A Human-Centred Framework for Retrieval-Grounded Cold-Start Recommendation

Stefano Valtolina[1,*,†], Ricardo Anibal Matamoros[2,†] and Francesco Epifania[2,†]

[1]Università degli Studi di Milano, via Celoria 18, Milano, Italy
[2]Social Things SRL, Viale Umbria 63, 20135 Milan, Italy

Abstract

Cold-start remains a fundamental limitation of recommender systems, particularly when no interaction history is available at first use. In knowledge-intensive domains, this issue is further exacerbated by the inability of existing approaches to exploit rich semantic and contextual information available at first interaction. In this paper, we address this limitation by reframing cold start as a knowledge orchestration problem and proposing a human-centred framework that integrates natural-language preference elicitation, retrieval-grounded candidate generation, and LLM-based reranking with explanation. The framework models recommendations as prompt-mediated, user-correctable interactions, enabling structured interpretation of first-session input and iterative refinement. The approach is designed to be domain-agnostic and applicable to contexts such as education, healthcare, and cultural heritage. For evaluation, we adopt a widely used benchmark dataset as a proxy testbed, enabling controlled comparison with standard baselines. Results show that combining retrieval grounding with explainable and interactive mechanisms improves first-session recommendation quality and user transparency. This work contributes a conceptual reframing of cold start, a transferable framework, and design insights for human-centred recommendation in semantically rich, behaviourally sparse settings.

Keywords

Cold-start recommendation, Explainable recommender systems, Large language models, Human-centred AI

1. Introduction

Recommender systems play a key role in helping users navigate large, complex information spaces. Despite their widespread adoption, these systems still face a fundamental limitation: their inability to provide reliable recommendations without prior interaction data. This condition, commonly referred to as cold start, arises when a user interacts with the system for the first time and no behavioral signals, such as clicks or ratings, are available.

A comparable challenge is increasingly emerging in recent AI-assisted creation environments, which have substantially lowered the barrier for non-technical users to design digital artifacts, including websites, chatbots, and interactive applications, through no-code, low-code, or prompt-based interfaces. In these contexts, systems are expected to provide meaningful guidance and context-aware support from the very first interaction, despite the absence of structured user models, historical behavior, or explicitly defined preference representations. In this context, cold-start is no longer merely a technical limitation of recommender systems, but a fundamental bottleneck in EUD-driven AI systems, where users themselves serve as system designers while the system lacks structured mechanisms to capture and operationalize situated preferences and contextual knowledge during the initial interaction phase.

This shift highlights the need for frameworks that support not only recommendations but also the co-construction of meaningful initial system behavior, aligning with principles of meta-design and cultures of participation [1, 2]. Traditional approaches address cold-start issues by leveraging auxiliary

Proceedings of the 10th International Workshop on Cultures of Participation in the Digital Age (CoPDA 2026): Exploring the Relationship between EUD, AI-Assisted Development, and Meta-Design, June 2026, Venice, Italy.

*Corresponding author.
†These authors contributed equally.
stefano.valtolina@unimi.it (S. Valtolina); ricardo.matamoros@socialthingum.com (R. A. Matamoros); francesco.epifania@socialthingum.com (F. Epifania)
0000-0003-1949-2992 (S. Valtolina); 0000-0002-1957-2530 (R. A. Matamoros); 0000-0003-1839-9366 (F. Epifania)

information, such as item metadata, user attributes, or cross-domain signals [3, 4]. While effective in partially mitigating data sparsity, these methods still treat cold start as a problem of missing interactions, relying on indirect evidence rather than directly leveraging user input. In many scenarios, users provide explicit descriptions of their needs at first interaction, including goals and constraints that are not captured by predefined features.

As a result, cold-start failures in such domains are not only due to the absence of behavioral data, but also to the inability of existing systems to exploit rich semantic and contextual information available at first interaction. Early recommendations are typically based on weak heuristics, such as popularity or shallow similarity, leading to outputs that are poorly aligned with user intent and difficult to justify.

