list is 0.07–0.22 versus 0.62–0.72 on natural turns. The per-turn driver label is STI-driven for 57–72% of natural turns and Jointly-driven for the remainder. LTP-driven turns account for less than 1%, and no turn is labelled Neither-driven. On synthetic conflicts and overrides, dependence on STI is strongest (overlap without STI of 0.03–0.04, with 81–84% of turns labelled STI-driven). On synthetic Uncertain turns, where the intent cue has been erased, the label shifts to Jointly-driven (81%). This is the intended behaviour. Agreement between the driver label and the selected action is 0.37 on natural turns and 0.43 on synthetic turns. These values are lower than those in the reduced-budget run because the model now fuses more frequently while the ranking remains intent-led. The probe confirms the findings of the ranking tables. The trained model relies primarily on the conversation and uses the profile as a secondary signal. The probe intervenes on the fitted model by measuring how the ranking changes when a pathway is removed. It does not provide a causal estimate of how users would respond, and the driver labels should be interpreted accordingly.

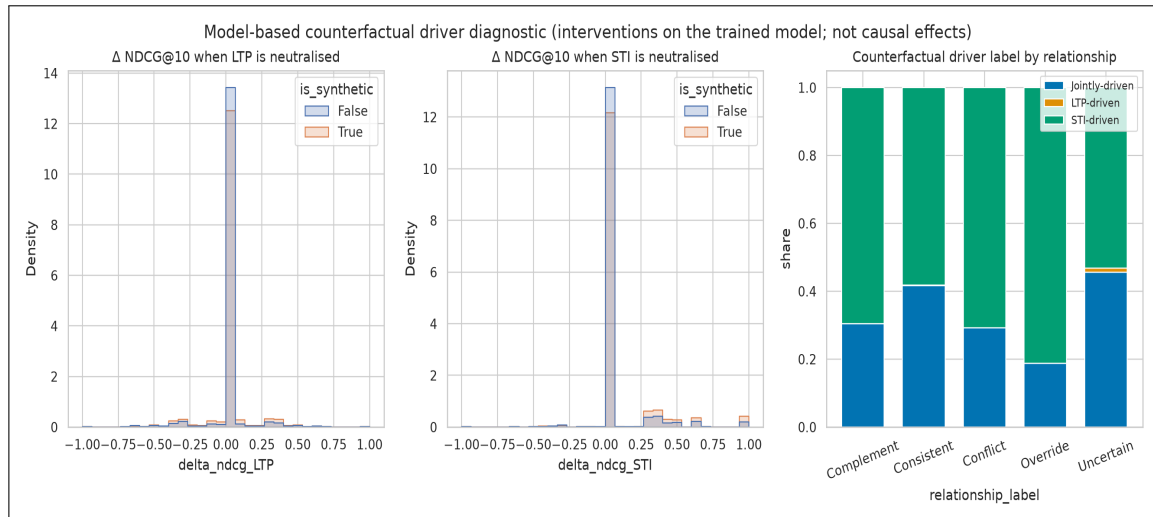


Figure 7. Driver labels and NDCG changes under signal removal

Table 6. Counterfactual probe by relationship class.

Source	Relationship	n	Δ NDCG@10		Top-10 overlap		Driver (%)			
			no LTP	no STI	no LTP	no STI	STI	LTP	Joint	Neither
natural	Complement	744	0.047	0.074	0.635	0.102	67	0	32	0
natural	Conflict	292	0.045	0.072	0.649	0.072	67	0	33	0
natural	Consistent	2070	0.035	0.055	0.619	0.176	58	0	41	0
natural	Override	39	0.048	0.087	0.692	0.092	72	1	28	0
natural	Uncertain	1327	0.018	0.049	0.715	0.216	57	1	42	0
synthetic	Complement	134	0.055	0.140	0.685	0.084	81	0	19	0
synthetic	Conflict	112	0.071	0.145	0.727	0.030	81	0	19	0

synthetic	Consistent	123	0.036	0.070	0.598	0.211	53	0	47	0
synthetic	Override	114	0.048	0.099	0.741	0.042	84	0	16	0
synthetic	Uncertain	141	0.028	0.029	0.520	0.380	15	4	81	0

