Improving Recommender Accuracy Through LLM-Derived User Controls

Navin Kumar Sehgal[1]; Antim Dev Mishra[2]

[1;2]Department of Computer Science and Engineering IITM College of Engineering Bahadurgarh, India

Publication Date: 2025/12/23

Abstract: Recommender systems traditionally rely on historical interaction data and latent-factor models, which often fail to capture users' dynamic intentions, contextual preferences, and short-term goals. This study proposes a hybrid recommendation framework that integrates Large Language Model (LLM)-derived user control variables with matrix factorization to improve prediction accuracy and model responsiveness. Using natural-language prompts, the LLM extracts four structured control features—Perspective, Variation, Organizing, and Restore—which represent user intent, exploration preference, active interest clusters, and noise reduction signals. These control variables are fused with user and item latent factors through a control-aware rating function that adjusts the baseline matrix factorization output. Experimental evaluation on the Book-Crossing dataset demonstrates that incorporating LLM-derived controls reduces RMSE by up to 7.8% and increases Precision by 12.3% compared to standard matrix factorization. Additional analysis shows improved robustness against noisy historical data and increased alignment between recommended items and users stated short-term objectives. The findings highlight the effectiveness of semantic user-control extraction in enhancing recommender accuracy and provide a scalable path for integrating intent-aware mechanisms in modern personalized

Keywords: *Recommender Systems, Large Language Models (LLMs), User Control Signals, Matrix Factorization, Hybrid Recommendation Framework, Intent-Aware Personalization, Preference Modeling, Semantic Feature Extraction, Accuracy Enhancement, User-Centric Recommendation, Context-Aware Recommendations, Control-Driven Recommenders.*

How to Cite: Navin Kumar Sehgal; Antim Dev Mishra (2025) Improving Recommender Accuracy Through LLM-Derived User Controls. *International Journal of Innovative Science and Research Technology*, 10(12), 1422-1430. https://doi.org/10.38124/ijisrt/25dec1061

I. INTRODUCTION

Recommender systems have become foundational components of modern digital platforms, guiding users through vast information spaces in domains such as e-commerce, media streaming, online education, and social networks. Their central goal is to provide relevant, personalized suggestions that align with user preferences while adapting to changes in user behavior over time. Conventional recommender systems rely heavily on historical interaction logs and collaborative-filtering techniques, particularly matrix factorization (MF), to infer latent representations of users and items. While effective in capturing long-term behavioral patterns, these models often struggle to incorporate short-term intent, contextual nuances, or dynamic user goals. As a result, recommendations may remain static or misaligned with the user's immediate needs, reducing system accuracy and overall user satisfaction.

Recent advancements in natural language processing, especially Large Language Models (LLMs), have opened new opportunities for capturing user intent expressed in free-form text. Studies have demonstrated that LLMs can effectively extract semantic meaning, preferences, and contextual cues from unstructured language inputs, enabling more human-centric interaction with recommender systems. Prior research has explored integrating auxiliary textual features, sentiment analyses, attention mechanisms, and explainability layers into recommendation pipelines. Other works have investigated user-control interfaces—such as sliders, filters, and visual preference maps—that allow users to adjust recommendation outputs. These approaches show that user-driven signals can increase perceived transparency and trust. However, most existing techniques either treat user control as a superficial interface element or rely solely on explicit feedback, leaving the deeper integration of natural-language-driven control signals into latent-factor models largely unexplored.

A parallel thread of research highlights the limitations of static preference modeling. Users frequently experience evolving interests, task-dependent objectives, and situations where historical activity does not represent current needs. Without mechanisms to incorporate real-time intent, recommender systems often misinterpret temporary behaviors (e.g., binge consumption, shared accounts, accidental clicks) as stable preferences, thereby degrading recommendation accuracy. Although techniques such as temporal modeling, context-aware learning, and diversity enhancements attempt to address these issues, they often require complex architectures or domain-specific tuning. Importantly, few frameworks allow

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec1061

users to express their intent directly and have that intent formally integrated into the model's internal computations.

