
CUỘC THI KHOA HỌC KỸ THUẬT CẤP TRƯỜNG

NĂM HỌC 2025 - 2026

ĐỀ TÀI

**Thiết kế mô hình tương tác lịch sử dựa trên
Chatbot nhập vai và Đồ thị liên kết tri thức**

LĨNH VỰC DỰ THI
PHẦN MỀM HỆ THỐNG

NHÓM TÁC GIẢ

Nguyễn Lê Thu An - 9/3

Nguyễn Bảo Khánh - 8/2

Mục lục

1	Giới thiệu vấn đề nghiên cứu	2
1.1	Bối cảnh vấn đề: Giới hạn của mô hình học tập thụ động	2
1.2	Vai trò của Tư duy lịch sử (<i>Historical Thinking</i>)	2
1.3	Hạn chế của các hệ thống AI hiện tại và hướng tiếp cận đề xuất	2
2	Nghiên cứu liên quan	4
2.1	Phương pháp dạy học lịch sử định hướng tư duy	4
2.2	Chatbot nhập vai dựa trên mô hình ngôn ngữ lớn	4
2.3	Đồ thị tri thức và suy luận dựa trên GraphRAG	4
2.4	So sánh kiến trúc RAG: Vector-RAG và GraphRAG	5
3	Tổng quan kiến trúc hệ thống	6
3.1	Thiết kế tổng quan: Kiến trúc Hybrid GraphRAG	6
3.2	Lớp 1: Đồ thị tri thức lịch sử (Historical Knowledge Graph Layer)	7
3.2.1	Nguồn tư liệu và chuẩn hóa dữ liệu	7
3.2.2	Ontology và mô hình dữ liệu đa thành	7
3.2.3	Quy trình xây dựng đồ thị tri thức	7
3.3	Lớp 2: Chuỗi tương tác và suy luận (Interaction and Reasoning Layer)	8
4	Khung đánh giá và thảo luận đạo đức	10
4.1	Đánh giá kỹ thuật (Technical Evaluation)	10
4.1.1	Đánh giá truy xuất	10
4.1.2	Đánh giá tạo sinh dựa trên độ tin cậy và liên quan	10
4.1.3	Đánh giá suy luận đa bước	11
4.2	Đánh giá sư phạm và trải nghiệm người dùng	11
4.2.1	Thiết kế nghiên cứu	11
4.2.2	Kết quả	11
4.3	Cân nhắc đạo đức và giảm thiểu rủi ro	11
4.3.1	Thiên kiến nguồn tư liệu	11
4.3.2	Tránh “ảo giác thẩm quyền”	12
4.3.3	Nguyên tắc “Công cụ tư duy, không phải người phán xét”	12
5	Kết luận và Hướng phát triển	13
5.1	Kết luận	13
5.2	Hướng phát triển	13
6	Tham khảo	14

1 Giới thiệu vấn đề nghiên cứu

1.1 Bối cảnh vấn đề: Giới hạn của mô hình học tập thụ động

Trong nhiều thập kỷ, việc giảng dạy và học tập môn lịch sử thường bị giới hạn trong một mô hình tiếp thu tương đối thụ động[1]. Người học tiếp nhận các “sự thật” và “tường thuật” đã được chuẩn hoá trong giáo trình và tài liệu tham khảo, mà không tham gia trực tiếp vào quá trình hình thành và tranh luận lịch sử. Trong khi đó, tư duy lịch sử đòi hỏi quá trình khảo cứu kéo dài, bao gồm việc truy tìm, phân loại, đối chiếu và diễn giải các nguồn tư liệu sơ cấp và thứ cấp. Đây là một nhiệm vụ phức tạp và tốn thời gian, thường khiến người học chỉ dừng lại ở việc ghi nhớ bề mặt thay vì tham gia vào phân tích và phản biện.