Recent advances in large language models (LLMs) have opened new opportunities for cold-start recommendations. LLMs can process natural language input and support semantic reasoning and explanation [5, 6], whereas conversational and interactive recommender systems emphasize user-driven refinement [7, 8]. However, purely generative approaches may lack grounding and controllability. Retrieval-grounded methods, such as Retrieval-Augmented Generation (RAG), address these limitations by anchoring outputs to external evidence [9]. At the same time, human-centred AI highlights the role of transparency and user control in high-stakes settings [10].

In this paper, we argue that cold-start recommendation should be reframed as a knowledge orchestration problem. Rather than treating cold start as a static condition caused by missing data, we conceptualize it as a process in which first-session input is progressively structured, grounded, and refined. Building on this perspective, we propose a human-centred framework that integrates semantic elicitation, retrieval-grounded candidate generation, and LLM-based reranking with explanation. Although the framework is domain-agnostic, its design is motivated by knowledge-intensive scenarios in which recommendations must be interpretable, context-aware, and aligned with user goals. For controlled evaluation, we adopt a widely used recommender-systems benchmark [11] and use interaction-based models as a warm-start reference [12].

The remainder of the paper is structured as follows. Section 2 reviews related work. Section 3 introduces the problem reframing. Section 4 presents the proposed framework. Section 5 describes the experimental setup, followed by results and discussion in Section 6. Section 7 concludes the paper, outlining limitations and future directions.

2. Background and Related Work

2.1. Cold-Start Recommendation and Hybrid Approaches

Recommender systems traditionally rely on collaborative filtering, content-based models, or hybrid approaches to infer user preferences [3, 4]. While effective in data-rich settings, these methods degrade significantly in cold-start scenarios, where interaction data is sparse or unavailable. Collaborative models, in particular, require sufficient user–item interactions to identify meaningful patterns. In contrast, content-based approaches depend on predefined features that often fail to capture the complexity of user needs.

To mitigate these limitations, prior work has explored the use of auxiliary information, including metadata, demographic attributes, and domain knowledge [4]. Knowledge-based and ontology-driven systems have been proposed to compensate for missing interactions, particularly in structured domains. Hybrid architectures attempt to combine multiple signals within multi-stage pipelines. These approaches improve robustness but still rely on indirect inference of user preferences rather than explicitly leveraging first-session input. In contrast, interaction-based models such as Neural Collaborative Filtering (NCF) achieve strong performance when behavioral data is available, providing a reference for recommendation quality in warm-start conditions [12, 13, 14].

2.2. LLM-Based, Retrieval-Grounded, and Human-Centred Recommendation

Recent advances in large language models (LLMs) have introduced new possibilities for recommendation, particularly in cold-start settings. LLMs can process natural language input and perform semantic reasoning, enabling relevance estimation without historical data [5]. They also support explanation generation [6]. However, LLMs remain black-box models. They offer limited technical transparency, but can improve perceived transparency by generating semantic justifications for recommendations.

To address these issues, retrieval-grounded methods such as Retrieval-Augmented Generation (RAG) have been proposed to anchor model outputs in external knowledge sources, improving factual consistency and reliability [9]. At the same time, human-centred AI research emphasizes the importance of transparency, user control, and iterative interaction, particularly in high-stakes domains [10]. In parallel, conversational and interactive recommender systems highlight the role of user-driven refinement and feedback in improving recommendation quality [8, 7].

Nevertheless, existing approaches typically address these aspects in isolation. Retrieval-based systems focus on grounding, while interaction-oriented systems focus on usability. This fragmentation reveals a key gap: the lack of a unified framework that integrates semantic elicitation, retrieval grounding, and interactive refinement within a single process, particularly in cold-start scenarios. Addressing this gap is the primary objective of the framework proposed in this work.

