Cross-Media Recommendation: A Multimodal Approach in Vector Recovery Systems

Pedro Paulo Carvalho Vieira[1,2], **Rodrigo Santana Esperidião**[1,2], **Luiz Felipe Alves Pinheiro**[1], **Willgnner Ferreira Santos**[1,2] and **Gustavo Siqueira Vinhal**[1]

[1] SENAI Fatesg College
R. 227-A, 95 – Setor Leste Universitário, Goiânia – GO, 74610-155

[2] SENAI Fatesg College – Applied Artificial Intelligence Core (NIAA)

{ppaulo030601, luizfelipepinheiro29}@gmail.com
rodrigo1700079@go.estudante.senai.br

Abstract. *Traditional recommendation systems operate in isolated domains, reinforcing filter bubbles that limit content discovery. This work proposes a content-based cross-media approach connecting films, series, games, anime, and manga in a unified semantic space. Multimodal representations were extracted from textual metadata and cover images using the SigLIP 2 vision-language model. The architecture combines Filtered Vector Search with a score boosting mechanism that re-ranks candidates by weighting popularity scores toward the user's consumption niche. Empirical evaluation over 20,993 latent vectors showed the multimodal approach outperforms purely textual and visual encodings in cross-media retrieval, achieving an inter-category Recall@10 of 0.7545.*

1. Introduction

A filter bubble is an information environment in which recommendation algorithms, combined with a user's prior behavior and preferences, select and prioritize content similar to what has already been consumed, reducing exposure to new perspectives [Pariser 2011]. As digital platforms have become central to entertainment consumption, recommendation systems directly influence individual discovery opportunities [Tasenţe 2025].

Despite the existence of broad cataloging platforms such as IMDb, Letterboxd, and MyAnimeList, they generally operate as isolated silos specialized in a single type of media [Fernández-Tobías et al. 2012]. Recent studies have attempted to break this fragmentation by introducing serendipity through Large Language Models (LLMs) [Sassi 2025], but the direct use of these models in production imposes severe operational challenges, including high latency, elevated costs, and the risk of algorithmic hallucinations [Zhang et al. 2025].

To overcome these limitations and mitigate format isolation, this work proposes a unified cross-media recommendation mechanism. The architecture is based on a Filtered Vector Search model supported by multimodal representations generated by the SigLIP 2 vision-language encoder [Tschannen et al. 2025]. By combining visual and textual information in a shared vector space, the system enables efficient semantic similarity computation across films, series, games, anime, and manga, providing the infrastructure needed for discovery beyond the ecosystems traditionally consumed by users.

2. Related Work

Vector projection of items is a well-established technique that evolved from foundational models such as Item2Vec [Barkan and Koenigstein 2017] to architectures capable of transferring knowledge across multiple domains, as in UniSRec [Hou et al. 2022]. More recently, to mitigate filter bubbles and promote serendipity, proposals such as ScientiaRec [Sassi 2025] and IL-Rec [Zhang et al. 2025] have adopted LLMs in the generation of recommendation policies.

However, given the latency, cost, and hallucination bottlenecks of real-time generative approaches, this work adopts pre-computed multimodal embeddings via SigLIP 2 [Tschannen et al. 2025]. In this setting, the main structural challenge is the recall drop caused by rigid filters in approximate vector search. To address this trade-off, the proposed architecture combines the Qdrant vector database with heuristic post-processing steps (Score Boosting), adjusting the relevance of filtered items without compromising retrieval stability and speed.

3. Methodology

The proposed recommendation model architecture aims to enable real-time cross-media discovery. The algorithmic flow is structured in four fundamental steps: multimodal vector representation, dynamic query generation, filtered vector search, and heuristic post-processing.

3.1. Data Collection and Processing

Metadata was extracted from multiple sources through integration with public APIs from reference providers in their respective niches: TMDB for films and series, IGDB for electronic games, and MyAnimeList for anime and manga. To prevent overlap and duplication of instances across platforms, strict filtering was applied during collection requests. This strategy guaranteed that each media type was assigned to a single source, ensuring, for example, that Japanese animations were extracted exclusively from MyAnimeList.

