Towards Self-Evolving Knowledge Systems: Enhancing Multimodal Agentic RAG with Hyperbolic Flows

Tendai Mukande
Insight Research Ireland Centre for Data Analytics
Dublin City University
Dublin, Ireland
tendai.mukande2@mail.dcu.ie

Noel O'Connor
Insight Research Ireland Centre for Data Analytics
Dublin City University
Dublin, Ireland
Noel.OConnor@dcu.ie

Abstract

Retrieval-Augmented Generation (RAG) has become a foundational paradigm for integrating AI agents with external knowledge. However, current RAG models remain largely constrained by static retrieval pipelines and limited capacity for adaptive reasoning over hierarchical knowledge structures. As AI agents increasingly operate in dynamic, information-rich environments, there is a growing need for models that can reason across modalities while continuously evolving their knowledge representations. We introduce **HFlow,** a self-evolving multimodal agentic RAG framework grounded in hypergraph representations and hyperbolic flow-based reasoning. In our formulation, heterogeneous modalities, including text, images, audio, and structured data, are modelled as nodes within a multimodal hypergraph, while hyperedges capture higher-order semantic and cross-modal relationships. By embedding this structure in hyperbolic space, the framework preserves hierarchical and compositional knowledge. This work proposes a shift from static retrieval to continuous knowledge navigation, where reasoning emerges through geometry-aware traversal of multimodal knowledge manifolds. The proposed framework unifies retrieval, reasoning, and adaptation within a single agentic architecture, offering a new direction for scalable, context-aware AI models. We discuss early empirical evidence demonstrating improved robustness and reasoning flexibility compared to conventional Euclidean and pipeline-based multimodal RAG approaches and outline future opportunities for self-improving knowledge agents.

CCS Concepts

• **Information systems** → **Search interfaces**; **Novelty in information retrieval**.

Keywords

Retrieval Augmented Generation; Hyperbolic Flows; Self-Evolving Models, Multimodal Recommender Systems.

ACM Reference Format:
Tendai Mukande and Noel O'Connor. 2026. Towards Self-Evolving Knowledge Systems: Enhancing Multimodal Agentic RAG with Hyperbolic Flows. In *International Conference on Multimedia Retrieval (ICMR '26), June 16–19, 2026, Amsterdam, Netherlands.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3805622.3810427

ACM ISBN 979-8-4007-2617-0/2026/06
https://doi.org/10.1145/3805622.3810427

1 Introduction

Retrieval-Augmented Generation (RAG) has emerged as a paradigm for improving AI agents by incorporating external knowledge into generative models [25]. RAG improves the accuracy, interpretability, and adaptability of facts [29]. However, conventional RAG models remain limited in complex real-world scenarios [5]. In particular, existing approaches typically rely on static embeddings, struggle to integrate heterogeneous multimodal information, and fail to capture hierarchical or higher-order relationships between knowledge entities [25]. These limitations restrict their ability to support the adaptive, goal-directed reasoning required in real-world scenarios [27]. The challenges are more pronounced in multimedia retrieval and agentic decision-making tasks such as multimodal planning and knowledge-guided reasoning [32]. Traditional multimedia retrieval frameworks are often designed for single modalities or pairwise relationships and therefore cannot fully model contextual dependencies across heterogeneous information channels [38]. As a result, most existing RAG models are limited when handling tasks that involve complex dynamic reasoning as new evidence or context becomes available [24].

Euclidean embedding models struggle to preserve hierarchical structure and semantic generalisation, as distances grow linearly and cannot efficiently represent exponentially expanding tree-like relationships [17]. Beyond geometric limitations, most existing RAG models often require retraining or fine-tuning to incorporate updated representations, limiting adaptability in dynamic environments [35]. For tasks such as *goal-directed planning*, static embedding models impose limitations on multi-step reasoning [32]. Effective planning requires modelling trajectories through structured state spaces, where actions induce transitions, and intermediate decisions dynamically reshape subsequent retrieval and inference [3]. This process is inherently sequential and relational: each step conditions the next, forming a dependency chain over evolving states rather than isolated queries in a fixed embedding space [27]. As a result, the model becomes limited in coherent reasoning, ensuring consistency between intermediate decisions, or adapting retrieval strategies as the plan unfolds [10, 25]. Consequently, trajectory-aware decision-making and structured multi-step planning remain under-supported in static embedding frameworks [16].

To address these issues, we propose the **HFlow** model as a new paradigm for agentic RAG models. Specifically, we integrate multimodal hypergraphs, hyperbolic embeddings, and flow-based reasoning. For multimodal reasoning, we aggregate text, image, audio, and structured knowledge via higher-order hyperedges with tangent-space attention. We hypothesise that this design allows agents to retrieve, reason, and generate knowledge within a coherent geometric representation, supporting contextual awareness, adaptability, and scalable multimodal integration.

Contributions

We propose a framework that integrates multimodal hyperbolic representation learning with flow-based embedding evolution. The framework provides the following key contributions:

- **Multimodal Representation Learning:** Knowledge is structured as a multimodal hypergraph, where nodes represent heterogeneous entities such as text, images, audio, and structured data. Hyperedges model higher-order relationships, capturing complex contextual interactions that span multiple modalities simultaneously.
- **Hyperbolic Embeddings for Hierarchical Semantics:** Node embeddings are learned by the model in hyperbolic space, which supports hierarchical and tree-like semantic relationships [17]. We hypothesise that this enables compact, expressive representations while improving generalisation across related concepts.
- **Flow-Based Agentic Reasoning:** A hyperbolic flow vector field governs the evolution of agent embeddings across the hypergraph. This allows agents to perform multi-step, goal-directed reasoning while dynamically adapting as new information is retrieved or inferred.
- **Self-Evolving Agent Behaviour:** The framework supports continuous learning by allowing agents to incorporate new nodes and hyperedges without full model retraining. This enables adaptive and scalable knowledge expansion.

The framework is positioned in a new paradigm for agentic Retrieval-Augmented Generation (RAG). Agents retrieve, reason, and generate knowledge within an end-to-end geometric representation that supports contextual awareness, adaptability, and scalable multimodal integration. This design enables the development of **self-evolving AI agents** capable of continuous learning, cross-modal reasoning, hierarchical knowledge modelling, and goal-directed planning.

2 Related Work

Recent advances in multimodal representation learning, hypergraph neural networks, and retrieval-augmented generation (RAG) have significantly improved the modelling and integration of heterogeneous information sources [10]. Multimodal foundation models, such as CLIP [26] and ALIGN [20], have shown potential in cross-modal alignment by learning shared embedding spaces in vision and language models [15]. Graph-based RAG frameworks capture higher-order relationships among entities beyond pairwise interactions, enabling richer relational modelling. In parallel, models such as RAG [21] and REALM [15] improve reasoning by incorporating external knowledge during inference.