7.5 Persistent Shifts

With $k = 2$, the tracker identified 32 persistent shifts per seed for AIPA (full) across approximately 20 seekers. Seeker 1087 shifted toward war films and horror, while seekers 1034 and 1035 shifted toward horror and romance. A shift applies only to instances after the second qualifying session. It affected 2,321 test instances from 30 seekers. For these instances, Hit@10 increased from 0.1166 without the tracker to 0.1172 with it (+0.0006, permutation $p = 1$). Across the entire natural dataset, the difference was +0.0003. Restricting the analysis to the 51 seekers with at least three sessions did not change the result. A sweep over $k \in \{1, 2, 3\}$ on the validation split produced 168, 7, and 0.2 shifts per seed, respectively, with no change in validation Hit@10 at any value of k . On the test split, $k = 1$ triggered 183 times and produced marginally worse results. The mechanism operates as designed and now triggers frequently enough for evaluation. However, the evaluation shows no improvement in ranking performance. The criterion requiring the tracker to improve performance on the instances it affects is not met.

7.6 Sensitivity and Fusion Weight

Figure 8 plots Hit@10 against history length, number of seeker turns and injected conflict intensity. Cold seekers are the easiest bucket (0.133). Short histories are the hardest (0.089 for 3–9 items). The mid and long buckets sit between (0.124 and 0.116). Every model shows the same shape, so it reflects the data, not the arbitration. On synthetic conflicts AIPA (full) reaches 0.23 at the mildest intensity (STI-only 0.24), 0.31 at the middle one (0.30) and 0.20 at the strongest (0.17). With 68–85 instances per cell none of these gaps means much. Table 7 shows the fixed-weight sweep for naive fusion. Anything between $\alpha_{LTP} = 0$ and 0.5 is close to the best (0.119–0.121). 0.75 already costs two points. Pure LTP weighting collapses to 0.031, which shows how little the profile alone can do here.

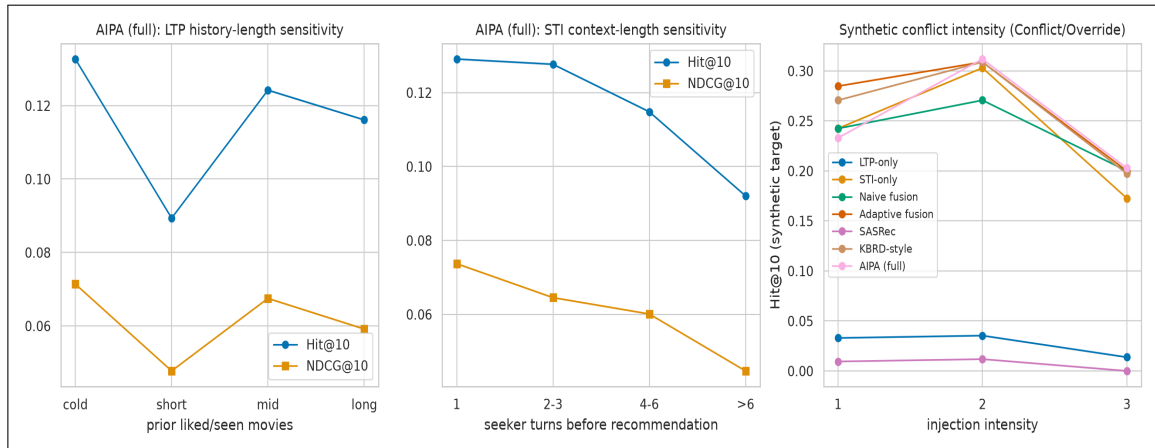


Figure 8. Hit@10 sensitivity

Table 7. Fixed fusion weight sweep

α_{LTP}	Hit@10	NDCG@10	Hit@20	Hit@10 (synthetic)
0.00	0.119	0.062	0.186	0.172
0.25	0.121	0.064	0.191	0.177
0.50	0.119	0.063	0.191	0.176
0.75	0.100	0.052	0.158	0.152
1.00	0.031	0.016	0.049	0.024