In pre-processing, synopses were standardized and translated to Portuguese via the LibreTranslate API. Although SigLIP 2 has native multilingual capability, recent evaluations of the model show uneven performance across languages and alignment failures in clustering tasks in the latent space [Sriratanawilai et al. 2025]. This limitation adds to the structural tendency of large models to partition the vector space and create language-specific subspaces, limiting the sharing of semantic representations [Wu et al. 2020]. Standardizing the catalog in a single language ensures that the cohesion of the latent space reflects exclusively the thematic and visual affinity of the works.

In parallel with textual normalization, original cover images were submitted directly to the native Transformers library pipeline, whose visual processor handled automatic resizing and tensor standardization, ensuring geometric adequacy for the encoder's input requirements.

Finally, numeric rating metadata also underwent structural alignment: since providers use divergent scoring scales (e.g., 0 to 10 in TMDB and 0 to 100 in IGDB), all collected scores were proportionally rescaled to a universal scale of 1 to 5, establishing a uniform comparative base for the recommendation heuristics.

3.2. Multimodal Representation

To enable cross-domain recommendation, all works are indexed using the SigLIP 2 model, an advanced encoder that aligns language and vision in a multilingual semantic space [Tschannen et al. 2025]. Textual metadata (synopsis) and visual attributes (covers and posters) are simultaneously converted into normalized dense vectors. This approach ensures that geometric proximity in the latent space reflects the similarity of works' thematic proposals, regardless of media format.

3.3. Query Vector Generation

The inference process starts from the definition of user seed media. These seeds can be defined explicitly (from a direct request) or implicitly inferred by the system based on interaction history (favorited items or items with maximum ratings).

The consolidated query vector ($\vec{q}$) is built by averaging the embeddings ($\vec{v}_i$) of all seed media, a technique known as mean pooling [Reimers and Gurevych 2019], and then normalizing the result using the L2 norm to keep distances geometrically consistent across search operations:

$$\vec{q} = \frac{\sum_{i=1}^{M} \vec{v}_i}{\| \sum_{i=1}^{M} \vec{v}_i \|_2} \tag{1}$$

where M represents the number of valid seed media retrieved.

3.4. Filtered Vector Search

Using the consolidated vector $\vec{q}$, the system executes an Approximate Nearest Neighbor (ANN) search in the vector database. Latent space partitioning is based on the HNSW (Hierarchical Navigable Small World) algorithm [Malkov and Yashunin 2016], operating under standard configurations optimized for an efficient balance between indexing time, memory consumption, and precision (such as $m = 16$ maximum edges per node and $ef_construct = 100$).

To support directed recommendation, the architecture employs Filtered Vector Search, which allows relational operators (that is, logical conditions such as restrictive media category filters) to be applied directly over the semantic search [Chronis et al. 2025]. Similarity between the query vector and candidates is quantified by the dot product of normalized vectors (cosine similarity) [Vechtomova 2009].

As pointed out in the literature, imposing rigid filters on approximate vector searches modifies the distribution of the search space and can cause early interruptions in HNSW graph traversal, resulting in severe recall drops [Chronis et al. 2025]. To address this architectural challenge and maintain retrieval stability, the system implements two combined strategies: an exact payload index for category filtering and a deliberate over-fetching policy.

In practical terms, the algorithm retrieves an expanded margin of candidates equivalent to 25 times the final limit requested by the application. This large pool ensures that, even with path pruning in the graph and subsequent exclusion of original seeds, the system provides abundant and semantically relevant material for the subsequent heuristic refinement step (Section 3.6), ensuring adequate volume for calculating B_{pop} and B_{rating} without the premature loss of relevant candidates.

3.5. Popularity Normalization and Alignment

Due to the heterogeneous nature of the data sources (TMDB, IGDB, and MyAnimeList), evaluation and engagement metrics have fundamentally different scales and variance distributions. To ensure that cross-media recommendations are not biased by platforms with inflated vote volumes, a popularity alignment was applied based on the Bayesian Mean, followed by standardized rescaling. The normalization process is divided into three mathematical steps.

Category Constants Definition: For each category C (which intrinsically separates data origins, such as films, games, or anime), two global statistical constants are defined:

- m_C: The 95th percentile of the number of ratings within category C. This value acts as a confidence threshold, representing the minimum number of votes required for an item to be considered statistically popular in its respective platform.
- c_C: The arithmetic mean of ratings (rating_avg) of all items in category C.