Despite these advances, several limitations remain. Multimodal foundation models mainly optimise for representation alignment rather than structured reasoning. Although they embed heterogeneous modalities in a shared latent space, they do not explicitly model higher-order relational dependencies or evolving knowledge structures [35]. Their reasoning capabilities remain implicit within dense vector representations and lack explicit mechanisms for relational composition or iterative refinement [9]. Graph-based methods partially address this limitation by modelling relational data; however, most approaches operate in Euclidean space, which constrains their ability to efficiently represent hierarchical or tree-like knowledge structures [25]. Hyperbolic models have been proposed to address this issue by enabling more compact representations of hierarchical structures. However, they still face limitations in effectively integrating and interacting with external knowledge sources [17]. Moreover, retrieval is typically driven by a single query and lacks iterative feedback loops, preventing the system from refining its reasoning strategy based on intermediate results [25].

To address these limitations, we propose a self-evolving agentic RAG framework that unifies geometric relational modelling, adaptive retrieval, and iterative reasoning. In this framework, knowledge is organised as a dynamic hypergraph rather than a collection of independent documents, allowing more effective representation of higher-order and cross-modal relationships. Embedding this hypergraph in an appropriate geometric space, such as a hyperbolic space for hierarchical structures, enables efficient modelling of compositional and tree-like knowledge. Retrieval operates on structured subgraphs, supporting relational reasoning rather than context accumulation [35]. We propose a RAG framework that transforms retrieval from a static augmentation mechanism into an adaptive self-improving process. Beyond passive retrieval, the model evolves through interaction, relational weights are adjusted according to task success, and knowledge is reorganised hierarchically to improve future reasoning efficiency.

3 Proposed Framework

Our proposed framework integrates multimodal representation learning and flow-based reasoning to enable adaptive model reasoning. We hypothesise that this design supports a geometry-aware, self-evolving agentic retrieval-augmented generation (RAG) framework, enhancing scalability, reasoning fidelity, and adaptability for complex multimodal tasks such as planning and contextual knowledge synthesis.

3.1 Hyperbolic Embedding Space

To preserve hierarchical and tree-like relational structures, nodes are embedded into a d-dimensional hyperbolic manifold $(\mathbb{H}^d, g)$:

$$x_v \in \mathbb{H}^d, \quad \forall v \in \mathcal{V}, \tag{1}$$

where g denotes the Riemannian metric tensor [17]. By embedding nodes on a hyperbolic manifold, the model organises high-level conceptual representations closer to the origin of the manifold while placing modality-specific or fine-grained entities near the boundary [13]. Using the Poincaré ball model [11], the geodesic distance between nodes u and v is defined as follows:

$$d_g(x_u, x_v) = \operatorname{arcosh}\left(1 + 2\frac{\|x_u - x_v\|^2}{(1 - \|x_u\|^2)(1 - \|x_v\|^2)}\right). \tag{2}$$

Hyperbolic geometry enables exponential volume growth, allowing efficient representation of hierarchical knowledge and multi-level semantic dependencies [24].

3.2 Tangent-Space Hyperedge Aggregation

For a node $x \in \mathbb{H}^d$, let $\mathcal{E}(x)$ denote the set of incident hyperedges. Since aggregation operations are not directly defined in curved manifolds, we project neighbouring embeddings into the tangent space $T_x\mathbb{H}^d$ using the logarithmic map [23]. For each hyperedge $e \in \mathcal{E}(x)$, the multimodal contextual representation is defined as follows:

$$h_e = \mathrm{Agg}\left(\{\log_x(x_j) \mid x_j \in e \setminus \{x\}\}\right), \tag{3}$$

where $\log_x(\cdot)$ denotes the logarithmic map and $\mathrm{Agg}(\cdot)$ is a permutation-invariant operator such as mean, sum, or attention pooling. To model contextual relevance, hyperedge importance weights are computed as follows:

$$\alpha_e = \frac{\exp\left(-\frac{1}{\tau}\|h_e\|_g\right)}{\sum_{e' \in \mathcal{E}(x)} \exp\left(-\frac{1}{\tau}\|h_{e'}\|_g\right)}, \tag{4}$$

where τ is a temperature parameter that controls the sharpness of the attention distribution.

3.3 Flow-Based Agentic Embedding Evolution

Agent behaviour is modelled as a time-dependent vector field defined on the tangent bundle of the hyperbolic manifold.

$$v_\theta(x, s, t) = \sum_{e \in \mathcal{E}(x)} \alpha_e \, W_\theta(h_e, b_e, s, t), \tag{5}$$

where b_e denotes the behavioural semantics associated with hyperedge e, and W_θ is a learnable mapping producing tangent vectors in $T_x\mathbb{H}^d$. The embedding evolution from time s to t is defined via the exponential map [23]:

$$X_{s,t}(x_s) = \exp_{x_s}\left((t - s)\, v_\theta(x_s, s, t)\right). \tag{6}$$

We hypothesise that this formulation enables multi-step reasoning by modelling embedding trajectories as goal-directed flows over hierarchical relational manifolds.

3.4 Cross-Modal Stochastic Interpolation

To model exploration, uncertainty, and knowledge expansion, we introduce stochastic interpolation along geodesic trajectories:

$$I_t(x_0, x_1) = \exp_{x_0}\left(\alpha_t \log_{x_0}(x_1)\right), \quad \alpha_t \in [0, 1]. \tag{7}$$

The corresponding velocity field is:

$$\dot{I}_t = \alpha_t' \frac{\log_{x_t}(x_1)}{1 - \alpha_t}. \tag{8}$$

We propose a mechanism to enable dynamic knowledge integration and facilitate continuous learning without requiring complete model retraining, by adding Gaussian noise [11] in the tangent space, creating a hyperbolic diffusion process along the geodesic:

$$I_t^{\text{noisy}} = \exp_{\mathbf{x}_0}\left(\alpha_t \, \log_{\mathbf{x}_0}(\mathbf{x}_1) + \boldsymbol{\epsilon}_t\right), \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \tag{9}$$

3.5 Multimodal Retrieval and Generation

After agentic flow evolution, the updated embedding is computed as follows:

$$\hat{x}_u = X_{0,1}(x_u). \tag{10}$$

Candidate items are ranked using hyperbolic similarity scoring:

$$\mathrm{score}(u, i) = -d_g(\hat{x}_u, x_i). \tag{11}$$

This distance-aware retrieval strategy jointly captures the alignment of modalities, hierarchical relationships, and contextual relevance.