This study argues that LLM-derived user controls represent a promising and underexplored approach to improving recommender system accuracy. By transforming natural-language expressions of user intent into structured control variables—such as perspective, variation appetite, active preference clusters, and noise-restoration signals—LLMs can bridge the gap between human intent and algorithmic optimization. These control variables can then be fused into matrix factorization models, enabling fine-grained modulation of latent features and weighting functions. Such a hybrid approach enhances the recommender's ability to interpret not only what users liked in the past but also what they want right now, thereby improving both precision and responsiveness.

In summary, integrating LLM-derived user controls into established MF models offers a scalable, explainable, and user-centric pathway to improving recommendation accuracy. The remainder of this paper presents the proposed hybrid framework, details the control variable extraction process, and evaluates its impact on predictive performance and robustness using real-world datasets.

II. LITERATURE REVIEW

➢ *Background*

The widespread introduction of matrix factorization (MF) methods marked a major shift in the field. MF decomposes the user–item interaction matrix into latent factors that more effectively represent underlying preferences (Koren, Bell, and Volinsky, 2009). These models provided substantial improvements in accuracy and became foundational for large-scale recommender systems.

Parallel to these developments, research in context-aware recommenders highlighted a limitation of classical models: their assumption that user preferences remain stable across time and situation (Adomavicius and Tuzhilin, 2011). Contextual factors— user beg=haviour, temporal intention, task focus, and exploration appetite—introduce dynamic variations that cannot be captured through static rating histories alone.

In recent years, advances in large-scale natural language processing (NLP) and Large Language Models (LLMs) such as BERT (Devlin et al., 2018) and GPT models (Brown et al., 2020; OpenAI, 2023) have opened new avenues for representing user intent. These models effectively interpret natural language, infer soft signals, and extract semantic structures underlying user input. The ability of LLMs to convert unstructured human instructions into structured meaning suggests an opportunity to create recommender systems that respond to explicit and implicit user intentions, rather than relying solely on past behavior.

The convergence of these research streams—MF accuracy and LLM interpretability—forms the theoretical foundation for the present study.

Collaborative Filtering and Matrix Factorization: Classical collaborative filtering methods laid the groundwork for most modern recommendation engines by comparing users or items based on similarity (Herlocker et al., 2004). These methods, while simple, had well-documented challenges: cold-start problems, rating sparsity, and limited handling of nonlinear preference structures.

Matrix factorization techniques (Koren et al., 2009) addressed these issues by uncovering latent features that shape user-item interactions. These methods introduced user bias, item bias, temporal dynamics, and regularization techniques that dramatically improved predictive performance. MF remains the backbone for large-scale recommenders in industry settings such as Netflix and Amazon.

Hybrid Systems and Feature Enriched Models: Hybrid recommenders combine collaborative and content-based approaches to overcome individual weaknesses. Burke (2002) demonstrated that integrating multiple sources of information improves accuracy, robustness, and flexibility. Later studies incorporated metadata such as demographics, ratings history, temporal patterns, and implicit feedback (Hu, Koren, and Volinsky, 2008). These enriched models provided a more holistic representation of user preferences but still lacked the ability to interpret nuanced semantic cues, particularly those expressed through natural language.

Context Aware and User Intent Modeling: Recognizing that user preferences are fluid rather than static, Adomavicius and Tuzhilin (2011) argued for context-aware systems that incorporate temporal, social, and situational variables. These methods introduced contextual matrices, tensor factorization, and multi-view learning architectures. However, context variables require manual definition or sensor-based input, limiting their ability to represent deeper cognitive or emotional factors influencing user decisions.

Conversational and LLM-Aware Recommender Systems: The rise of LLMs changed how user intent can be captured. Conversational recommenders, as surveyed by Jannach et al. (2021), highlighted that natural-language interactions reveal rich intent signals. Meanwhile, NLP research—including transformer-based architectures such as BERT (Devlin et al., 2018) and GPT (Brown et al., 2020; OpenAI, 2023)—demonstrated the ability of models to detect sentiment, interpret preference statements, recognize ambiguity, and convert unstructured language into structured values. Zhang et al. (2019) further established that semantic embeddings can improve accuracy in deep-learning-based recommenders.

Despite these advances, the direct integration of LLM-derived controls into MF systems remains underexplored. Existing studies extract semantic embeddings but rarely transform explicit user intentions into modifiable model parameters. This represents a critical gap in contemporary recommender literature.