Bayesian Mean Adjustment: To mitigate distortion caused by items with extremely high ratings but very few votes, an adjusted score (BR_i) is calculated for each item i using the Bayesian Mean formula:

$$BR_i = \left(\frac{v_i}{v_i + m_C} \right) R_i + \left(\frac{m_C}{v_i + m_C} \right) c_C \tag{2}$$

where v_i is the total number of ratings that item i received, and R_i is the original average rating of item i. This formulation acts similarly to a Gaussian prior: items with few votes ($v_i \ll m_C$) have their ratings pulled toward the category mean (c_C), while items with many votes ($v_i \gg m_C$) retain their original rating (R_i).

Min-Max Normalization on a Universal Scale: Since the Bayesian mean distributions (BR) may still vary across categories due to differences in original rating scales (e.g., 0–10 in TMDB vs. 0–100 in IGDB), intra-category Min-Max normalization is applied to align popularity on a universal scale from 0 to 100:

$$Popularity_i = \frac{BR_i - \min(BR_C)}{\max(BR_C) - \min(BR_C)} \times 100 \tag{3}$$

where $\min(BR_C)$ and $\max(BR_C)$ represent the minimum and maximum values of the adjusted score within category C, respectively. This mapping ensures a fair and comparable popularity metric across the entire catalog, feeding the score boosting mechanism.

3.6. Heuristic Relevance Adjustment (Score Boosting)

The vector recommendation engine aims not simply to suggest the most popular items in the catalog, but to respect the user's niche consumption profile. For this, a boost factor based on a radial basis function (Gaussian) is applied.

Given a query built from a set of seed items S, the target popularity μ_P is calculated as the mean popularity of the seeds:

$$\mu_P = \frac{1}{|S|} \sum_{s \in S} Popularity_s \tag{4}$$

For any candidate item i returned by the vector search, the popularity deviation is defined as $\Delta_i = Popularity_i - \mu_P$. The multiplicative popularity boost factor (B_{pop}) is then calculated as:

$$B_{pop}(i) = 0.15 \cdot \exp\left(-\frac{\Delta_i^2}{2\sigma^2}\right) \tag{5}$$

where 0.15 defines the maximum bonus amplitude (15%) and $\sigma = 20.0$ is the standard deviation of the curve. The maximum boost occurs when $\Delta_i = 0$, i.e., when the candidate has the same popularity as the seed mean. A standard deviation of 20.0 creates a comfort zone, ensuring that items with a difference of up to ≈ 20 points still receive a reasonable bonus, while candidates with much higher or lower popularity levels (e.g., $|\Delta_i| > 40$) receive a boost close to zero.

In parallel with niche matching, a rating bonus (B_{rating}) is applied to give slight priority to high-quality works. Using the Bayesian Mean adjusted score (BR_i), already on the normalized scale of 1 to 5, a linear gain of up to 10% is calculated for the best-rated items:

$$B_{rating}(i) = 0.10 \cdot \left(\frac{BR_i}{5.0}\right) \tag{6}$$

The final adjusted similarity score of an item ($S_{adjusted}$) combines the raw vector similarity (S_{vector}) with both factors additively:

$$S_{adjusted} = S_{vector} \times \Big(1.0 + B_{rating}(i) + B_{pop}(i)\Big) \tag{7}$$

3.7. Evaluation Protocol and Thematic Alignment

To enable quantitative evaluation of cross-media recommendation effectiveness, a common semantic vocabulary had to be established. Since works' metadata was collected from different providers (e.g., TMDB for audiovisual, IGDB for games, and MyAnimeList for Japanese animations), genre taxonomies vary substantially in granularity and nomenclature. For example, while cinema commonly uses the broad label "Science Fiction," niche catalogs adopt highly specific markers for the same theme, such as "Mecha" or "Space."

To resolve this divergence and allow mathematical cross-referencing of the data, a macro-genre mapping strategy was developed. This process consisted of grouping granular and specific labels into universal thematic supercategories shared across all media. The detailed mapping of subgenres to defined macro-genres is available in the project repository[1].

With the evaluation space consolidated, latent space effectiveness was measured using the Recall@K metric. The model retrieves the K nearest neighbors of a seed work, and the recommendation is considered semantically coherent (true positive) if the retrieved candidate shares at least one macro-genre with the original seed. This standardization ensures that similarity evaluation is fair and consistent when crossing the different structural boundaries of media formats.