3. Problem Reframing: Cold Start as Knowledge Orchestration

Cold-start recommendation is traditionally framed as the inability to provide reliable suggestions without historical interaction data. In many knowledge-intensive settings, such as education, healthcare, or cultural heritage, however, users provide explicit first-session input, goals, constraints, and contextual requirements that remain largely underutilized in conventional recommendation pipelines.

We therefore reinterpret cold-start recommendation as a knowledge orchestration problem. The focus shifts from compensating for missing data to structuring heterogeneous first-session input into a representation that supports candidate selection, semantic alignment, and explanation. In this perspective, first-session evidence becomes the primary driver of recommendation, rather than a secondary auxiliary signal.

This shift has important implications for system design. First, it requires moving from implicit inference to explicit elicitation, where user needs are captured through natural-language interaction. Second, it introduces the need for grounding mechanisms, ensuring that recommendations are anchored in retrievable, verifiable data rather than generated unconstrained. Third, it highlights the role of iterative refinement, in which recommendations are progressively adjusted based on user feedback. Finally, it emphasizes explainability as a core requirement, since early interactions must establish trust in the absence of prior system experience.

From an EUD and meta-design perspective, this reframing has a broader implication. Cold-start can be interpreted as the initial condition in which end users, acting as system designers, define the operational semantics of an AI system through natural-language input. Rather than configuring explicit models or rules, users provide situated knowledge that the system must interpret, ground, and operationalize.

In this sense, the proposed knowledge orchestration process can be seen as an enabling layer for participatory AI systems, where users progressively shape both the recommendation outcomes and the underlying relevance criteria through interaction. This idea aligns with the vision of cultures of participation, in which users are not passive recipients of recommendations but active contributors to the system's evolving behavior.

Figure 1 contrasts the traditional view of cold start with the proposed process-oriented perspective.

Based on this perspective, we conceptualize cold-start recommendation as a three-phase process: elicitation, where user needs are expressed; grounding, where candidate items are selected based on semantic alignment; and refinement, where recommendations are explained and iteratively adjusted. The interaction between these phases is depicted in Figure 2, which illustrates the orchestration pipeline underlying the proposed framework.

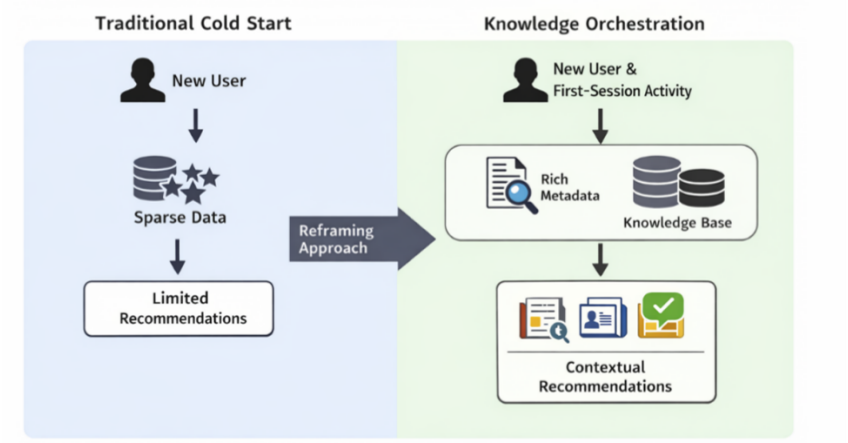

Figure 1: Cold Start Reframing: Conceptual comparison between traditional cold-start formulation (data sparsity) and the proposed knowledge orchestration perspective, highlighting the role of first-session evidence.