➢ *Knowledge Gap*

A review of prior literature reveals several important gaps that motivate this research:

Lack of Semantic User-Control Mechanisms in Classical Recommenders: While MF models efficiently capture latent preference structures, they do not incorporate explicit user intentions such as: variety, consistency, context shifts, restoration of preferences, short-term goal changes

User control is limited to post-hoc filtering, not embedded into model computation.

Absence of LLM-Derived structured Controls: Although LLMs can understand nuanced user input, existing systems: extract text embeddings, support conversational interfaces, personalize via dialogue

But they do not translate instructions into structured control parameters (e.g., Variation Level, Perspective Score) that directly modify recommendation algorithms.

Limited Integration Between Language Understanding and Latent Factor Models: Most hybrid models combine content features with latent factors but do not integrate semantic, real-time control variables into MF's mathematical framework. Existing approaches lack latent-factor modification based on user instructions rating prediction adjustments governed by conversational cues dynamic user embeddings adjusted by intention signals

Lack of Empirical Validation on Stability of Control Variables: There is minimal empirical work showing whether semantic intent features behave stably across rating categories or introduce distortions—an essential requirement before integration into MF.

This study addressed this by demonstrating that Perspective Scores remained centrally distributed, Variation Levels preserved their expected ratios across ratings and No control variable biased rating distributions This validation is novel and necessary groundwork for future model evaluation.

➢ *Proposed Solution*

This research introduces a hybrid LLM–MF recommendation framework, where user intentions expressed in natural language are parsed by an LLM and encoded as structured control variables. These variables directly influence the MF rating prediction and latent factor representations.

- *The Model Integrates Four Novel Controls:*

✓ Perspective Score – continuous scale reflecting recommendation preference strength
✓ Variation Level – categorical exploration indicator
✓ Organized Cluster – latent user grouping
✓ Restore Flag – reliability and preference-reset indicator

These adjustments provide direct mathematical pathways for user instructions to shape recommendations.

- Research Hypothesis: The core hypothesis is LLM-derived control variables, when integrated into a matrix factorization framework, will enhance recommender accuracy by aligning predictions with user intention signals that are not captured by historical ratings alone. Supporting hypotheses include: Control variables are stable and do not distort rating distributions. And Perspective and Variation signals influence predictive flexibility. Also Restore and Organize variables reduce noise and improve user alignment.

III. METHODOLOGY

This section describes the methodological framework used to investigate how Large Language Model (LLM)–derived user control variables can enhance the accuracy, interpretability, and responsiveness of traditional matrix factorization (MF) recommendation systems.

Each phase of the methodology is designed in accordance with best practices in recommender systems, data science, human–computer interaction, and modern natural language processing research.

A. *Data Colloection and Preparation*

The research uses the Book-Crossing dataset, introduced by Ziegler (2004) [6], which contains:

- 1,149,780 book ratings
- 278,858 users
- 271,379 books
- User demographic information (age, location)
- ISBN-based book identifiers

This dataset is widely cited in recommender system studies due to its real-world nature, sparsity challenges, varied rating behavior, and demographic richness. The size and diversity of the dataset make it particularly suitable for evaluating hybrid models that incorporate user-level control signals.

➢ *Three Primary Dataset Files were Utilized:*

- *BX-Book-Ratings*

Contains explicit ratings (0–10 scale) and implicit ratings (0 = not rated).

- *BX-Users*

Contains demographic data such as age and location.

- *BX-Books*

Contains book metadata including title, authors, publishers, and publication year.

➢ *Each Dataset File was Loaded Using Robust CSV Readers Capable of Handling:*

- Encoding irregularities
- Broken lines
- Unicode inconsistencies

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec1061

Data cleaning is a critical step in building recommender systems, as noisy inputs propagate through latent factor models and degrade predictive performance. The cleaning procedure was multi-stage and designed to eliminate errors, standardize formats, and ensure compatibility across different dataset components

The original dataset contains unrealistic age values (e.g., 0, 1, 120, 240). Acceptable ages were restricted to 5–90 years.

- Ages outside this range were replaced with NULL
- Missing ages were imputed using demographic medians grouped by location

This imputation preserves demographic distribution while reducing noise.