The Recall@10 evaluation reported in this study reflects exclusively the raw vector retrieval step, intentionally omitting the heuristic adjustment phase (Score Boosting)

[1] `https://github.com/Ppaulo03/mylib_recommendation`

described in Section 3.6. This methodological decision isolates and validates the actual capability of the multimodal latent space to semantically connect works from different media, without interference from external variables. In the full system, injecting popularity boosting (B_{pop}) and rating boosting (B_{rating}) alters the ordering based purely on cosine similarity. Although this heuristic reordering may theoretically introduce slight variations in the strict macro-genre cohesion at Top-K, it was designed to optimize retrieval quality from a user-centered perspective. By aligning the popularity level of recommended works with the user's consumption niche, the boosting is proposed as a mechanism to introduce serendipity in recommendations. Empirical validation of this qualitative factor, associated with beyond-accuracy metrics [Sassi 2025], is reserved as a direction for future work focused on practical user acceptance of the system.

4. Results and Discussion

To validate the constructed latent space and the effectiveness of the proposed approach, the system's performance was empirically evaluated through two main metrics: the Silhouette Score, which measures intra-cluster cohesion and inter-cluster separation, and Recall@10, which measures the proportion of retrieved neighbors sharing thematic attributes (macro-genres) with the seed work.

The analysis is divided into two steps. First, an ablation study compares the proposed architecture against purely textual and purely visual vectorization baselines. Then, the model's capability in mitigating filter bubbles through cross-media recommendation is evaluated.

4.1. Experimental Setup and Dataset

The consolidated catalog after pre-processing reached approximately 210,000 works. To conduct the empirical evaluation and ensure methodological viability of quantitative tests without introducing biases, a stratified random sample corresponding to exactly 10% of works from each category was extracted. This strategy ensures that categories with lower production volumes maintain adequate and proportional representation relative to massively dominant formats.

The test dataset totaled 20,993 multimodal embeddings extracted using the SigLIP 2 model. The final distribution of this sample reflects the natural asymmetries of the entertainment industry, consisting of: 16,563 Films, 2,200 Series, 1,526 Games, 601 Anime, and 103 Manga.

4.2. Ablation Study: Intra-Category Evaluation

Table 1 presents the Recall@10 performance broken down by media category.

The analysis reveals that exclusive use of text shows the worst performance across all categories, with particular difficulty in Films (0.3895), suggesting that synopses alone do not adequately capture the semantic diversity of cinematographic subgenres. In contrast, visual (Image Only) and Multimodal representations clearly outperform text, with an average gain of approximately 35%. The **Games** category demonstrated the most coherent semantic representation in the latent space, reaching a Recall@10 of 0.8679 with the Multimodal approach, indicating strong visual and thematic identity for this medium.

Table 1. Intra-Category and Global Recall@10 by Embedding Modality

Category	Text Only	Image Only	Multimodal
Games	0.7320	0.8653	**0.8679**
Series	0.5168	**0.7298**	0.7211
Manga	0.5311	**0.7010**	0.7000
Films	0.3895	**0.6213**	0.6209
Anime	0.4416	0.5828	**0.5908**
Global (Macro-genres)	0.5306	0.7057	**0.7074**

Table 2. *Silhouette Score* by Embedding Modality (Intra-Category and Global)

Category	Text Only	Image Only	Multimodal
Games	-0.0402	-0.0407	**-0.0385**
Series	-0.4774	**-0.0503**	-0.0677
Manga	-0.3023	-0.2967	**-0.2908**
Films	-0.2349	-0.0160	**-0.0128**
Anime	-0.1415	-0.1314	**-0.1169**
Global (Macro-genres)	-0.0935	**-0.0138**	-0.0169

Beyond Recall@10, Table 2 details the Silhouette Score results, which evaluate intra-cluster cohesion and inter-cluster separation in the latent space.

Since genres and narrative archetypes are multi-label and organically shared across all media, natural overlaps in the continuous space are expected, reflected in predominantly negative Silhouette values. Table 2 shows that the Multimodal and Image Only approaches demonstrated a superior ability to group similar items compared to textual representation, reducing dispersion. The Films category stands out: clustering improved from -0.2349 (Text) to -0.0128 (Multimodal), approaching ideal cohesion.