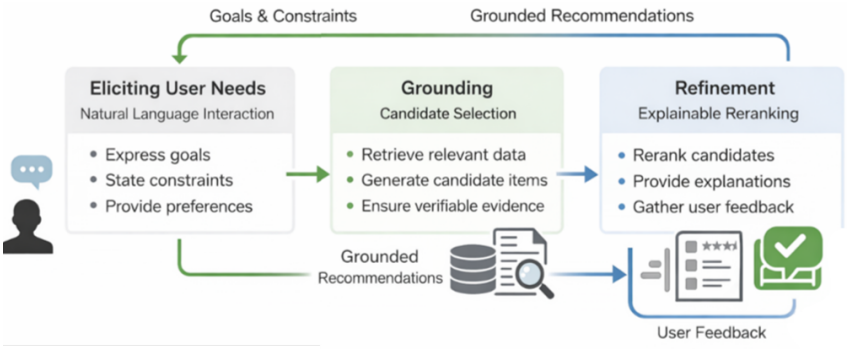

Figure 2: Framework Architecture: Overview of the proposed three-stage recommendation framework, illustrating the orchestration pipeline from user input to retrieval-grounded reranking and explanation.

Overall, this reframing shifts the focus from data availability to interaction, representation, and grounding. It provides the conceptual foundation for designing recommender systems that operate effectively from the first interaction, leveraging semantic inputs and user feedback rather than relying solely on historical behavior.

4. Proposed Framework

This section describes the proposed framework for cold-start recommendation, focusing on its implementation as a structured and grounded pipeline. The framework operationalizes the knowledge orchestration perspective through a hybrid architecture that combines interaction-based and semantic components within a unified process.

Before detailing the architecture, it is important to clarify how the proposed approach differs from existing conversational RAG systems. While conversational RAG approaches combine natural-language interaction with retrieval, they typically operate as loosely coupled pipelines focused on response generation. In contrast, the proposed framework introduces a structured orchestration model explicitly designed for cold-start recommendation, where (i) elicitation, (ii) grounded candidate generation, and (iii) constrained LLM-based reranking are integrated as interdependent stages of a single decision process. This formulation emphasizes controllability, bounded reasoning, and iterative refinement, distinguishing the framework from purely generative or conversational approaches.

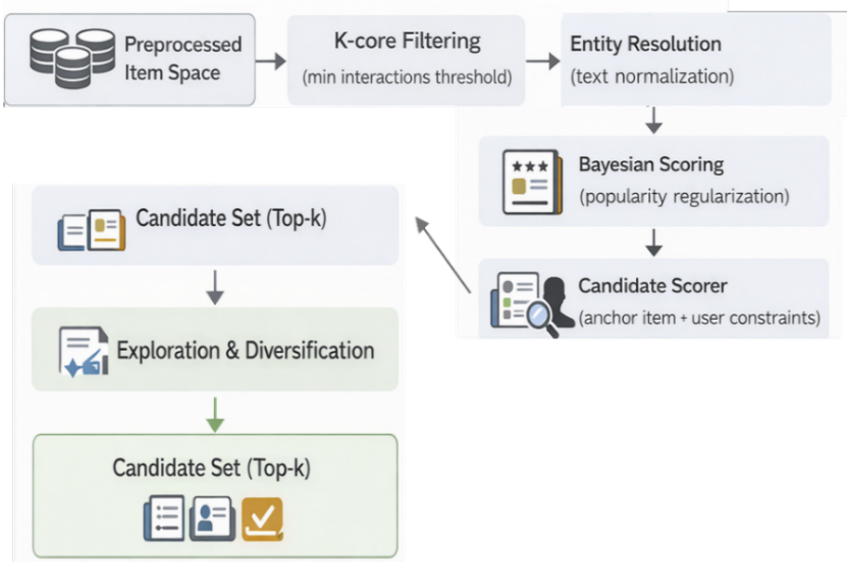

Figure 3: Candidate Generation and Selection Trade-off: Candidate generation pipeline combining preprocessing, statistical filtering, and semantic scoring. Items are filtered through k-core constraints and entity resolution, ranked using Bayesian-adjusted popularity, and refined via a candidate scorer guided by anchor items and user constraints. A diversification step ensures balanced coverage before selecting the final top-k candidate set.