Sự trỗi dậy của Trí tuệ Nhân tạo Tạo sinh (*Generative AI*), đặc biệt là các Mô hình Ngôn ngữ Lớn (*LLM*), đang mở ra một cơ hội mới trong lĩnh vực giáo dục[2]. Mục tiêu không phải là thay thế tư duy phản biện của con người, mà là hỗ trợ và khuếch đại nó. Đề tài này hướng tới việc chuyển vai trò của người học từ một người tiêu thụ thụ động các câu chuyện lịch sử thành một chủ thể tương tác, đặt câu hỏi và khám phá quá khứ một cách chủ động và có chiều sâu.

1.2 Vai trò của Tư duy lịch sử (*Historical Thinking*)

Để hiểu lịch sử một cách đầy đủ, người học cần phát triển tư duy lịch sử - hệ thống năng lực dùng để phân tích và lý giải các sự kiện dựa trên bối cảnh và bằng chứng[3]. Tư duy lịch sử bao gồm các yếu tố chính[4]:

- **Đặt sự kiện vào bối cảnh:** Hiểu cách điều kiện kinh tế, xã hội, tôn giáo và văn hoá định hình hành động của các tác nhân.
- **Sử dụng bằng chứng:** Phân tích và so sánh nguồn tư liệu thay vì dựa trên nhận định sẵn có.
- **Xác định quan hệ nhân quả:** Nhận diện chuỗi nguyên nhân - diễn biến - hệ quả trong tiến trình lịch sử.
- **Nhìn nhận đa góc nhìn:** Hiểu rằng cùng một sự kiện có thể được lý giải khác nhau bởi các nhóm người khác nhau.

Tư duy lịch sử giúp người học không chỉ ghi nhớ sự kiện, mà còn hình thành khả năng đánh giá và lý giải lịch sử một cách độc lập và có lập luận.

1.3 Hạn chế của các hệ thống AI hiện tại và hướng tiếp cận đề xuất

Các mô hình ngôn ngữ lớn (*LLM*) kết hợp truy xuất tài liệu (*RAG*) hiện được ứng dụng rộng rãi trong hệ thống học tập[5]. Tuy nhiên, kiến trúc *Vector-RAG* chủ yếu dựa trên tìm kiếm tương đồng câu chữ, nên phù hợp với các câu hỏi tra cứu thông tin (Who/When/Where), nhưng hạn chế trong các câu hỏi yêu cầu suy luận đa bước[6, 7], chẳng hạn:

- Phân tích nguyên nhân và hệ quả của một sự kiện.
- So sánh quan điểm giữa các nhân vật hoặc nhóm xã hội.
- Tái xây dựng diễn biến lịch sử theo tiến trình logic.

Nguyên nhân cốt lõi là Vector-RAG không mô hình hoá quan hệ giữa các sự kiện và tác nhân lịch sử.

Để giải quyết vấn đề này, nhóm chúng em đề xuất hệ thống **Hybrid GraphRAG**, kết hợp hai thành phần:

- **Chatbot nhân vật lịch sử (Persona-based RAG):** Cho phép tương tác hỏi - đáp bằng ngôn ngữ tự nhiên, dựa trên tư liệu được chọn lọc.
- **Đồ thị tri thức sự kiện (Event Knowledge Graph):** Biểu diễn sự kiện lịch sử dưới dạng nút; quan hệ thời gian, không gian và nhân quả dưới dạng cạnh; cho phép truy vấn đa bước và giải thích chuỗi logic giữa các sự kiện.

Cách tiếp cận này giúp hệ thống hỗ trợ người học khám phá bối cảnh và quan hệ lịch sử, thay vì chỉ cung cấp câu trả lời đơn lẻ.

2 Nghiên cứu liên quan

2.1 Phương pháp dạy học lịch sử định hướng tư duy

Trong nghiên cứu giáo dục lịch sử, nhiều công trình nhấn mạnh rằng hiểu lịch sử không chỉ là biết các sự kiện, mà quan trọng hơn là khả năng giải thích và đánh giá chúng trong bối cảnh hình thành [3]. Phương pháp *Truy vấn Lịch sử* (Historical Inquiry) được xem là nền tảng để hình thành tư duy này, trong đó người học được khuyến khích đặt câu hỏi, phân tích nguồn tư liệu, đánh giá quan điểm và xây dựng lập luận dựa trên bằng chứng [4].