The macro-genre grouping yielded a global Recall@10 of 0.7074 for the Multimodal model, compared to 0.5306 for the Text model. The global Silhouette Score of -0.0169 confirms that, despite cross-media overlap, thematic mapping is maintained. Figure 1 illustrates the continuous clustering and distribution of media in the multimodal latent space, projected using t-SNE [van der Maaten and Hinton 2008].

Although the Image Only approach achieved highly competitive intra-category performance, its isolated adoption is insufficient for the main scope of this work. The limitation of purely visual models emerges when attempting to cross different media formats, a challenge explored and quantified in Section 4.3.

4.3. Cross-Media Recommendation Effectiveness

Traditional algorithms tend to reinforce users' consumption with redundant content within the same structural silo (e.g., recommending only action films to action film viewers), restricting the informational horizon [Pariser 2011]. This work argues that content-based cross-media recommendation offers a path toward expanding these horizons.

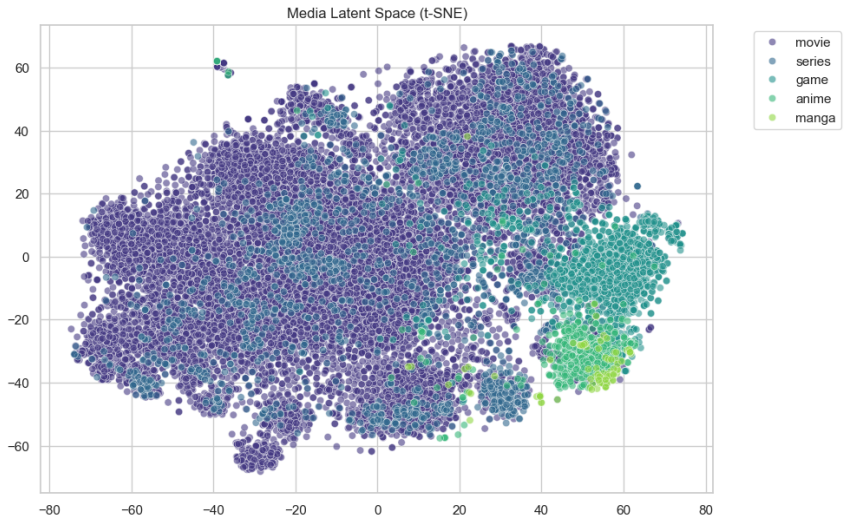


Figure 1. t-SNE projection of the multimodal latent space, showing the continuous and organic mapping across different media categories.

Table 3. Inter-Category Recall@10 Averages by Modality

Modality	Same-Type Mean (Diagonal)	Cross-Type Mean (Off-diagonal)
Text Only	0.70	0.63
Image Only	**0.80**	0.69
Multimodal	0.79	**0.75**

The organic transition between formats, such as recommending a game based on a film, introduces media diversification, while the multimodal latent space ensures that the thematic essence is preserved, securing the relevance of the suggestion. To measure the technical effectiveness of this semantic preservation across media boundaries, we evaluated the Inter-Category Recall@10, which analyzes the degree of genre overlap when the search crosses distinct formats.

The evaluation is divided under two perspectives: the granularity of connections in the similarity matrix (Figure 2) and the comparison of aggregated cross-media retrieval averages (Table 3 and Figure 3).

As shown in Figure 2, the Multimodal model stood out not only in intra-category performance (main diagonal with mean 0.78), but also in its cross-media discovery capability.

Correspondences in the matrix include the strong semantic affinity between **Manga and Games**, which maintained high cross Recall@10 scores. This result indicates that the latent representation connects these media in a highly cohesive way.

To quantify the improvement brought by our approach in format crossing, Table 3 compares retrieval averages for same-type items (matrix diagonal) with the cross-media average for different-type items (off-diagonal).

The model's global effectiveness in crossing domains is evidenced in Table 3 and Figure 3. The model achieved an **off-diagonal (cross-type) mean of 0.7545**, significantly