4.1. Hybrid Orchestration and Input Modeling

An orchestrator activates the recommendation strategy according to the availability of interaction data. When sufficient history is available, the system can rely on an interaction-based model; however, this work focuses on the cold-start setting, where recommendations depend entirely on first-session input. User input is modeled through an anchor item and optional natural-language constraints, which define the semantic space for recommendation.

4.2. Data Preparation and Candidate Generation

The candidate generation stage operates over a preprocessed item space designed to reduce sparsity and improve statistical robustness. A k-core filtering strategy is applied, requiring a minimum number of interactions per item (e.g., 50), thus retaining only statistically reliable items and reducing noise.

A critical preprocessing step involves entity resolution between interaction data and metadata. In the absence of a shared unique identifier, items are matched through text normalization, which converts text to lowercase and removes punctuation and special characters. This enables a consistent representation across heterogeneous data sources.

Candidate ranking combines popularity and metadata-based similarity. Popularity is computed using a Bayesian-adjusted rating that regularizes local statistics with global averages and mitigates bias toward sparsely rated items. A heuristic Candidate Scorer then evaluates alignment with the anchor item and user constraints, assigning greater weight to items that share relevant attributes.

To avoid overspecialization, the framework introduces a controlled exploration mechanism. Instead of selecting only the highest-scoring items, candidates are sampled from a broader pool of relevant items, resulting in a bounded candidate set (top-k). This design balances relevance, diversity, and computational efficiency. The trade-off between candidate set size and performance is illustrated in Figure 3.

4.3. LLM-Based Reranking and Interactive Refinement

The candidate set is passed to a reranking stage based on a large language model. The model operates as a semantic ranker rather than a generator, receiving as input a structured prompt that includes the

anchor item, user constraints, and candidate items.

The reranking task is formulated as a listwise ranking problem, where candidates are evaluated based on semantic alignment with the user's request. By constraining the model to a bounded candidate set, the framework ensures grounding in real data while leveraging LLMs' reasoning capabilities. Prompt design plays a central role in this stage. In addition to direct prompting, structured reasoning strategies (e.g., Chain-of-Thought) are used to improve consistency and interpretability. A distinctive feature of the framework is the integration of explanation within the ranking process: for each recommended item, the model generates a concise natural-language justification, improving perceived transparency.

The framework also supports iterative refinement through user interaction. Additional constraints can be provided in natural language and incorporated into subsequent reranking steps. This interaction loop, illustrated in Figure 2, enables progressive alignment between system output and user intent. From a system perspective, this introduces a stateful process that refines user preferences across iterations without relying on historical data.

5. Experimental Setup

This section describes the experimental setting adopted to evaluate the proposed framework under cold-start conditions, including dataset selection, baseline methods, and evaluation protocol.

5.1. Dataset and Evaluation Protocol

Evaluating cold-start recommendations in knowledge-intensive domains is inherently challenging, as datasets combining rich semantic descriptions with sufficient interaction data are scarce and often lack reproducibility. To address this limitation, we adopt a widely used benchmark dataset in recommender systems research [11].

The dataset combines large-scale interaction data and textual metadata, enabling evaluation of both statistical robustness and semantic reasoning within a unified setting. Such a setup does not aim to reproduce domain-specific behavior, but rather to isolate the contribution of semantic elicitation and retrieval grounding under controlled conditions. For the warm-start configuration, we follow a standard leave-one-out (LOO) protocol, where the most recent interaction is used as ground truth and ranked against sampled negative items. Cold-start conditions are simulated by removing all historical interactions and generating recommendations exclusively from first-session input, forcing the system to rely solely on semantic signals.

5.2. Baselines and Evaluation Metrics

We compare the proposed framework against baseline approaches representing different recommendation paradigms. A popularity-based baseline ranks items according to global interaction frequency. A metadata-based baseline performs direct item-description matching, while a dense retrieval baseline uses embedding-based similarity. An LLM-only baseline generates recommendations directly from user input without retrieval, allowing us to isolate the contribution of grounding. An interaction-based model, Neural Collaborative Filtering (NCF), is included as a warm-start reference.