3.6 Scalability, Expressivity, and Efficiency Analysis

Beyond computational efficiency, our proposed framework is designed to balance *representational expressivity*, *adaptive reasoning*, and *scalability*. We analyse these aspects jointly. Let $N = |\mathcal{V}|$ denote the number of nodes, $E = |\mathcal{E}|$ the number of hyperedges, $\bar{k}$ the average cardinality of the hyperedge, d the embedding dimension and T the number of flow evolution steps.

$$\text{Memory Complexity} = O(dN + d\bar{k}E), \tag{12}$$

$$\text{Flow Complexity} = O(Td\bar{k}E). \tag{13}$$

Compared to pairwise graph models ($O(d|E|)$), the additional factor $\bar{k}$ reflects higher-order relational modelling, which is essential to capture cross-modal dependencies but introduces moderate overhead.

Expressivity vs. Euclidean Representations. Unlike Euclidean embeddings, hyperbolic space provides exponential volume growth, enabling compact encoding of hierarchical and long-tail structures. This is beneficial in recommendation settings where item distributions are highly skewed. The gain is not necessarily reflected in aggregate ranking metrics alone, but becomes more pronounced in cold-start scenarios and multi-hop relational reasoning [41].

Flow-Based Reasoning vs. Static Retrieval. Conventional retrieval models (e.g., dual encoders or CLIP-style similarity) perform *single-step alignment.* In contrast, the proposed flow formulation models retrieval as a *trajectory optimisation process* in latent space:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{v}_\theta(\mathbf{x}(t), \mathcal{H}, t). \tag{14}$$

This enables iterative refinement of representations, approximating multi-step reasoning without explicit discrete retrieval loops. Importantly, the current implementation uses a fixed number of steps T, which limits the observable advantage over strong static baselines. Increasing T or introducing adaptive stopping criteria is expected to widen this gap.

Dynamic Update Efficiency. A key advantage of the framework is that new nodes or hyperedges can be incorporated without retraining. The cost of an update is:

$$O(d \cdot \deg(e_{\text{new}})), \tag{15}$$

where $\deg(e_{\text{new}})$ is the size of the new hyperedge. This contrasts with standard RAG or embedding models that require periodic retraining to reflect new knowledge. The framework introduces three main trade-offs: (i) Higher-order modelling (hypergraphs) vs. computational overhead, (ii) Flow-based reasoning vs. increased inference cost (T steps), and (iii) Hyperbolic geometry vs. optimisation complexity. However, these trade-offs are justified in scenarios that require adaptive reasoning, dynamic knowledge integration, and multimodal relational understanding.

4 Discussion

The proposed framework establishes a unified paradigm for next-generation multimedia retrieval, integrating multimodal information and enabling agentic reasoning through flow-based embedding evolution [5, 22, 24]. The model provides a scalable foundation for multimedia retrieval in heterogeneous data ecosystems [38]. In terms of scalability and efficiency, let $N = |\mathcal{V}|$ denote the number of nodes, $E = |\mathcal{E}|$ the number of hyperedges, $\bar{k}$ the average cardinality of the hyperedge, d the embedding dimension, and T the number of flow evolution steps. Memory requirements scale as $O(dN + d\bar{k}E)$, accounting for both node embeddings and hyperedge contextual representations, and the computational cost of agentic flow evolution is $O(Td\bar{k}E)$. By integrating hyperbolic flows [2], stochastic diffusion, and multimodal hypergraphs, the proposed framework establishes an **agentic, adaptive RAG paradigm**. This formulation enables self-evolving agents capable of multi-step reasoning, continuous knowledge integration, and context-aware generation.

5 Discussion

The proposed framework introduces a unified perspective on multimodal retrieval by jointly modelling *geometry*, *structure*, and *reasoning dynamics*. Rather than treating retrieval as a static similarity matching problem, we reformulate it as a *continuous trajectory optimisation process* over a structured latent space. The model is designed to address key limitations of existing approaches in scenarios such as: (i) **dynamic environments**, where new items or events can be incorporated without retraining; (ii) **long-tail recommendation**, where hyperbolic geometry improves generalisation in sparse regions; (iii) **multi-hop reasoning**, where flow-based trajectories approximate iterative retrieval; and (iv) **cross-modal interaction**, where hyperedges capture higher-order relationships beyond pairwise alignment.

Future work will focus on incorporating adaptive flow depth, improving hyperedge construction strategies, and extending evaluation to more challenging settings. In particular, evaluating the model on multi-hop and temporal retrieval benchmarks will provide a more comprehensive assessment of its reasoning and adaptation capabilities.

6 Use Case: Self-Evolving Travel Assistant

To illustrate the applicability of our framework, we consider an AI travel recommendation task that generates personalised city itineraries based on user interests, budget, and time constraints. Contextual information, such as weather updates, newly introduced attractions, or scheduling conflicts, dynamically affects recommendations, allowing the model to maintain a contextually optimised and adaptive travel experience (Figure 1). We position this use case in an emerging class of agent-based recommender systems (RSs) that continually improve through user interactions, environmental feedback, and newly available knowledge [8].

Step 1: User queries "*3-day art + music trip in Paris*" → agent embedding flows toward relevant hyperedges.

Step 2: New attraction added (e.g., "*Musée d'Orsay exhibit*") → new hyperedge added dynamically.

Step 3: The flow vector field adapts to integrate this new node; the agent embedding now attends to the new experience without retraining everything.

Step 4: Future queries evolve with the new knowledge, planning richer itineraries or generating multimodal guides.

Insight: The agent *self-evolves* by growing its hypergraph, updating flow dynamics, and continuously adapting embeddings. Multimodal reasoning is enabled through higher-order hyperedges and tangent-space attention [22], overcoming the restrictions of pairwise embeddings or unimodal attention [33]. Hyperbolic embeddings preserve hierarchical knowledge and semantic generalisation, while flow-based embedding evolution and stochastic interpolation allow adaptive, self-evolving behaviour, and continuous learning without retraining [40]. Agent-directed, goal-directed reasoning is achieved by embedding trajectories guided by hyperbolic flow, which supports multi-step decision-making [12]. Finally, tangent-space aggregation and hyperedge-level attention are applied for scalable cross-modal interaction [36]. This design aims to provide an adaptive and scalable agentic RAG framework capable of multimodal reasoning, allowing dynamic updates and knowledge adaptation [24].

6.1 Model Overview

The proposed self-evolving travel assistant is designed as a **multimodal agentic RAG model** augmented with **hyperbolic flows**, allowing it to adapt to new information and provide contextually rich recommendations. The model integrates three key components: a multimodal hypergraph knowledge representation, hyperbolic embeddings for hierarchical reasoning, and flow-based embedding evolution for agentic, self-evolving behaviour. The architecture (Figure 1) implements an adaptive multimodal agentic RAG pipeline in which reasoning and retrieval are coupled, as shown in Figure 2.