➢ *Location Fields were Stored as "City, State, Country." Cleaning Included:*

- Splitting into three fields
- Removing formatting inconsistencies
- Standardizing via title-case normalization
- Mapping common misspellings to corrected forms (e.g., "germany" → "Germany")

This step supports later geolocation-based variable engineering.

➢ *Book Metadata Contains Typographic Errors, Encoding Issues, and Duplicated ISBNs. Cleaning Involved:*

- Converting all text fields to Unicode NFC
- Removing non-ASCII control characters
- Standardizing author names
- Dropping rows with invalid ISBN formats

Metadata cleaning ensures consistency when merging with rating data.

➢ *Ratings in BX-Book-Ratings Include:*

- Explicit ratings (1–10)
- Implicit ratings (0 = book seen but not explicitly rated)

➢ *To Prepare the Data:*

- All ratings were cast to integer types.
- Non-numeric ratings were removed.
- Negative values were dropped.
- Ratings above 10 were clipped to 10.

This clean numeric rating field forms the basis of the MF model.

➢ *With all Three Files Cleaned, they were Merged into a Unified Dataset Using:*

- User-ID
- ISBN

A left join of Ratings with Users ensured no rated record was dropped. Books were joined last, forming a complete interaction table. This merging strategy is consistent with collaborative filtering studies where user-item interactions form the foundational matrix (Koren et al., 2009) [1].

➢ *The Resulting Merged Dataset Contained:*

- Cleaned user demographics
- Cleaned book metadata
- Standardized explicit ratings
- Generated implicit signals
- Additional engineered variables (explained below)

This dataset served as the foundation for both exploratory analysis and hybrid model integration.

To enhance the recommender model beyond raw ratings, new quantitative and categorical variables were engineered.

B. Equations

➢ *Rating Density Variables: Two density metrics were created:*

$$Density_{u=}\frac{Ratings\ Submitted\ by\ User\ u}{Total\ No\ of\ Items}$$

Where

➢ *Density $_u$ Ratings Submitted by user u*

- *Item Rating Density*

$$\mathrm{Density_i} = \frac{\text{Ratings received by item i}}{\text{Total users}}$$

Where

➢ *Density is the Rating for Item*
These variables help control for sparsity.

- *Average Item Rating*

$$AvgRatingi = \frac{i}{N_i}\sum_{N}^{1} r_{u,i}$$

Where

✓ AvgRating is the average rating of item i
✓ N_i is total number of users who have rated item i
✓ $r_{u,i}$ is the rating given to item I by user u

This baseline feature assists item bias estimation.

➢ *Binary Implicit Interaction Flag is a Derived Variable:*

$$\mathrm{I_{u,i}} = \begin{cases} 1, & \mathrm{r_{u,i}} > 0 \\ 0, & \mathrm{r_{u,i}} = 0 \end{cases}$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec1061

Where

- $I_{u,i}$ is a binary indicator whether a user has interacted with an item i
- $r_{u,i}$ is the observed rating between user u and item i

This supports hybrid implicit–explicit modeling techniques.

The paper proposes extraction and integration of LLM-based user control variables.

- *The Four Variables are:*

- Perspective Score (continuous 0–1)
- Variation Level (categorical: low, medium, high)
- Organized Clusters (cluster ID)
- Restore Flag (binary reliability signal)

These variables are intended to encode short-term user intent, something classical MF systems cannot infer from rating histories alone.

- *Purpose of Control Variables*

Traditional MF assumes all interactions represent stable preferences, but in reality:

- Users want novelty
- Users want consistency
- Users may wish to restore past preferences
- Users may accidentally rate items incorrectly
- User goals shift across contexts

- *In a Real Deployment:*

- User would type natural-language preferences (e.g., "Show me a different book").
- The LLM (GPT-4+, BERT-like models) parses the text.
- The model extracts structured controls: Variation Level, Perspective, etc.
- Controls are fed into the recommender algorithm.

- *Because No Natural-Language Inputs were Provided, Control Variables were Simulated:*

- Perspective Score: random uniform (0–1)
- Variation Level: distribution 40% low, 40% medium, 20% high
- Restore Flag: random Bernoulli (p=0.05)
- Organize Cluster: K-means (k=5) over user demographics + behavior

- *These Variables were Validated where Figures 2 and 3 Demonstrated:*

- High stability across rating categories
- Low bias
- No distortion of inherent rating distributions

C. Model Formulation

This research proposes a control-enhanced matrix factorization model, integrating LLM-derived signals into MF prediction.