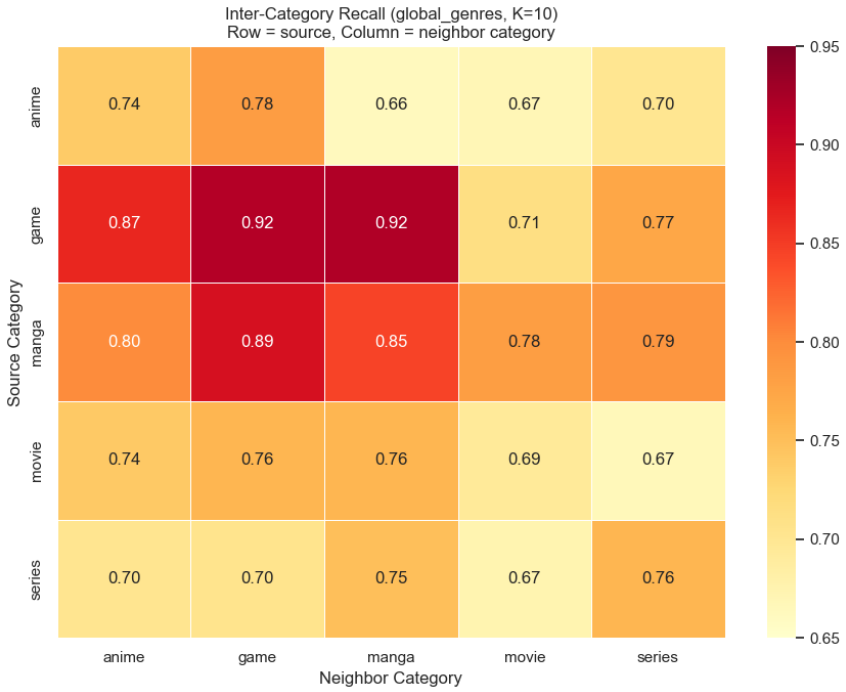


Figure 2. Inter-Category Recall@10 heatmap ($K = 10$) using Multimodal embeddings. The Y axis represents the source media and the X axis the retrieved neighboring media.

above the 0.62 obtained by the purely textual approach.

Although the Image Only approach has a slight edge in intra-category retrieval, purely visual models tend to group works by aesthetic similarities alone (e.g., color palettes or animation styles). They consequently fail when connecting visually distinct but thematically identical media, such as a book cover and a live-action film poster. The Multimodal approach addresses this gap: by fusing synopsis semantics with visual features via SigLIP 2, the model trades a marginal advantage in same-format recommendation for a meaningful gain in cross-media capability.

The chart shows that the average cross-media Recall@10 (dashed line) remains high and competitive relative to same-type item recommendations (blue bars). These results empirically confirm that fusing visual and textual signals via SigLIP 2 is the most effective mechanism for capturing thematic similarities across media boundaries, enabling the cross-format connections proposed in this work.

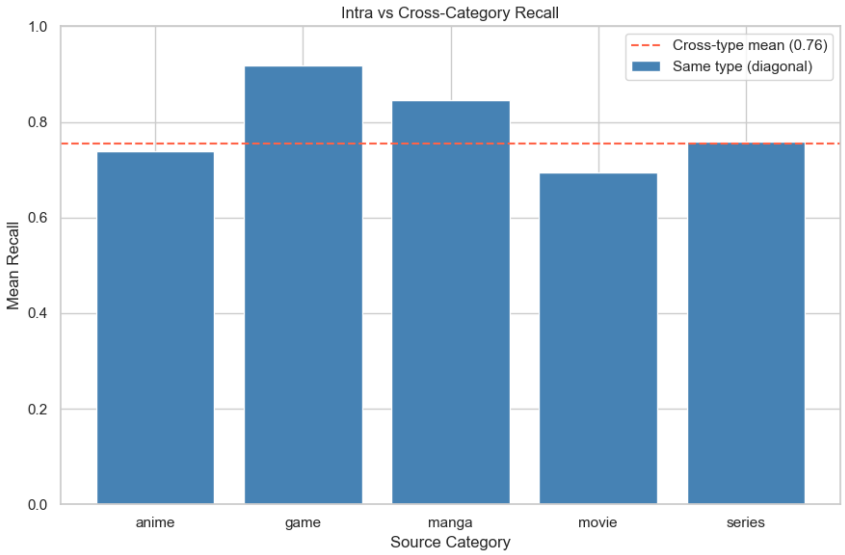


Figure 3. Comparison of intra-category performance versus the cross-type mean (dashed line).

5. Conclusion

This work presented and evaluated a unified content-based cross-media recommendation model. By integrating multimodal representations generated by SigLIP 2 into a Filtered Vector Search infrastructure and a dynamic post-processing mechanism (Score Boosting), the approach demonstrated the ability to infer semantic similarities in a latent space shared by films, series, games, anime, and manga.