6.1.1 Query Embedding → Hyperbolic Flow Module. The embedded user query initialises the agent state within the hyperbolic latent space. The Hyperbolic Flow Module then evolves this embedding to represent the agent's reasoning trajectory.

6.1.2 Multimodal Hypergraph → Retrieval. This component retrieves knowledge nodes that are relevant both to the user query and to the agent's reasoning trajectory. The final embedding of the agent $\hat{\mathbf{x}}^u$ is used to identify and retrieve nearby nodes in the hypergraph in multiple modalities, including text nodes (descriptions, schedules), image nodes (photos, maps), and audio nodes (music, narration, ambient soundscapes).

6.1.3 Retrieval → Generation. Retrieved data is processed through a generator that provides contextual representations in the tangent space, ensuring coherent and context-aware responses. The generator produces outputs across text, image, and audio modalities, e.g., textual itineraries, annotated maps, and audio samples that depict real-world user experiences.

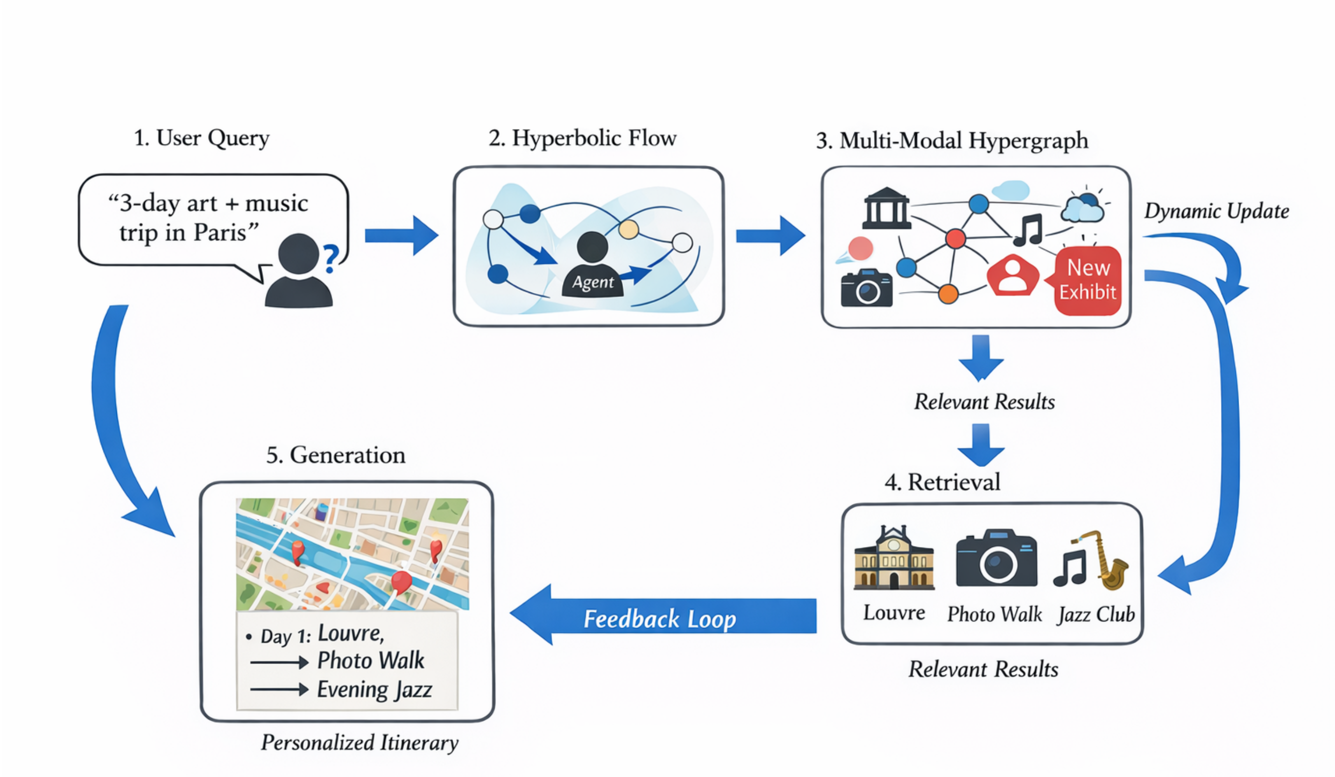

Figure 1: Visualisation of the Agentic RAG framework integrating hyperbolic flow, stochastic diffusion, and multimodal reasoning. The query embedding evolves through deterministic and stochastic hyperbolic flows (Equations shown below modules) to enable multimodal retrieval and generation with iterative feedback.

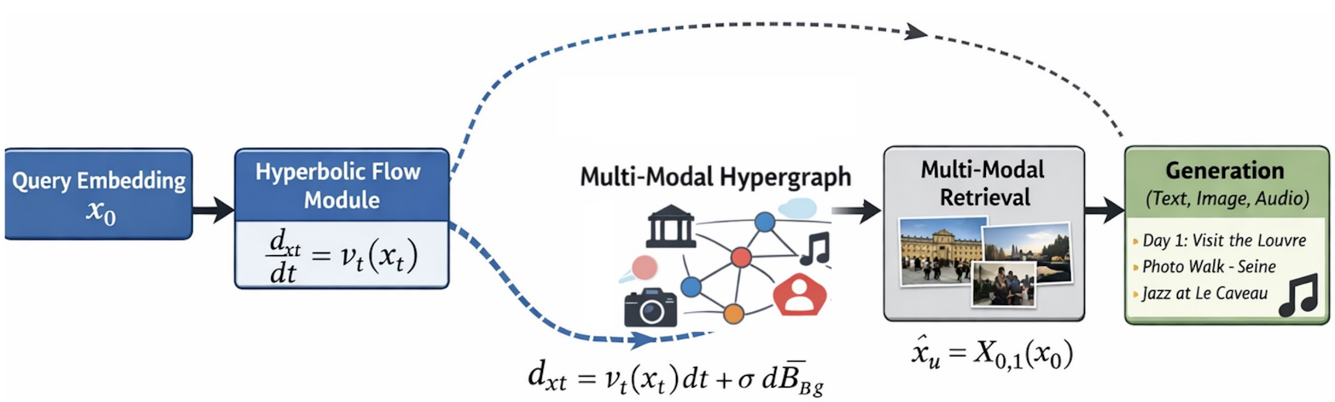

Figure 2: Overview of the proposed HFlow framework. A query embedding x_0 is evolved through a Hyperbolic Flow Module governed by continuous-time dynamics, enabling geometry-aware trajectory modelling in hyperbolic space. The resulting representation interacts with a multimodal Hypergraph that encodes higher-order cross-modal relationships. Structured retrieval supports adaptive reasoning, and the refined representation is used for multimodal generation (text, image, audio).