Performance is evaluated using standard ranking metrics, including Hit Rate @10 (HR@10) and Normalized Discounted Cumulative Gain @10 (NDCG@10), which capture both retrieval accuracy and ranking quality. In the warm-start setting, relevance is defined using held-out interactions. In the cold-start setting, evaluation is inherently limited to proxy-based measures and qualitative comparison across methods, due to the absence of behavioral data. For completeness, the same ranking metrics are considered under simulated cold-start conditions to support qualitative comparison across baselines.

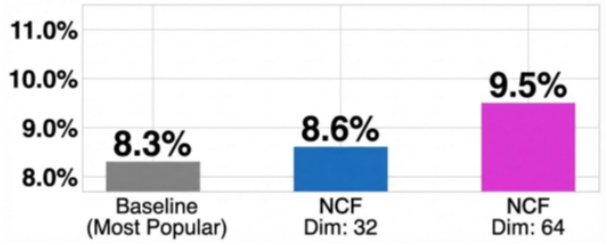

Figure 4: Warm-Start Performance (HR@10): Comparison between a popularity-based baseline and Neural Collaborative Filtering (NCF) models with increasing embedding dimensionality. Interaction-based models significantly outperform the baseline, confirming their effectiveness in data-rich scenarios and providing an upper bound for cold-start evaluation.

6. Results and Discussion

This section presents the results of the experimental evaluation and discusses their implications for cold-start recommendation. We first establish a reference point using interaction-based models, and then analyze the behavior of the proposed framework under cold-start conditions.

6.1. Quantitative Results and Performance Gap

We first analyze the interaction-based configuration as an upper-bound reference. As shown in Figure 4, the Neural Collaborative Filtering (NCF) model achieves an HR@10 of 0.095, outperforming the popularity baseline (HR@10 = 0.083) by approximately 14%.

This result confirms the effectiveness of interaction-driven models in capturing user preferences when sufficient behavioral data is available. At the same time, it highlights the performance gap between data-rich and cold-start conditions.

In cold-start scenarios, where no historical data is available, recommendations rely entirely on semantic input and retrieval mechanisms. In this setting, the proposed framework produces more stable, semantically consistent recommendations than both popularity-based and LLM-only baselines. In particular, the retrieval stage ensures that recommendations remain grounded in relevant items, while the reranking stage captures semantic nuances that are not accessible through similarity alone.

Although cold-start methods do not fully match the performance of interaction-based models, the results indicate that the proposed approach reduces the gap by leveraging structured input and semantic alignment. This suggests that first-session elicitation can partially compensate for the absence of behavioral data.

The impact of candidate set size is illustrated in Figure 5. Increasing the number of candidates improves performance up to a threshold, after which gains become marginal while computational cost increases significantly. Larger candidate sets also introduce noise that negatively affects the reasoning stage. These results justify the use of a bounded candidate set, balancing effectiveness and efficiency.

6.2. Qualitative Analysis and Implications

Beyond ranking performance, the proposed framework introduces qualitative improvements in explainability and interaction, particularly relevant in cold-start scenarios. In the absence of historical data, users cannot rely on prior system behavior to calibrate trust, making transparency and controllability essential.

A key contribution of the framework is the integration of explanation within recommendation. Each item is accompanied by a concise natural-language justification, improving perceived transparency and enabling users to inspect the rationale behind suggestions.

The framework also supports iterative refinement through natural-language interaction. Users can progressively modify their requests by introducing constraints or preferences, which are incorporated into subsequent recommendation cycles. This interaction loop transforms a recommendation from