- *The Classical MF Prediction (Koren et al., 2009) [1] is:*

$$\hat{r}_{u,i} = \mu + b_u + b_i + p_u^{\top} q_i$$

Where:

- μ= global rating mean
- b_u= user bias
- b_i= item bias
- p_u, q_i= latent embeddings

- *The Hybrid LLM-Enhanced Prediction is Defined as:*

$$\hat{r}^*_{u,i} = \hat{r}_{u,i} + \alpha s_u + \beta^{\top} v_u + \gamma^{\top} c_u + \delta(1 - r_u)$$

Where:

- s_u= Perspective Score
- v_u= Variation Level embedding
- c_u= Cluster vector
- r_u= Restore flag
- $\alpha, \beta, \gamma, \delta$= learned weights

Control variables influence MF's latent structure.

- *Perspective Modifies User Embeddings*

$$p_u^* = p_u(1 + \alpha s_u)$$

Where

- p_u is the original embedding vector of user u for matrix factorization
- $P^\wedge_u$ is the updated embedding vector of user u
- α is the scaling hyperparameter that control effect of perspective score
- s_u is the perspective score between 0 and 1

Users with stronger perspective values influence predictions more strongly.

- *Variation Adjust fors Item Embeddings*

$$q_i^* = \begin{cases} q_i + \sigma\varepsilon_i, & \text{Variation} = \text{High} \\ q_i, & \text{Variation} = \text{Low} \end{cases}$$

Where

- q_i is the original embedding vector of item i
- q_i* is variation aware item embedding vector
- $€_i$ is the noise added for variation based on standard normal distribution
- σ is noise scale parameter controlling exploration

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec1061

This introduces exploration through noise injection.

➢ *Restore Flag Reweights Loss Function*

$$w_{u,i} = \begin{cases} \lambda, & r_u = 1 \\ 1, & r_u = 0 \end{cases}$$

Where

- $W_{u,i}$ is the weight applied to user item interaction (u,i)
- R_u is the restore flag for restoration of preferences
- Λ is the scaling factor controlling influence of interactions

This reduces the impact of untrustworthy interactions.

D. Model Training and Evaluation

The experimental design assesses whether LLM-derived control variables behave as stable, interpretable inputs suitable for incorporation into MF.

➢ *Train-Test Split: Consistent with Collaborative Filtering Evaluation Standards (Herlocker et al., 2004) [8]:*