The empirical evaluation of 20,993 vectors confirmed the superiority of the multimodal approach over purely textual encoding, which failed to capture the aesthetic and narrative nuances of various works. The proposed model achieved a cross-type (off-diagonal) average Recall@10 of 0.7545, surpassing the text baseline (0.62). This result demonstrates the system's effectiveness in crossing format boundaries to connect semantically relevant items, as illustrated by the strong affinity detected between domains such as Manga and Games.

These results suggest that shared multimodal representations constitute a promising foundation for recommendation systems aimed at exploring content beyond the traditional limits of each platform or format. Cross-media recommendation can thus serve as a fundamental component in future strategies to broaden discovery diversity and reduce the effects of excessive specialization in recommendation mechanisms.

Future work includes transitioning to hybrid models via collaborative filtering and fine-tuning the multimodal encoder for the semantically dispersed Films category. To address current limitations, upcoming evaluations will feature statistical significance tests, comparative analysis against state-of-the-art (SOTA) baselines, and an isolated impact analysis of the score boosting mechanism. Finally, controlled user studies utilizing beyond-accuracy metrics (serendipity, diversity) will validate the practical mitigation of filter bubbles.

6. Acknowledgments

We would like to thank SENAI Fatesg College for the institutional support and laboratory infrastructure that enabled the development of this prototype. We also extend our gratitude to the faculty advisors for their academic and technical guidance, which informed the architectural decisions made throughout this project.

References

Barkan, O. and Koenigstein, N. (2017). Item2vec: Neural item embedding for collaborative filtering.

Chronis, Y., Caminal, H., Papakonstantinou, Y., Özcan, F., and Ailamaki, A. (2025). Filtered vector search: State-of-the-art and research opportunities. *Proceedings of the VLDB Endowment*, 18(12):5488–5492.

Fernández-Tobías, I., Cantador, I., Kaminskas, M., and Ricci, F. (2012). Cross-domain recommender systems: A survey of the state of the art. *Proceedings of the 2nd Spanish Conference on Information Retrieval*.

Hou, Y., Mu, S., Zhao, W. X., Li, Y., Ding, B., and Wen, J.-R. (2022). Towards universal sequence representation learning for recommender systems.

Malkov, Y. A. and Yashunin, D. A. (2016). Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *CoRR*, abs/1603.09320.

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press, New York.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *CoRR*, abs/1908.10084.

Sassi, I. B. (2025). Scientiarec: a scientific article recommendation system with llm-driven feature extraction. In *Proceedings of the 2nd Workshop on Evaluating and Applying Recommendation Systems with Large Language Models (EARL 2025)*, Prague, Czech Republic. Jointly held with the 19th ACM Conference on Recommender Systems (RecSys 2025).

Sriratanawilai, S., Thongwat, J., Chumpu, R., Payoungkhamdee, P., Nutanong, S., and Limkonchotiwat, P. (2025). Distilling multilingual vision-language models: When smaller models stay multilingual.

Tasenţe, T. (2025). Understanding the dynamics of filter bubbles in social media communication: A literature review. *Vivat Academia*, pages 1–21.

Tschannen, M., Gritsenko, A., Wang, X., Naeem, M. F., Alabdulmohsin, I., Parthasarathy, N., Evans, T., Beyer, L., Xia, Y., Mustafa, B., Hénaff, O., Harmsen, J., Steiner, A., and Zhai, X. (2025). Siglip 2: Multilingual vision-language encoders with improved semantic understanding, localization, and dense features.

van der Maaten, L. and Hinton, G. (2008). Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 9(nov):2579–2605. Pagination: 27.

Vechtomova, O. (2009). Introduction to information retrieval christopher d. manning, prabhakar raghavan, and hinrich schütze (stanford university, yahoo! research, and

university of stuttgart) cambridge: Cambridge university press, 2008, xxi+482 pp; hardbound, isbn 978-0-521-86571-5. *Computational Linguistics*, 35:307–309.

Wu, S., Conneau, A., Li, H., Zettlemoyer, L., and Stoyanov, V. (2020). Emerging cross-lingual structure in pretrained language models.

Zhang, Y., Xie, L., Qiu, R., Liu, J., and Wang, S. (2025). Beyond static llm policies: Imitation-enhanced reinforcement learning for recommendation.