7.7 Order-Permutation Control

Cross-session histories rely on the assumption that conversation identifiers follow chronological order. ReDial does not provide timestamps, so this assumption cannot be verified directly. A control experiment evaluates the effect of this assumption on the results. The same split, target instances, and models are retained, with only the order of each seeker's earlier sessions changed. For each target dialogue, eligible sessions are those with lower identifiers, as in the primary run. A seeded shuffle then reorders these sessions before truncation and decay are applied. The session set, count, and content remain unchanged under both orders. No later session or target session can enter the history. LTP-only, STI-only, KBRD-style, and AIPA (full) were retrained for all five seeds using this ordering. Table 8 compares the two runs on the natural test set. No model exhibits a meaningful change. Table 8 reports means over five seeds. "Chron." denotes the primary run using conversation identifier order, while "Perm." denotes the reordering of the same earlier sessions using a fixed seed before truncation and decay. Δ represents the permuted score minus the chronological score, with a 95% bootstrap interval and a Holm-corrected paired permutation p-value. LTP-only changes by -0.000 in Hit@10 and AIPA (full) by +0.000, both with $p_{Holm} = 1.00$. STI-only improves by 0.003 ($p_{Holm} = 0.22$), and KBRD-style improves by 0.004 ($p_{Holm} = 0.03$). The KBRD-style change is small, approximately 3% of its score, and contradicts the expected direction of a recency effect.

Thus, the within-history ordering imposed by the identifiers provides little usable signal at this history length. Applying recency truncation and decay to the same sessions in either order produces similar outcomes. The ranking of the leading models remains unchanged, with AIPA performing below STI-only and KBRD-style under both orders. The control therefore supports the main findings within a limited scope. It shows that the results do not depend on which earlier session is treated as the most recent. However, it cannot establish whether a lower identifier corresponds to an earlier session and therefore cannot validate the leak-free split.

Table 8. Order-permutation control

Model	Hit@10					NDCG@10	
	chron.	perm.	Δ	95% CI	p_{Holm}	chron.	perm.
LTP-only	0.055	0.054	-0.000	[-0.003, +0.002]	1.00	0.030	0.029
STI-only	0.125	0.129	+0.003	[+0.000, +0.007]	0.22	0.066	0.067
KBRD-style	0.128	0.132	+0.004	[+0.001, +0.008]	0.03	0.067	0.069
AIPA (full)	0.117	0.117	+0.000	[-0.003, +0.004]	1.00	0.060	0.061

7.8 Ablations and Efficiency

Considering the AIPA results in Tables 2 and 3 as an ablation study, no component removal changes natural Hit@10 by more than 0.003, and none of the differences is statistically