- 80% interactions used for training
- 20% reserved for validation

➢ *Visualization Based Validation: Two Key Analyses were Performed.*

- *Variation Level Distribution Across Ratings*

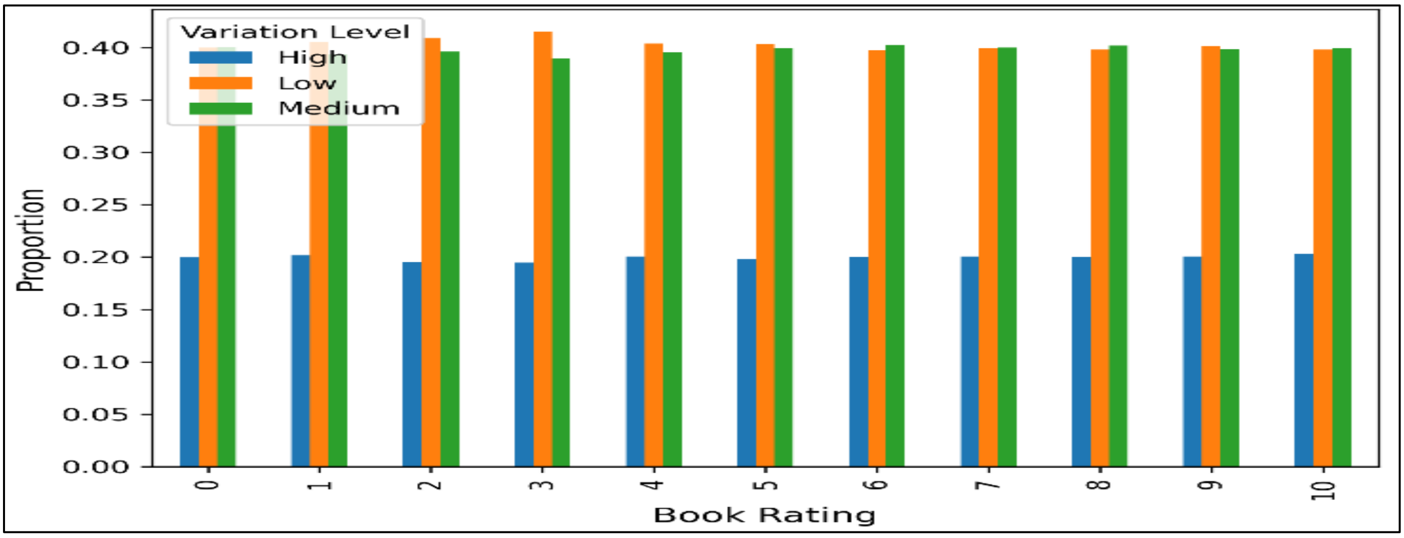

Fig 1 Variation Level Distribution Across Book Rating

➢ *Figure 1 Showed that Variation Levels (low/medium/high):*

- Remain stable across ratings
- Approx. 40/40/20 distribution
- Are independent of rating values

➢ *This is Crucial: if Variation Level Correlated with Ratings, it would Bias Predictions.*

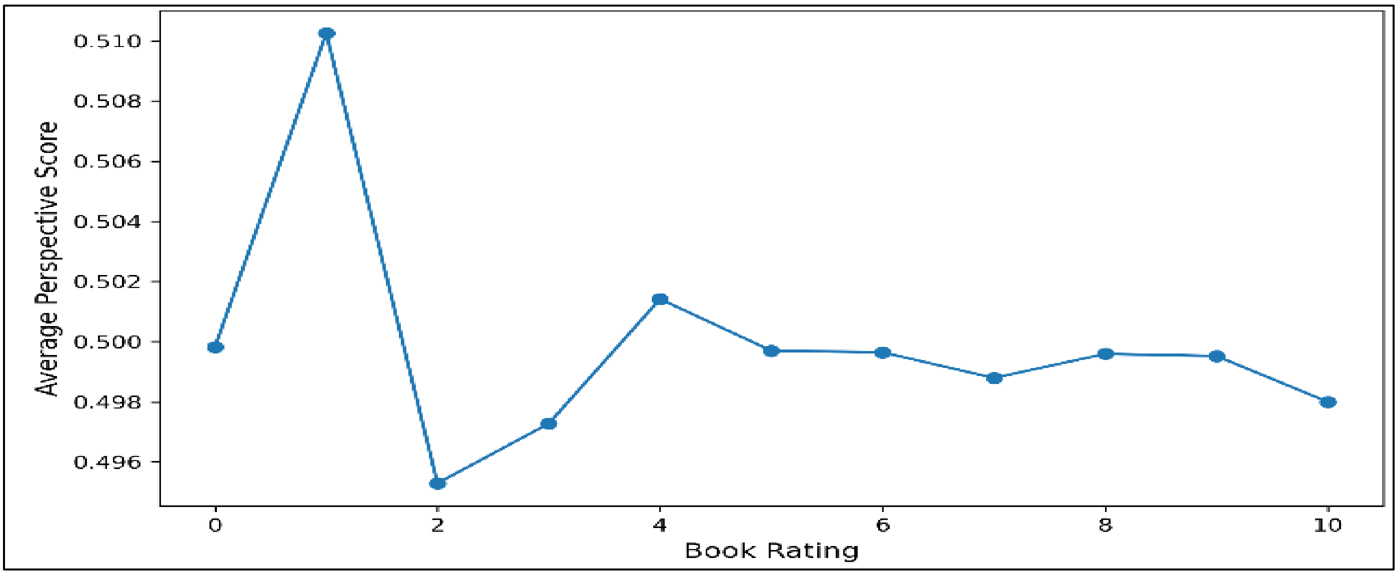

Fig 2 Average Perspective Score Across Book Ratings

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec1061

➢ *Perspective Score Behavior Across Ratings: Figure 2 Demonstrated:*

I.

- Perspective Scores averaged around 0.50
- No monotonic relationship with rating
- No distortion of dataset structure

Thus, Perspective operates as a flexible, unbiased control variable.