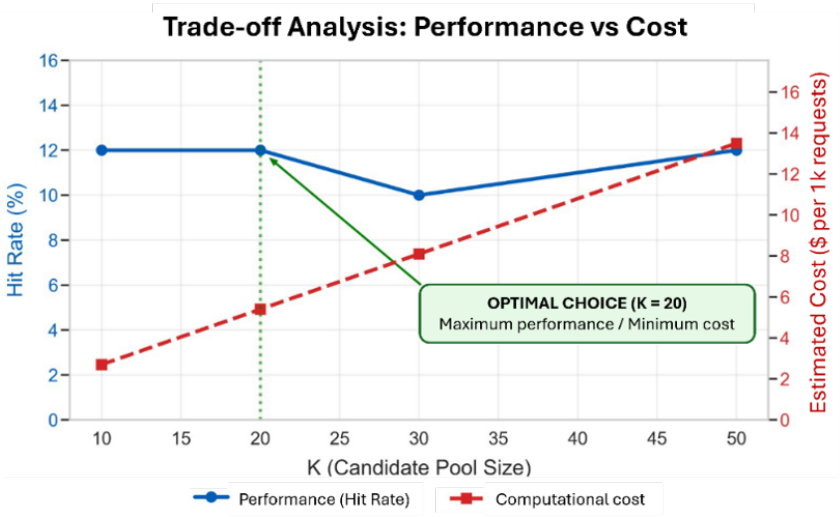

Figure 5: Candidate Set Optimization Analysis: Trade-off between recommendation performance (HR@10) and computational cost across different candidate set sizes. Results show that performance improves with larger candidate pools up to a saturation point, beyond which additional candidates introduce diminishing returns. The selected configuration (k = 20) provides the best balance between effectiveness and efficiency.

a static prediction task into a dynamic process that adapts to user input. To assess the impact of these mechanisms, we consider evidence from a user-based evaluation conducted on a comparable recommendation pipeline. A user-based evaluation with 32 participants familiar with digital search and recommendation systems provided indicative evidence of user preferences in a blind-comparison setting.

Results show a consistent preference for explanation-aware recommendations. In particular, structured reasoning approaches (e.g., Chain-of-Thought prompting) were preferred in 50% of cases, compared to 42% for direct prompting. Furthermore, iterative refinement proved effective, with the system successfully adapting to user feedback in over 90% of interactions.

Taken together, these findings indicate that recommendation quality in cold-start scenarios cannot be evaluated solely through ranking metrics. These results are consistent with prior work on interactive recommender systems, which emphasizes the importance of user control and iterative feedback in uncertain settings [8, 7]. In knowledge-intensive domains, where decisions require justification and contextual alignment, these interaction capabilities become essential for system adoption.

An important implication of this framework concerns the evolution of the knowledge base and relevance criteria over time. Although the current implementation assumes a static item space, the interaction loop naturally supports user-driven knowledge refinement, with constraints, preferences, and feedback from successive interactions interpreted as incremental contributions to the system's knowledge structure.

From a participatory perspective, this suggests a transition from static recommendation models to co-evolving systems, in which users influence not only individual outputs but also the criteria for defining relevance. This aspect opens promising directions for integrating mechanisms of user-driven knowledge curation, validation, and adaptation, in line with the principles of symbiotic and participatory AI systems.

7. Conclusion and Future Work

In this paper, we have reframed cold-start recommendation as a knowledge orchestration problem and proposed a framework that integrates semantic elicitation, retrieval grounding, and LLM-based reasoning within a unified pipeline. The results show that structuring first-session input and constraining generation via retrieval enable coherent, explainable recommendations, even in the absence of behav-

ioral data. While the proposed approach does not fully match the performance of interaction-based models, it reduces the gap between cold-start and warm-start conditions.

Beyond quantitative performance, our findings highlight the importance of interaction and explainability in cold-start scenarios. The user-based evaluation indicates that recommendation quality is not determined solely by ranking accuracy, but also by the system's ability to expose reasoning and support iterative refinement. These aspects are particularly relevant in knowledge-intensive domains, where recommendations must be interpretable, controllable, and aligned with user goals.

The experimental evaluation presents important limitations. The proxy dataset enables controlled comparison but does not fully capture the constraints of real professional domains, and the user-based evaluation provides indicative rather than statistically generalizable evidence. Future work will validate the framework in domain-specific settings, such as education, healthcare, and cultural heritage, by integrating domain-specific constraints, refining evaluation protocols, and conducting in-situ user studies.