significant ($p_{\text{Holm}} \geq 0.77$). Removing the relationship head, counterfactual features, or persistence tracker leaves the score unchanged to three decimal places. Removing clarification yields the highest score among the variants (0.118), while the rule policy achieves the highest score overall (0.120). Neither difference is statistically significant. On the strict conflict subset, the results follow the expected direction: removing persistence or using the rule policy reduces performance by 0.005 – 0.006, but these effects remain within the noise. All variants have 0.51 – 0.63M parameters and train in under a minute on a CPU, except SASRec. The inference time per turn is below 0.1 ms. Although the absolute computational cost is small, the relative cost is higher. AIPA (full) requires 53 s for training and 0.071 ms to score a sample. The KBRD-style model requires 14 s and 0.013 ms, while the conversation-aware model requires 12 s and 0.010 ms. Thus, AIPA is approximately four to five times more computationally expensive than the two strongest baselines, with a 15-20% increase in parameter count. The additional cost arises from the two encoders, relationship and action heads, and counterfactual pass, which scores each batch three times. Parameter count and model size do not establish deployment memory requirements. Therefore, Table 9 reports peak process memory, batched throughput, end-to-end single-turn latency, and catalogue scaling for the strongest models. Profiling was performed on an 8-core CPU using seed 42. Peak RSS represents the entire process after training. Forward-pass cost per turn was measured with the catalogue tiled $1\times$, $4\times$ and $16\times$. End-to-end latency includes MiniLM encoding of the three text fields, feature construction, and the forward pass for a single turn with a cold embedding cache. For AIPA (full), the median end-to-end latency is 13.6 ms, and the 95th percentile is 25.5 ms. Sentence encoding accounts for nearly all of this latency, which is shared by all text-reading models. Peak resident memory after training is 2,378MB for the entire Python process, including the model and data. Batched scoring achieves a throughput of 23,117 turns per second on the 8-core CPU. Scoring cost increases approximately linearly with catalogue size, from the 6,924 ReDial items to 16 times that size, as expected for a full dot product over the item tower. Several aspects remain unmeasured. No GPU, mobile, or edge-device testing was performed. Energy consumption, latency under concurrent users, candidate retrieval for very large catalogues, clarification text generation, and the cost of persistence updates across sessions were not evaluated. The supported claim is lightweight CPU inference in an offline setting, not production readiness.

Table 9. Deployment profile

	Peak RSS	Throughput	End-to-end turn (ms)		Forward ms/turn vs. catalogue		
Model	(MB)	(turns/s)	median	p95	$1\times$	$4\times$	$16\times$
STI-only	1678	33,580	13.3	24.9	0.012	0.036	0.305
Conversation-aware	2378	109,485	13.7	23.7	0.004	0.011	0.098
KBRD-style	2378	83,135	13.0	24.3	0.006	0.015	0.172
AIPA (full)	2378	23,117	13.6	25.5	0.026	0.091	0.546

7.9 Qualitative Cases and Error Analysis

Representative test turns illustrate the aggregate pattern. Reference labels are weak-rule labels, and rank denotes the rank of the true target under AIPA (full). When a seeker explicitly states a genre (“inspirational war movies like PT 109”), the model identifies the departure from a comedy-heavy profile as a conflict, prioritises STI, and ranks the target seventh.

When the request contains no genre cue (“anything recent that you have seen?”), the model asks a clarifying question, and the target ranks much lower. The most instructive failure involves negation. “I like pretty much anything but horror!” produces a horror-dominated STI vector. The weak rule labels this as an override, and the model follows the incorrect intent, ranking the target 144th. The most common error is a weak-rule Complement or Conflict turn being predicted as Consistent and fused. In the natural dataset, 44% of Complement and 64% of Conflict references are misclassified, down from 75% and 81%, respectively, in the reduced-budget run. Their targets have median ranks of 165 and 144, respectively, compared with 186 for Consistent turns.

8. Discussion

8.1 Evidence Supported by the Results

Four findings are supported by this run. (i) Any model that reads the conversation significantly outperforms a profile-only model on ReDial. The profile-only approaches, whether pooled (LTP-only) or sequential (GRU, SASRec), achieve 5–8% Hit@10, compared with 12–13% for the intent-reading models. (ii) The learned policy behaves as intended. It asks when necessary, rarely asks when unnecessary, and rarely overrides in the wrong direction. With the full corpus, the relationship head is a usable classifier (macro-F1 0.63, Conflict-F1 0.41), rather than a majority-class predictor. (iii) On explicitly stated departures from habitual preferences in the synthetic dataset, AIPA follows the user's intent, and the counterfactual probe confirms that the ranking is driven by the stated intent. (iv) The framework has a low absolute computational cost. It runs on a CPU, trains in under a minute, and scores a turn in a fraction of a millisecond. However, it requires four to five times the training and scoring time of the two strongest baselines, while end-to-end latency is dominated by sentence encoding in all cases.