6.1.4 Agents as Hyperbolic Embeddings. In this framework, an **agent** is a computational entity that represents the reasoning and preference state of the model for a given user query. Each agent is initialised by embedding a user query q into hyperbolic space:

$$\mathbf{z}_q = f_\phi(q), \tag{16}$$

where f_ϕ is a trainable embedding function that maps queries to the hyperbolic latent space. The agent then *traverses* the knowledge hypergraph, attending to relevant nodes and hyperedges corresponding to points of interest, events, and contextual signals. Unlike static embeddings in conventional RAG models, these agents are dynamic trajectories in hyperbolic space. They can *self-evolve* as the hypergraph is updated with new nodes or hyperedges, such as newly introduced attractions or events, without requiring retraining.

6.1.5 Agentic Retrieval and Generation. After evolving through the hyperbolic flow, the agent embedding $\hat{\mathbf{x}}^u$ retrieves the relevant k top nodes from the hypergraph. Retrieved nodes are then fed into a generation module, producing multimodal outputs such as textual itineraries (day-by-day plans), annotated maps highlighting POIs, images of landmarks or events and audio samples representing music or ambient soundscapes. Agent behaviour is modelled as a time-dependent hyperbolic flow vector field [37]. For example, if a user submits the query q = *"3-day art + music trip in Paris"*, the agent embedding $\mathbf{z}_q$ propagates through $\mathcal{H}$, activating relevant hyperedges representing locations, events, and cultural experiences. When a new attraction becomes available (e.g., a *"Musée d'Orsay exhibit"*), a corresponding hyperedge e_{new} is added dynamically. The hyperbolic flow vector field automatically incorporates the new node, allowing $\mathbf{z}_q$ to attend to the attraction without retraining. Future queries leverage the expanded hypergraph $\mathcal{H}'$ and updated flow $\mathbf{v}'_\theta$ to generate richer itineraries and comprehensive travel guidance.

6.2 Experimental Setup

To enable self-evolving, multimodal recommendations, we construct a multimodal knowledge space by integrating multiple publicly available datasets that cover Point-Of-Interest (POI), user behaviour, events, and contextual information in Paris. TripAdvisor[1] and Yelp[2] provide POI information, including textual descriptions, user reviews, ratings, and images. These datasets form the backbone of nodes in the hypergraph. *Foursquare*[3] and *Gowalla*[4] *check-in* datasets were used to model user behaviour and preferences, providing user nodes and user-item interactions for personalised itineraries. Live events, such as concerts, exhibitions, and festivals, were integrated from Paris Open Data[5] and Eventbrite[6], serving as dynamic updates to the knowledge space. Structured geographic and accessibility information for POIs was obtained from *OpenStreetMap*[7], enabling location-based recommendations.

6.2.1 Encoding Multimodal Data. Each modality is encoded into a fixed-length vector embedding to serve as a node in the hypergraph: **Textual data:** Reviews, descriptions, and event information are encoded using **Sentence Transformers** [28], producing embeddings $\mathbf{z}^{\text{text}} \in \mathbb{R}^d$. **Images:** POI or event images are encoded using **CLIP** [26], yielding embeddings $\mathbf{z}^{\text{image}} \in \mathbb{R}^{d_v}$. **Audio:** Ambient sounds or music clips are encoded using **OpenL3** [7], producing embeddings $\mathbf{z}^{\text{audio}} \in \mathbb{R}^{d_a}$. **Videos:** Video clips are encoded by extracting frame-level embeddings with CLIP and aggregating temporally, combined with audio embeddings from OpenL3. The resulting embedding $\mathbf{z}^{\text{video}} \in \mathbb{R}^{d_v}$ captures visual, temporal, and auditory information. **Context:** Structured features such as time of day, day of week, weather, and accessibility are projected into embeddings $\mathbf{z}^{\text{context}} \in \mathbb{R}^{d_c}$.

6.2.2 Agentic Capability Mapping. Table 1 maps each metric to the corresponding capability of the HFlow model. Mean Reciprocal Rank (MRR) reflects the agent's goal-directed reasoning by measuring how quickly it retrieves the most relevant item. Normalised Discounted Cumulative Gain (NDCG) captures multi-step, structured reasoning by evaluating the quality and ordering of top-ranked items. Hit Ratio (HR) quantifies exploration and coverage, indicating whether the agent retrieves multiple relevant items per query. Novelty measures self-evolving adaptation, showing the agent's ability to recommend newly added or unseen items without retraining [8]. Finally, the improvements in all metrics collectively reflect the agent's ability to integrate information across multiple modalities.

Ground Truth

The **ground truth** represents the set of relevant items for a given user query. Formally, for a query q:

$$G_q = \{i_1, i_2, \ldots, i_m\}, \tag{17}$$

where each i_j is a relevant point of interest (POI), event, or media item that aligns with user preferences and contextual constraints. Ground truth items are determined from multiple complementary sources to ensure relevance and context alignment. These include existing datasets, such as POIs, user reviews, events, and geographic information obtained from TripAdvisor, *Yelp, Foursquare, Gowalla, OpenStreetMap, Paris Open Data*, and *Eventbrite*. User preferences, reflected in historical *check-ins*, ratings, and interactions, provide additional signals of relevance for each individual. Contextual factors, such as travel dates, time, budget, weather, and accessibility, are also considered to filter and prioritise relevant items.

To evaluate our model, we distinguish between two types of ground truth. The static ground truth consists of items known during the training phase, and metrics computed using this set assess traditional recommendation performance, such as quality and coverage. Dynamic ground truth, on the other hand, includes newly added POIs, events, or media in the updated hypergraph. The metrics computed using this set, such as novelty, capture the agent's ability to adapt to new information and evaluate the model's ability to capture self-evolving scenarios.

$$\mathcal{H} = (\mathcal{V}, \mathcal{E}), \quad \mathcal{V} = \mathcal{U} \cup \mathcal{I} \cup \mathcal{C} \cup \mathcal{M}, \tag{18}$$

where $\mathcal{U}$ are user nodes, $\mathcal{I}$ are item or content nodes (POIs, events), $\mathcal{C}$ are contextual nodes (time, weather, accessibility).

$$\mathcal{M} = \{\mathbf{z}^{\text{text}}, \mathbf{z}^{\text{image}}, \mathbf{z}^{\text{audio}}, \mathbf{z}^{\text{video}}\} \tag{19}$$