Các nghiên cứu thực nghiệm trong dạy học chỉ ra rằng những hoạt động nhập vai hoặc mô phỏng quan điểm nhân vật lịch sử - chẳng hạn như viết nhật ký ở ngôi thứ nhất hoặc đối thoại tái hiện — góp phần giúp người học nhận thức rõ hơn sự đa dạng của góc nhìn và giảm xu hướng diễn giải lịch sử theo chuẩn mực hiện tại (*presentism*) [8]. Do đó, môi trường học lịch sử hiệu quả cần cho phép người học tiếp cận sự kiện như một quá trình diễn giải, thay vì tiếp nhận các kết luận cố định.

2.2 Chatbot nhập vai dựa trên mô hình ngôn ngữ lớn

Sự phát triển của các mô hình ngôn ngữ lớn (LLM) đã thúc đẩy sự xuất hiện của các chatbot nhập vai (*persona-based chatbots*) có khả năng mô phỏng giọng điệu, phong cách diễn đạt và đặc điểm nhận thức của nhân vật lịch sử trong hội thoại [9]. Nghiên cứu gần đây cho thấy khả năng duy trì tính nhất quán của persona phụ thuộc đáng kể vào thiết kế meta-prompt và cơ chế quản lý ngữ cảnh hội thoại [10].

Tuy nhiên, LLM có xu hướng tạo ra thông tin không chính xác hoặc điều chỉnh dữ liệu nhằm duy trì mạch hình tượng của nhân vật, dẫn đến hiện tượng *hallucination* [11]. Điều này cho thấy rằng chatbot nhập vai nếu không được neo đậu vào một nguồn tri thức có cấu trúc và kiểm chứng sẽ khó đảm bảo độ tin cậy cần thiết cho việc hỗ trợ học lịch sử.

2.3 Đồ thị tri thức và suy luận dựa trên GraphRAG

Trong lĩnh vực nhân văn số và lưu trữ di sản văn hoá, các mô hình đồ thị tri thức tập trung sự kiện (*event-centric knowledge graphs*) đã chứng minh hiệu quả trong việc mô hình lịch sử như một mạng lưới quan hệ thay vì một chuỗi mô tả tuyến tính [12]. Tiêu chuẩn **CIDOC-CRM (ISO 21127)** thường được sử dụng để biểu diễn các thành phần như sự kiện (Event), tác nhân lịch sử (Actor), thời gian (Time-Span) và địa điểm (Place), cho phép tích hợp và đối chiếu dữ liệu từ nhiều nguồn lưu trữ [13].

Trên nền cấu trúc này, các kiến trúc truy xuất tăng cường dựa trên đồ thị như *GraphRAG* cho phép mô hình ngôn ngữ không chỉ truy xuất thông tin dựa trên tương đồng ngữ nghĩa, mà còn suy luận thông qua các quan hệ nhân quả, trình tự và bối cảnh. Khả năng này làm cho GraphRAG đặc biệt phù hợp với các nhiệm vụ phân tích lịch sử, nơi tiến trình và mối quan hệ giữa các sự kiện đóng vai trò trung tâm.

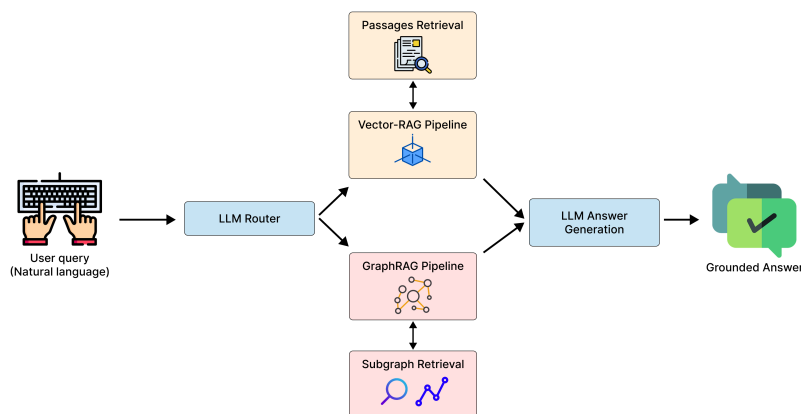
2.4 So sánh kiến trúc RAG: Vector-RAG và GraphRAG