8.2 Findings Not Supported by the Results

The central claim that explicit arbitration outperforms simpler fusion on conflict turns is not supported. With the full corpus, this is no longer attributable solely to limited statistical power. On the strict natural conflict subset, AIPA performs two percentage points below the intent-only and KBRD-style models, with overlapping intervals. On the broad subset, the gap is 1.5 percentage points and is statistically significant. On synthetic conflicts, AIPA performs comparably to adaptive fusion and the straightforward intent-following policy. On the full natural dataset, the intent-only and KBRD-style models outperform AIPA by approximately one percentage point, with statistically significant differences. The ablation results are also unchanged: removing any individual arbitration component has no meaningful effect on performance. Thus, these components do not determine the model's ranking. Three explanations are likely, with the most important discussed first.

First, the labels. Weak-rule labels serve as the reference for the relationship head, arbitration metrics, and conflict subset. If the rules misclassify many turns, training is affected by label noise, and evaluation uses a subset that may not accurately represent the intended cases. The negation example in Section 7.9 illustrates this problem. The near-perfect synthetic results, compared with the moderate natural results, provide further evidence.

Second, the signal. The counterfactual probe indicates that the trained model is intent-driven on 60–70% of natural turns and profile-driven on fewer than 1%. When one

pathway already dominates, a layer that determines how much to trust the other has little to arbitrate. The flat ablation results are consistent with this finding. ReDial seekers are crowd workers who discuss a limited number of films per session. Their cross-session profiles are short, noisy, and weakly predictive, achieving only 5% Hit@10 independently. A corpus containing genuine long-term histories is needed to determine whether arbitration can provide benefits.

Third, the representation. MiniLM embeddings improved every model compared with the TF-IDF experiment (STI-only: 0.110 to 0.125; AIPA: 0.095 to 0.117) but did not change the ranking. The STI encoder still represents an utterance using genre proportions and a pooled sentence vector. Cues such as “for my brother” or “anything but horror” are not adequately captured. The Uncertain class accounts for 30% of natural turns, where the strongest baselines gain their advantage.

8.3 Three Factors may Explain These Findings

1. A corpus with genuine long-term histories. ReDial's cross-session profiles are derived from crowd workers who each completed a few tasks. A profile that achieves only 5% Hit@10 independently provides little useful signal for an arbitration layer. Review-grounded dialogue corpora or platform logs containing genuine return visits are suitable next testbeds.
2. Human-annotated relationship labels for several hundred test turns. The pipeline already accepts such a file and reports all metrics against it separately. This is the most valuable addition for determining whether weak-rule labels explain the Conflict-F1 score of 0.41.
3. An STI encoder that handles negation and pragmatics. “Anything but horror” must not be interpreted as a preference for horror. A fine-tuned classifier using sentence embeddings or an instruction-tuned language model prompted to identify genre preferences and polarity could address this failure.
4. A faithful knowledge-grounded baseline. A KBRD or KGSP implementation using its original knowledge graph is needed. The current KBRD-style reimplementations already outperform AIPA, and this finding should be tested against the original models.
5. User-based evaluation of clarification or, at minimum, simulated responses. Offline precision does not establish whether users welcome a question. An answered clarification is the point at which arbitration can acquire information unavailable to a single-shot baseline.