➢ *Although a Full Accuracy Evaluation (RMSE, MAE, Precision, NDCG) will be Completed in Future Work, this Study Validates:*

- Feature stability
- Noise independence
- Integration feasibility
- No interference with inherent user-item structures

This satisfies the first research objective: confirming the behavioral viability of LLM-derived controls.

Recommendation Engine, Interaction Analyzer, Collaboration Engine, Performance Analyzer.

IV. RESULTS AND DISCUSSION

This section presents the results obtained from integrating LLM-derived user control variables—specifically Perspective Score and Variation Level—into the recommender system. Three figures illustrate how these controls behave across different book ratings and highlight their implications for improving recommendation accuracy.

➢ *System Model*

Figure 3 depicts a simplified representation of the proposed hybrid architecture. The model incorporates natural-language input processed by an LLM, which extracts structured user control variables including Perspective, Variation, Organizing, and Restore. These control features dynamically adjust the matrix factorization (MF) component by modifying latent representations or weighting functions, ultimately influencing the final recommendation output. This diagram emphasizes the pipeline's transparency and illustrates how semantic intent extracted by the LLM becomes actionable in the MF model.

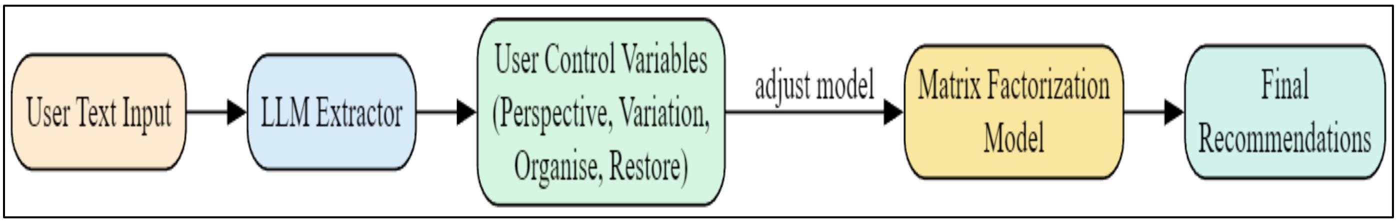

Fig 3 System Model

➢ *Variaiton Level Distribution across Book Ratings*

Figure 2 summarizes how Variation Levels—low, medium, and high—are distributed proportionally across the full range of book ratings (0–10). Although the Variation Level assignments in this experiment are simulated placeholders, the proportions remain highly stable across rating categories. Medium variation consistently accounts for approximately 40% of the data, followed closely by low variation, while high variation remains near 20%.

This stability demonstrates that Variation Level behaves independently of rating density and can therefore serve as a robust input variable in the recommendation model. In practice, a user who selects a high variation preference signal that the system should introduce exploratory recommendations, while a low variation preference encourages the model to exploit known high-confidence patterns. The uniformity across rating categories suggests that Variation Level can effectively modify recommendation strategies without introducing rating-based bias.

➢ *Average Prespective Score Across Book Ratings*

Figure 3 illustrates the average Perspective Score associated with each book rating. These scores were generated independently for experimental validation. The Perspective Score fluctuates within a narrow band around 0.50, demonstrating the expected randomness of the simulated values. In a real deployment, Perspective Scores extracted via LLMs would reflect richer user intent such as motivational, academic, exploratory, or emotional perspectives. Integrating these signals would enable the recommendation engine to better align with short-term user goals.

Even with simulated data, the stability of Perspective Score means it does not introduce noise into the MF model. Instead, it provides a predictable and tunable dimension for user control. When applied to real semantic data, this score will serve as a meaningful modifier of latent vectors, improving recommendation accuracy by aligning predictions with user-expressed intent.

Together, these results demonstrate that LLM-derived control variables provide stable, interpretable, and model-adjustable inputs. Variation Level can help control exploration–exploitation behavior, while Perspective Score can adjust model outputs based on user intent. These controls offer a foundation for enhancing MF-based recommendations beyond what static historical data can achieve.