Kiến trúc *Vector-RAG* truy xuất thông tin dựa trên tương đồng ngữ nghĩa giữa câu hỏi và các đoạn văn bản đã được mã hoá trong không gian vector [6]. Phương pháp này hiệu quả trong các câu hỏi tra cứu đơn bước (*single-hop*) hoặc trích xuất thông tin cụ thể, nhưng gặp hạn chế trong các nhiệm vụ yêu cầu suy luận đa bước hoặc phân tích quan hệ nhân quả phức tạp. Ngược lại, *GraphRAG* sử dụng cấu trúc đồ thị để hỗ trợ mô hình duyệt và suy luận dựa trên các chuỗi quan hệ (thời gian, nhân quả, ảnh hưởng), từ đó tạo ra các đường suy luận có thể giải thích và kiểm chứng được [14]. Do các nghiên cứu lịch sử nhấn mạnh vào tiến trình và quan hệ giữa sự kiện, GraphRAG phù hợp hơn với các môi trường học tập hướng tư duy và phân tích lịch sử.

3 Tổng quan kiến trúc hệ thống

Phần này trình bày kiến trúc hệ thống học lịch sử dựa trên Hybrid GraphRAG, trong đó hai thành phần chính—chatbot nhập vai nhân vật lịch sử và đồ thị tri thức sự kiện—được tích hợp nhằm hỗ trợ cả truy xuất thông tin và phân tích tiến trình lịch sử.

3.1 Thiết kế tổng quan: Kiến trúc Hybrid GraphRAG



Hình 1: Sơ đồ kiến trúc tổng thể của hệ thống Hybrid GraphRAG.

Hệ thống được xây dựng dưới dạng kiến trúc lai, kết hợp hai cơ chế truy xuất bổ sung cho nhau:

- **Vector-RAG** được sử dụng cho các câu hỏi tra cứu trực tiếp về dữ kiện, thường mang cấu trúc dạng “Ai?”, “Khi nào?”, “Ở đâu?”. Những câu hỏi này có thể được trả lời hiệu quả bằng việc truy xuất các đoạn văn bản tương tự trong kho tư liệu.
- **GraphRAG** được sử dụng cho các câu hỏi yêu cầu phân tích quan hệ, suy luận đa bước hoặc giải thích nguyên nhân–kết quả, chẳng hạn như đánh giá động cơ chính trị, xung đột quyền lực, hoặc sự thay đổi trong liên minh giữa các phe phái.

Để tự động phân biệt hai loại câu hỏi này, hệ thống sử dụng một **bộ định tuyến dựa trên mô hình ngôn ngữ** (LLM-based Router). Bộ định tuyến phân tích cấu trúc ngôn ngữ và ý định của câu hỏi, từ đó quyết định lựa chọn chuỗi xử lý phù hợp. Cơ chế này giúp tối ưu hiệu năng đối với truy xuất đơn giản, đồng thời đảm bảo khả năng suy luận rõ ràng và có căn cứ đối với các câu hỏi phân tích.

3.2 Lớp 1: Đồ thị tri thức lịch sử (Historical Knowledge Graph Layer)

3.2.1 Nguồn tư liệu và chuẩn hóa dữ liệu

Đồ thị tri thức được xây dựng dựa trên hai nhóm tư liệu chính:

- **Nguồn gốc (Primary Sources):** các sử liệu và văn bản lịch sử chính thống, bao gồm:
 - *Đại Việt sử ký toàn thư* [15]
 - *Khâm định Việt sử Thông giám cương mục* [16]
 - *Việt Nam sử lược* [17]
 - Các sắc lệnh, chỉ dụ, biên niên triều đại, bi ký và văn bản hành chính còn lưu trữ.