8.4 Threats to Validity

Conversation identifiers are assumed to follow chronological order. The order-permutation control in Section 7.7 bounds the effect of this assumption on the reported results but cannot recover the actual session order. ReDial seekers are crowd workers rather than organic users, so their cross-session preferences may partly reflect task instructions. The MovieLens join fails to match 13.3% of catalogue titles and approximately 6% of mentions, leaving these items without genre information. Peak memory and end-to-end latency were measured offline on a single machine and do not establish performance under interactive workloads. The sequential,

SASRec, conversation-aware, and KBRD-style baselines are small in-house reimplementations. These limitations are documented in the released code and generated experiment report.

9. Conclusion

This work aimed to make the reconciliation of long-term preferences and short-term intent explicit in a conversational recommender and to determine whether this approach improves recommendations when the two signals disagree. A reproducible pipeline was developed using an English corpus. It includes cross-session histories designed to prevent leakage and two separate label sources covering all five relationship classes. It incorporates a relationship classifier trained using conflict-weighted and self-trained weak labels, a learned arbitration policy with learned fusion weights and clarification, a persistence tracker, and a counterfactual probe. The framework was evaluated on the full corpus using five seeds and eight baselines, with confidence intervals, corrected paired tests, and predefined success criteria. The ranking results are negative. Although the arbitration components behave as intended and make the model's decisions interpretable, they do not improve recommendations over an intent-only model or a KBRD-style baseline on ReDial, either on conflict turns or across the full dataset. Removing these components also produces no meaningful change. The study therefore provides a tested framework, a measurement protocol for evaluating conflict cases, and a diagnosis: the profile signal in this corpus is too weak for arbitration to provide measurable benefits. The next step is not to introduce another model component, but to evaluate the framework using a corpus containing genuine long-term histories and human-annotated relationship labels. The code is ready for both.

References

- [1] K. Christakopoulou, F. Radlinski, and K. Hofmann, "Towards Conversational Recommender Systems," in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, 815–824.
- [2] Y. Sun and Y. Zhang, "Conversational Recommender System," in Proceedings of the 41st international ACM SIGIR conference on research and development in information retrieval, 2018, 235–244.
- [3] W. Lei et al., "Estimation-Action-Reflection: Towards Deep Interaction Between Conversational and Recommender Systems," in Proceedings of the 13th international conference on web search and data mining (WSDM), 2020, 304–312.
- [4] Y. Zhang, X. Chen, Q. Ai, L. Yang, and W. B. Croft, "Towards Conversational Search and Recommendation: System Ask, User Respond," in Proceedings of the 27th ACM international conference on information and knowledge management (CIKM), 2018, 177–186.
- [5] Q. Chen et al., "Towards Knowledge-Based Recommender Dialog System," in Proceedings of the 2019 conference on empirical methods in natural language processing (EMNLP-IJCNLP), 2019, 1803–1813.
- [6] K. Zhou, W. X. Zhao, S. Bian, Y. Zhou, J.-R. Wen, and J. Yu, "Improving Conversational Recommender Systems via Knowledge Graph Based Semantic Fusion," in Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery and data mining, 2020, 1006–1014.