[1]https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews
[2]https://www.yelp.com/dataset
[3]https://sites.google.com/site/yangdingqi/home/foursquare-dataset
[4]https://snap.stanford.edu/data/loc-gowalla.html
[5]https://opendata.paris.fr
[6]https://www.eventbrite.com/platform/api
[7]https://www.openstreetmap.org

Table 1: Mapping of Metrics to Agentic Capabilities

Benefit	Metric(s)	Quantification
Goal-directed reasoning	MRR	Rank of first relevant node
Structured (multi-step) reasoning	NDCG	Discounted relevance of top-k ordered items
Exploration/coverage	HR	Fraction of queries with more than 1 relevant node in top-k
Self-evolving adaptation	Novelty (%)	Fraction of recommended items unseen during training
multimodal integration	All metrics	Improvements indicate cross-modal reasoning

are modality-specific nodes. Hyperedges $e \in \mathcal{E}$ connect related nodes across modalities to determine higher-order semantic relationships. For example, a hyperedge connects:

$$e = \{\text{POI text, POI image, POI audio, POI video, context}\}. \quad (20)$$

.

Hyperedges $e \in \mathcal{E}$ connect arbitrary subsets of nodes, capturing higher-order semantic relationships between modalities and context. During agentic reasoning, embeddings attend to relevant hyperedges using tangent-space aggregation to enable integration of multimodal information.

6.2.3 Agentic Integration via Hyperbolic Flows. Agent embeddings propagate through the hypergraph according to a hyperbolic flow vector field and integrate multimodal information in a unified formulation:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{v}_\theta(\mathbf{x}(t), \mathcal{H}, t), \quad \hat{\mathbf{x}}^u = X_{0,1}(\mathbf{x}_0), \quad (21)$$

where $X_{0,1}(\cdot)$ denotes the flow map from $t = 0$ to $t = 1$, and $\mathbf{v}_\theta$ identifies the relevant hyperedges connecting multimodal nodes. This formulation ensures that the text, image, audio, video, and context nodes are fully integrated into the agent's reasoning trajectory.

6.2.4 Dynamic Updates and Self-Evolving behaviour. New POIs, events, or videos can be dynamically added as nodes or hyperedges to the hypergraph. Hyperbolic flows automatically incorporate these new nodes, allowing agent embeddings to adapt in real-time. Queries such as "3-day art and music trip in Paris" propagate through the hyperbolic flow module, which evolves the embeddings according to Equation (21). The evolved embeddings attend over multimodal hyperedges and are used to retrieve the top-k most relevant nodes.

6.2.5 Self-Evolving Dynamics. The HFlow model supports self-evolving behaviour. (1) New nodes or hyperedges (e.g., newly opened attractions or live events) are dynamically added to the hypergraph. (2) Hyperbolic flows automatically integrate the new information into existing embeddings. (3) Future queries leverage the expanded knowledge, producing richer, updated itineraries without retraining. Each agent acts as a *dynamic knowledge traverser*, continuously reasoning over a multimodal, hierarchical, and self-evolving knowledge space. The combination of hypergraph structure, hyperbolic embeddings, and flow-guided evolution enables both reasoning and *lifelong adaptation*, fulfilling the criteria of a self-evolving agentic RAG model. Table 2 presents the performance of HFlow compared with baseline representation learning models.

The results demonstrate that the self-evolving hyperbolic flow-based RS outperforms baseline models on all metrics. Improvements in MRR and NDCG@5 indicate that relevant items are ranked higher and overall ranking quality is improved. HR@5 shows that a greater proportion of relevant items appear in the top recommendations, reflecting better retrieval coverage. Importantly, the novelty metric is substantially higher for the HFLow model, demonstrating its ability to recommend newly introduced attractions and events without retraining. These results confirm that the combination of hyperbolic flow and a dynamic multimodal hypergraph supports self-evolving travel recommendations.

Table 2: Performance Comparison of Travel Recommendation Models on Real-World Dataset (Paris POIs)

Model	MRR	NDCG@5	HR@5	Novelty (%)
MBHT [34]	0.37	0.39	0.48	8
PBAT [30]	0.41	0.44	0.52	6
HyperLLM [6]	0.52	0.56	0.61	22
HRDM [39]	0.39	0.48	0.57	17
ALIGN [20]	0.42	0.45	0.60	12
CLIP [26]	0.53	0.57	0.68	20
HFlow	**0.65**	**0.68**	**0.80**	**48**

6.3 Ablation Study

To evaluate the contribution of each component to the self-evolving hyperbolic flow-based travel model, we conduct an ablation study. Specifically, we evaluate three variants of the model: **w/o Flow:** We remove the hyperbolic module from the model framework. **w/o Hypergraph:** The hypergraph structure is removed, and items are connected only via pairwise embeddings; flow is retained. **w/o Updates:** We simulate a static knowledge base, without adding new information to the model. The performance of these variants was evaluated on the same metrics as the full model, as summarised in Table 3.

Table 3: Ablation Study of Self-Evolving Hyperbolic Flow RS on Paris POIs

Model Variant	MRR	NDCG@5	HR@5	Novelty (%)
w/o Flow	0.57	0.60	0.72	40
w/o Hypergraph	0.50	0.53	0.65	30
w/o Updates	0.63	0.65	0.77	18
HFlow	**0.65**	**0.68**	**0.80**	**48**

The results validate the contribution of each component of the model. For example, removing the hyperbolic flow module reduces the ranking quality as indicated by the degradation of the MRR and NDCG results. Excluding the multimodal hypergraph substantially degrades both relevance and novelty, indicating that higher-order connections across modalities are critical for rich recommendations. When dynamic updates are disabled, novelty drops drastically, confirming that the model's ability to recommend newly introduced attractions relies on the self-evolving hypergraph. The ablation study demonstrates that each component—hyperbolic flow, multimodal hypergraph, and dynamic updates—contributes meaningfully to the modelling of self-evolving scenarios.

7 Conclusion

We introduce **Hflow,** a novel conceptual framework for agentic multimedia retrieval that integrates multimodal hypergraph representations with hyperbolic flow-based embedding evolution. The proposed architecture rethinks retrieval-augmented generation by enabling agents to reason over higher-order multimodal relationships while preserving hierarchical knowledge structures through curvature-aware embeddings.We envision that this approach enables a new generation of multimedia retrieval agents that operate on structured, continuously evolving multimodal memory spaces rather than static modality-specific indexes. The proposed framework represents a promising research roadmap that aims to stimulate new theoretical and practical developments in multimedia retrieval. We hope that this work stimulates a shift from retrieval as information access toward retrieval as structured cognitive reasoning over heterogeneous knowledge ecosystems.