Beyond recommender systems, this work contributes to the broader discourse on EUD and AI-assisted development by highlighting cold-start as a fundamental challenge in user-driven system design. Addressing this challenge is essential for enabling reliable, transparent, and participatory AI systems that operate effectively from their very first interaction.

Declaration on Generative AI

While preparing this work, the author used ChatGPT and Grammarly to check grammar and spelling. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] G. Fischer, End-user development and meta-design: Foundations for cultures of participation, in: V. Pipek, M. B. Rosson, B. de Ruyter, V. Wulf (Eds.), End-User Development, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 3–14.

[2] G. Fischer, E. Giaccardi, Meta-design: A Framework for the Future of End-User Development, Springer Netherlands, Dordrecht, 2006, pp. 427–457. URL: https://doi.org/10.1007/1-4020-5386-X_19. doi:10.1007/1-4020-5386-X_19.

[3] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Transactions on Knowledge and Data Engineering 17 (2005) 734–749. doi:10.1109/TKDE.2005.99.

[4] R. Burke, Hybrid recommender systems: Survey and experiments, User Modeling and User-Adapted Interaction 12 (2002) 331–370. URL: https://doi.org/10.1023/A:1021240730564. doi:10.1023/A:1021240730564.

[5] Y. Hou, J. Zhang, Z. Lin, H. Lu, R. Xie, J. McAuley, W. X. Zhao, Large language models are zero-shot rankers fornbsp;recommender systems, in: Advances in Information Retrieval: 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24–28, 2024, Proceedings, Part II, Springer-Verlag, Berlin, Heidelberg, 2024, p. 364–381. URL: https://doi.org/10.1007/978-3-031-56060-6_24. doi:10.1007/978-3-031-56060-6_24.

[6] N. Tintarev, J. Masthoff, Explaining Recommendations: Design and Evaluation, Springer US, Boston, MA, 2015, pp. 353–382. URL: https://doi.org/10.1007/978-1-4899-7637-6_10. doi:10.1007/978-1-4899-7637-6_10.

[7] D. Jannach, P. Resnick, A. Tuzhilin, M. Zanker, Recommender systems — beyond matrix completion, Commun. ACM 59 (2016) 94–102. URL: https://doi.org/10.1145/2891406. doi:10.1145/2891406.

[8] D. Jannach, A. Manzoor, W. Cai, L. Chen, A survey on conversational recommender systems, ACM Comput. Surv. 54 (2021). URL: https://doi.org/10.1145/3453154. doi:10.1145/3453154.

[9] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, D. Kiela, Retrieval-augmented generation for knowledge-intensive nlp tasks, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20, Curran Associates Inc., Red Hook, NY, USA, 2020.

[10] B. Shneiderman, Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered ai systems, ACM Trans. Interact. Intell. Syst. 10 (2020). URL: https://doi.org/10.1145/3419764. doi:10.1145/3419764.

[11] J. McAuley, J. Ni, J. Li, Amazon reviews data (2018), https://cseweb.ucsd.edu/~jmcauley/datasets.html, 2018. Accessed: 2026-05-31.

[12] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, Neural collaborative filtering, in: Proceedings of the 26th International Conference on World Wide Web, WWW '17, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2017, p. 173–182. URL: https://doi.org/10.1145/3038912.3052569. doi:10.1145/3038912.3052569.

[13] H. Abdollahpouri, R. Burke, B. Mobasher, Controlling popularity bias in learning-to-rank recommendation, in: Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 42–46. URL: https://doi.org/10.1145/3109859.3109912. doi:10.1145/3109859.3109912.

[14] W. Krichene, S. Rendle, On sampled metrics for item recommendation, Commun. ACM 65 (2022) 75–83. URL: https://doi.org/10.1145/3535335. doi:10.1145/3535335.