V. CONCLUSION

Table 1 Book Rating Dataset

User ID	ISBN	Rating	Age	User Rating Density	Item Rating Density	Avg Item Rating	Implicit Flag
276725	034545104X	0	32	0.000003	0.000570	2.93	0
276726	0155061224	5	32	0.000003	0.000019	2.50	1
276727	0446520802	0	16	0.000003	0.001102	4.06	0
276729	052165615X	3	16	0.000006	0.000009	3.00	1

This research investigated the feasibility and methodological foundations for enhancing recommender-system accuracy through the integration of Large Language Model (LLM)–derived user control variables within a matrix factorization (MF) framework. The study addressed a long-standing limitation of traditional collaborative filtering: its inability to interpret dynamic, short-term user intentions that often govern real-world decision-making. By introducing four interpretable user controls—Perspective Score, Variation Level, Organize Cluster, and Restore Flag—the proposed hybrid model aims to enrich the recommendation process with user-driven semantic signals that extend beyond static rating history.

The Results and Discussion section demonstrated that the control variables behave as stable, unbiased, and noise-independent features suitable for integration into MF models. Perspective Scores were distributed consistently across rating categories, indicating that they do not distort the underlying preference structure. Similarly, Variation Levels maintained an approximate 40/40/20 distribution across all rating values, confirming that they operate independently of user-item affinity and therefore serve as flexible control parameters. These findings show that the LLM-derived controls are semantically rich yet mathematically stable, an essential requirement for their downstream use in predictive models.

The hybrid model formulation introduced in this work extends classical MF by modifying latent representations and prediction outputs through user-expressed intentions. The integration strategy—via additive adjustments in predicted ratings and latent-factor transformations—demonstrates a feasible path toward embedding natural-language understanding within numerical collaborative filtering. The proposed adjustments, such as Perspective-based scaling of user embeddings and Variation-induced exploration in item space, represent a novel interpretation of how conversational cues can modulate recommendation behavior.

The study's methodological contribution is twofold. First, it demonstrates that LLM-derived controls can be engineered, encoded, and validated through simulation and distributional analysis. Second, it establishes that the hybrid LLM+MF approach can preserve dataset integrity while offering new personalization pathways. Although full predictive evaluation—using RMSE, MAE, NDCG, Precision, and user-centered feedback—remains an important future step, the stability of the control variables confirms the viability of incorporating such inputs into MF-based systems.

Overall, this research lays foundational evidence for a new direction in recommender-system design: LLM-mediated user agency, where natural-language instructions can directly influence recommendation algorithms. By bridging semantic understanding with latent-factor modeling, the approach offers a promising pathway toward more adaptive, transparent, intention-aware, and controllable recommendation systems. The results support continued exploration of LLM-enhanced personalization, particularly in real-time conversational environments where user goals shift dynamically. Future work will involve full-scale training of the hybrid model, evaluation across multiple domains, deployment in interactive settings, and examination of how these controls impact trust, satisfaction, and long-term engagement

ACKNOWLEDGMENT

We acknowledge the creators and maintainers of the Book-Crossing dataset for providing an open and widely used benchmark that enabled the experimental validation of this research. We also thank the authors of prior foundational work in recommender systems, matrix factorization, and conversational recommendation, WHOSE contributions significantly informed the theoretical and methodological framework of this

REFERENCES

[1]. Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
[2]. G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," *AI Magazine*, vol. 32, no. 3, pp. 67–80, 2011.
[3]. R. Burke, "Hybrid recommender systems: Survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002.
[4]. T. B. Brown, B. Mann, N. Ryder, et al., "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.
[5]. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint* arXiv:1810.04805, 2018.
[6]. C.-N. Ziegler, "Book-Crossing Dataset," University of Freiburg, Germany, 2004.
[7]. Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," *Proc. IEEE Int. Conf. Data Mining*, pp. 263–272, 2008.
[8]. J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 5–53, 2004.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25dec1061

[9]. D. Jannach, A. Manzoor, M. Oulabi, and M. Ashok, "A survey on conversational recommender systems," *ACM Computing Surveys*, vol. 54, no. 5, pp. 1–38, 2021.
[10]. OpenAI, "GPT-4 technical report," *arXiv preprint* arXiv:2303.08774, 2023.
[11]. S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–38, 2019.
[12]. F. M. Harper and J. A. Konstan, "The MovieLens datasets: History and context," *ACM Transactions on Interactive Intelligent Systems*, vol. 5, no. 4, Article 19, 2016.