8 Appendix

Graph-based RSs typically rely on discrete message passing defined over incidence matrices. In contrast, we formulate the recommendation task as a continuous transport process between preference distributions defined on a Riemannian manifold.

Proposition 1 (Existence and Uniqueness of Preference Transport Flow) [1]. Assume that the vector field $v_\theta(x, s, t)$ is Lipschitz continuous in x and continuous in (s, t). Then the differential equation

$$\frac{dx_t}{dt} = v_\theta(x_t, s, t) \tag{22}$$

admits a unique solution trajectory on $\mathbb{H}^d$ for any initial embedding x_0. This guarantees that preference evolution is well-defined as a continuous transformation on the manifold, compared to discrete, layer-wise propagation.

8.1 Consistency of the Generalised Flow Map

The proposed model directly integrates the global flow operator $X_{s,t}$ instead of numerically integrating the local dynamics.

Proposition 2 (Tangent Space Closure [14]). Let $h_e \in T_x\mathbb{H}^d$. If $W_\theta(h_e, b_e, s, t) \in T_x\mathbb{H}^d$, then:

$$v_\theta(x, s, t) \in T_x\mathbb{H}^d. \tag{23}$$

This guarantees that the exponential map produces valid manifold embeddings [19], preserving geometric consistency during training and inference. Stochastic interpolation implicitly defines a Monge transport map [31] between ρ_0 and ρ_1. This provides a probabilistic interpretation of the recommendation as a structured distribution alignment across interaction contexts.

8.2 Proof-of-Concept

To evaluate the proposed framework in an agentic RAG environment, consider the travel assistant use case, where queries involve heterogeneous sources including textual requests, images, audio guides, and structured entities such as attractions, events, and schedules. These form higher-order relationships, which we represent as hyperedges in the multimodal hypergraph. In agentic RAG, a single query typically involves heterogeneous information sources, including textual documents, visual content, audio signals, and structured knowledge entities. Hyperbolic space provides a geometric representation for hierarchical and tree-like relational structures commonly observed in multimodal knowledge environments [9]. This framework integrates hyperbolic flows, stochastic diffusion, and multimodal hypergraphs to enable an *agentic, adaptive RAG model.* The design tightly couples reasoning, retrieval, and generation within a continuous hyperbolic latent space.

8.3 Hyperbolic Flow as Geometry-Aware Diffusion

Let $\mathbf{x}_t$ denote the agent embedding at time t. In classical diffusion models, the evolution of embeddings is defined as follows:

$$d\mathbf{x}_t = \mathbf{v}_t(\mathbf{x}_t)\, dt + \sigma\, d\mathbf{W}_t \tag{24}$$

where $\mathbf{v}_t(\mathbf{x}_t)$ is the deterministic drift that guides the embeddings towards the target distribution, $d\mathbf{W}_t$ denotes the Brownian motion that introduces stochastic noise [4], and σ controls the magnitude of the noise. In the proposed agentic RAG framework, embeddings evolve on a hyperbolic manifold $\mathbb{H}^d$. The deterministic Riemannian probability flow is defined as follows:

$$\frac{d\mathbf{x}_t}{dt} = \mathbf{v}_t(\mathbf{x}_t) \tag{25}$$

To allow exploration of alternative reasoning paths, stochastic properties can be incorporated, yielding hyperbolic diffusion:

$$d\mathbf{x}_t = \mathbf{v}_t(\mathbf{x}_t)\, dt + \sigma\, d\mathbf{B}_t^g \tag{26}$$

where $d\mathbf{B}_t^g$ denotes Brownian motion defined on the hyperbolic manifold [18]. Equation 25 moves embeddings along high-probability reasoning trajectories, while Equation 26 enables stochastic exploration of alternative knowledge or reasoning paths.

Acknowledgments

This research is supported by Taighde Éireann – Research Ireland under Grant Number SFI/12/RC/2289 P2 (Insight Research Ireland Centre for Data Analytics), co-funded by the European Regional Development Fund.

References

[1] David Alvarez-Melis, Youssef Mroueh, and Tommi Jaakkola. 2020. Unsupervised hierarchy matching with optimal transport over hyperbolic spaces. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 1606–1617.

[2] Joey Bose, Ariella Smofsky, Renjie Liao, Prakash Panangaden, and Will Hamilton. 2020. Latent variable modelling with hyperbolic normalizing flows. In *International conference on machine learning*. PMLR, 1045–1055.

[3] Davide Caffagni, Federico Cocchi, Nicholas Moratelli, Sara Sarto, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. 2024. Wiki-llava: Hierarchical retrieval-augmented generation for multimodal llms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1818–1826.

[4] Pavel Castro-Villarreal. 2010. Brownian motion meets Riemann curvature. *Journal of Statistical Mechanics: Theory and Experiment* 2010, 08 (2010), P08006.

[5] Wenhu Chen, Hexiang Hu, Xi Chen, Pat Verga, and William Cohen. 2022. Murag: Multimodal retrieval-augmented generator for open question answering over images and text. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 5558–5570.

[6] Wentao Cheng, Zhida Qin, Zexue Wu, Pengzhan Zhou, and Tianyu Huang. 2025. Large language models enhanced hyperbolic space recommender systems. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1944–1953.

[7] Aurora Linh Cramer, Ho-Hsiang Wu, Justin Salamon, and Juan Pablo Bello. 2019. Look, listen, and learn more: Design choices for deep audio embeddings. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3852–3856.

[8] Yashar Deldjoo, Zhankui He, Julian McAuley, Anton Korikov, Scott Sanner, Arnau Ramisa, René Vidal, Maheswaran Sathiamoorthy, Atoosa Kasirzadeh, and Silvia Milano. 2024. A review of modern recommender systems using generative models (gen-recsys). In *Proceedings of the 30th ACM SIGKDD conference on Knowledge Discovery and Data Mining*. 6448–6458.

[9] Karan Desai, Maximilian Nickel, Tanmay Rajpurohit, Justin Johnson, and Shanmukha Ramakrishna Vedantam. 2023. Hyperbolic image-text representations. In *International Conference on Machine Learning*. PMLR, 7694–7731.

[10] Shanshan Feng, Guoxin Yu, Dawei Liu, Han Hu, Yong Luo, Hui Lin, and Yew-Soon Ong. 2025. MHR: A Multi-Modal Hyperbolic Representation Framework for Fake News Detection. *IEEE Transactions on Knowledge and Data Engineering* (2025).

[11] Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. 2018. Hyperbolic neural networks. *Advances in neural information processing systems* 31 (2018).