- [7] R. Li, S. E. Kahou, H. Schulz, V. Michalski, L. Charlin, and C. Pal, “Towards Deep Conversational Recommendations,” in *Advances in neural information processing systems* (NeurIPS), 2018.
- [8] S. A. Hayati, D. Kang, Q. Zhu, W. Shi, and Z. Yu, “INSPIRED: Toward Sociable Recommendation Dialog Systems,” in *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, 2020, 8142–8152.
- [9] K. Zhou, Y. Zhou, W. X. Zhao, X. Wang, and J.-R. Wen, “Towards Topic-Guided Conversational Recommender System,” in *Proceedings of the 28th international conference on computational linguistics (COLING)*, 2020, 4128–4139.
- [10] D. Jannach, A. Manzoor, W. Cai, and L. Chen, “A Survey on Conversational Recommender Systems,” *ACM Computing Surveys*, vol. 54, no. 5, 2021, 1–36.
- [11] C. Gao, W. Lei, X. He, M. de Rijke, and T.-S. Chua, “Advances and Challenges in Conversational Recommender Systems: A Survey,” *AI Open*, vol. 2, 2021, 100–126.
- [12] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, “Session-based Recommendations with Recurrent Neural Networks,” in *International conference on learning representations (ICLR)*, 2016.
- [13] W.-C. Kang and J. McAuley, “Self-attentive Sequential Recommendation,” in *2018 IEEE international conference on data mining (ICDM)*, 2018, 197–206.
- [14] M. Quadrana, P. Cremonesi, and D. Jannach, “Sequence-Aware Recommender Systems,” *ACM Computing Surveys*, vol. 51, no. 4, 2018, 1–36.
- [15] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge University Press, 2009.
- [16] N. Reimers and I. Gurevych, “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)*, 2019, 3982–3992.
- [17] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, “BPR: Bayesian Personalized Ranking from Implicit Feedback,” in *Proceedings of the 25th conference on uncertainty in artificial intelligence (UAI)*, 2009, 452–461.
- [18] I. Loshchilov and F. Hutter, “Decoupled Weight Decay Regularization,” in *International conference on learning representations (ICLR)*, 2019.
- [19] F. M. Harper and J. A. Konstan, “The MovieLens datasets: History and context,” *ACM Transactions on Interactive Intelligent Systems*, vol. 5, no. 4, pp. 1–19, 2015.
- [20] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On Calibration of Modern Neural Networks,” in *Proceedings of the 34th international conference on machine learning (ICML)*, 2017, 1321–1330.
- [21] S. Holm, “A Simple Sequentially Rejective Multiple Test Procedure,” *Scandinavian Journal of Statistics*, vol. 6, no. 2, 1979, 65–70.
- [22] N. Cliff, “Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions,” *Psychological Bulletin*, vol. 114, no. 3, 1993, 494–509.

Appendix:

Supplementary Tables

The table below is referenced from the main text.

Table A. Notation

Symbol	Meaning
$\mathcal{I}, \mathcal{G}$	item catalogue and genre set ($ \mathcal{G} = 19$)
$G(i) \subseteq \mathcal{G}$	MovieLens genres of item i (may be empty)
u, c, t	seeker, dialogue (conversation id), turn index
$\mathcal{H}_u(c)$	dialogues of u strictly before c , Eq. (1)
$h = (h_1, \dots, h_L)$	liked/seen items from $\mathcal{H}_u(c)$, $L \leq L_{\max}$
$\mathcal{P}$	profile sentences of u from $\mathcal{H}_u(c)$
R_t, l_t	recent seeker text and last seeker utterance before t
$\mathcal{M}_{<t}, \mathcal{M}_{<t}^+$	movies mentioned / liked by the seeker before t
$\pi_L, \pi_S \in \Delta^{ \mathcal{G} }$	LTP and STI genre vectors, Eqs. (3), (4)
$f \in \mathbb{R}^8$	lexical and length flags, Eq. (5)
$\phi(\cdot) \in \mathbb{R}^{d_t}$	text encoder, Eq. (6)
e_i, b_i	item vector and item bias, Eq. (7)
$h_L, h_S \in \mathbb{R}^d$	LTP and STI encodings
$s_L, s_S, s \in \mathbb{R}^{ \mathcal{I} }$	component and fused score vectors
$\mathcal{R}, \mathcal{A}$	relationship classes, actions ($ \mathcal{R} = 5, \mathcal{A} = 4$)
$p \in \Delta^5, q \in \Delta^4$	relationship posterior, action distribution
$W_A \in \mathbb{R}^{4 \times 2}$	action-to-weight table (learned), Eq. (17)
$c \in \mathbb{R}^5$	counterfactual features, Eq. (23)
$\gamma, \theta, \tau, \rho, k, \eta$	decay, rule threshold, driver threshold, dominance ratio, persistence count, persistence gain