[12] Mingyang Gao, Suyang Zhou, Wei Gu, Jili Fan, Haiquan Liu, Aobo Guan, and Honghua Xu. 2026. Pre-Trained Large Model as Distributed PV Power Forecaster: Integrating Scenario-Consistent Retrieval-Augmented Generation and Cross-Modal Semantic Alignment. *IEEE Transactions on Sustainable Energy* (2026).

[13] Zhi Gao, Yuwei Wu, Yunde Jia, and Mehrtash Harandi. 2022. Hyperbolic feature augmentation via distribution estimation and infinite sampling on manifolds. *Advances in neural information processing systems* 35 (2022), 34421–34435.

[14] Mohammad Ghomi. 2013. Tangent lines, inflections, and vertices of closed curves. (2013).

[15] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Mingwei Chang. 2020. Retrieval augmented language model pre-training. In *International conference on machine learning*. PMLR, 3929–3938.

[16] Rui Han, Qun Liu, Xu Gong, Guoyin Wang, Li Liu, and Xingping Xian. 2025. MHGCL: A Multi-Modal Hypergraph Contrastive Learning Framework for Molecular Property Prediction. *Applied Soft Computing* 184, Part A (2025), 113645. doi:10.1016/j.asoc.2025.113645

[17] Neil He, Hiren Madhu, Ngoc Bui, Menglin Yang, and Rex Ying. 2025. Hyperbolic deep learning for foundation models: A survey. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*. 6021–6031.

[18] Pei Hsu. 1988. Brownian motion and Riemannian geometry. *Contemp. Math* 73 (1988), 95–104.

[19] Guillaume Huguet, Daniel Sumner Magruder, Alexander Tong, Oluwadamilola Fasina, Manik Kuchroo, Guy Wolf, and Smita Krishnaswamy. 2022. Manifold interpolating optimal-transport flows for trajectory inference. *Advances in neural information processing systems* 35 (2022), 29705–29718.

[20] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. 2021. Scaling up visual and vision-language representation learning with noisy text supervision. In *International conference on machine learning*. PMLR, 4904–4916.

[21] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems* 33 (2020), 9459–9474.

[22] Qian Li, Cheng Ji, Shu Guo, Kun Peng, Qianren Mao, and Shangguang Wang. 2025. Variational Multi-Modal Hypergraph Attention Network for Multi-Modal Relation Extraction. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)*. ijcai.org, 8159–8167. doi:10.24963/ijcai.2025/907

[23] Qi Liu, Maximilian Nickel, and Douwe Kiela. 2019. Hyperbolic graph neural networks. *Advances in neural information processing systems* 32 (2019).

[24] Paolo Mandica, Luca Franco, Konstantinos Kallidromitis, Suzanne Petryk, and Fabio Galasso. 2024. Hyperbolic learning with multimodal large language models. In *European Conference on Computer Vision*. Springer, 382–398.

[25] Boci Peng, Yun Zhu, Yongchao Liu, Xiaohe Bo, Haizhou Shi, Chuntao Hong, Yan Zhang, and Siliang Tang. 2025. Graph retrieval-augmented generation: A survey. *ACM Transactions on Information Systems* 44, 2 (2025), 1–52.

[26] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PmLR, 8748–8763.

[27] Md Mahbubur Rahman, Shufeng Hao, Chongyang Shi, An Lao, and Jinyan Liu. 2025. Exploring Hyperbolic Hierarchical Structure for Multimodal Rumor Detection. In *Findings of the Association for Computational Linguistics: EMNLP 2025*. 115–134.

[28] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[29] Bhaskarjit Sarmah, Dhagash Mehta, Benika Hall, Rohan Rao, Sunil Patel, and Stefano Pasquali. 2024. Hybridrag: Integrating knowledge graphs and vector retrieval augmented generation for efficient information extraction. In *Proceedings of the 5th ACM International Conference on AI in Finance*. 608–616.

[30] Jiajie Su, Chaochao Chen, Zibin Lin, Xi Li, Weiming Liu, and Xiaolin Zheng. 2023. Personalized behavior-aware transformer for multi-behavior sequential recommendation. In *Proceedings of the 31st ACM International Conference on Multimedia*. 6321–6331.

[31] Théo Uscidda and Marco Cuturi. 2023. The monge gap: A regularizer to learn all transport maps. In *International Conference on Machine Learning*. PMLR, 34709–34733.

[32] Weiye Xu, Min Wang, Wengang Zhou, and Houqiang Li. 2024. P-rag: Progressive retrieval augmented generation for planning on embodied everyday task. In *Proceedings of the 32nd ACM International Conference on Multimedia*. 6969–6978.

[33] Xing Xu, Tan Wang, Yang Yang, Lin Zuo, Fumin Shen, and Heng Tao Shen. 2020. Cross-modal attention with semantic consistence for image–text matching. *IEEE transactions on neural networks and learning systems* 31, 12 (2020), 5412–5425.

[34] Yuhao Yang, Chao Huang, Lianghao Xia, Yuxuan Liang, Yanwei Yu, and Chenliang Li. 2022. Multi-behavior hypergraph-enhanced transformer for sequential recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2263–2274.

[35] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. In *The eleventh international conference on learning representations*.

[36] Linwei Ye, Mrigank Rochan, Zhi Liu, and Yang Wang. 2019. Cross-modal self-attention network for referring image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10502–10511.

[37] Yilin Ye, Shishi Xiao, Xingchen Zeng, and Wei Zeng. 2024. ModalChorus: Visual probing and alignment of multi-modal embeddings via modal fusion map. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[38] Qinhan Yu, Zhiyou Xiao, Binghui Li, Zhengren Wang, Chong Chen, and Wentao Zhang. 2025. MRAMG-Bench: a comprehensive benchmark for advancing multimodal retrieval-augmented multimodal generation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3616–3626.

[39] Meng Yuan, Yutian Xiao, Wei Chen, Chou Zhao, Deqing Wang, and Fuzhen Zhuang. 2025. Hyperbolic diffusion recommender model. In *Proceedings of the ACM on Web Conference 2025*. 1992–2006.

[40] Ding-Chu Zhang, Yida Zhao, Jialong Wu, Liwen Zhang, Baixuan Li, Wenbiao Yin, Yong Jiang, Yu-Feng Li, Kewei Tu, Pengjun Xie, et al. 2025. Evolvesearch: An iterative self-evolving search agent. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. 13134–13147.

[41] Qing Zhang, Jing Zhang, Xiangdong Su, Feilong Bao, and Guanglai Gao. 2026. Hyperbolic-Based Cross-Modal Semantic Remodeling Network for Zero-Shot Sketch-Based Image Retrieval. *ACM Transactions on Multimedia Computing, Communications and Applications* 22, 1 (2026), 1–23.