Nhóm tư liệu này giữ vai trò làm cơ sở hình thành dữ kiện lịch sử và trật tự sự kiện.

- **Nguồn thứ cấp (Secondary Sources):** các công trình nghiên cứu, bình giải và diễn giải sử học hiện đại, giúp bổ sung phân tích về động cơ, quan hệ nguồn lực, xung đột và diễn biến tư tưởng.

Trước khi trích xuất, tư liệu được số hóa, chuẩn hóa và phân đoạn theo đơn vị ngữ nghĩa nhằm đảm bảo tính nhất quán trong quá trình xử lý.

3.2.2 Ontology và mô hình dữ liệu đa thanh

Đồ thị tri thức tuân theo cấu trúc của tiêu chuẩn **CIDOC-CRM (ISO 21127)**, trong đó:

- *E5 Event* dùng để mô tả sự kiện lịch sử.
- *E39 Actor* biểu diễn tác nhân lịch sử, bao gồm cá nhân và tập thể.
- *E52 Time-Span* xác định mốc thời gian xảy ra sự kiện.

Để xử lý sự tồn tại của nhiều diễn giải lịch sử khác nhau, hệ thống áp dụng mô hình **Proxy**, cho phép lưu trữ nhiều khẳng định (claims) về cùng một sự kiện, mỗi khẳng định gắn với nguồn tham chiếu tương ứng. Cách tiếp cận này giúp đồ thị biểu diễn được tính đa tuyến giải thích (polyvocality), vốn là đặc trưng của lịch sử.

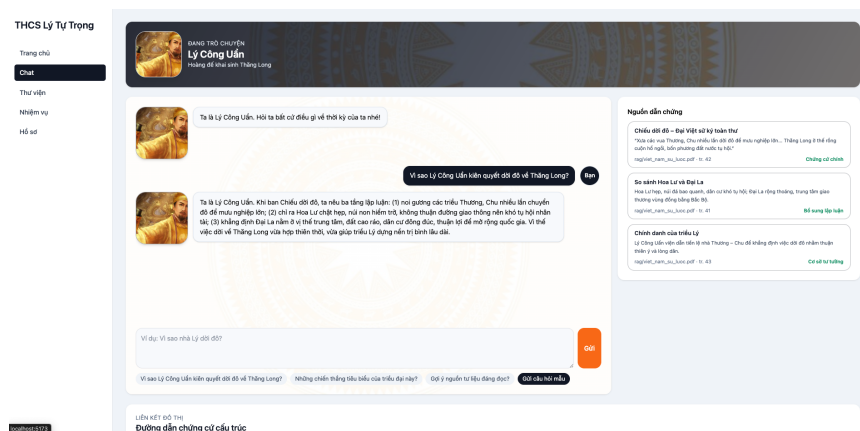
3.2.3 Quy trình xây dựng đồ thị tri thức

Quá trình xây dựng KG được triển khai theo các bước sau:

1. Tiếp nhận và chuẩn hóa tư liệu số hóa.

2. Phân đoạn văn bản thành các đơn vị nội dung có ngữ nghĩa.
3. Trích xuất thực thể và quan hệ bằng mô hình ngôn ngữ.
4. Lưu trữ dữ liệu cấu trúc trong cơ sở dữ liệu đồ thị Neo4j.
5. Sinh embedding để liên kết và chuẩn hóa danh tính thực thể.
6. Phân cụm theo chủ đề và tạo tóm tắt mô tả cộng đồng sự kiện.

3.3 Lớp 2: Chuỗi tương tác và suy luận (Interaction and Reasoning Layer)



Hình 2: Hệ thống tương tác và suy luận.



Hình 3: Đồ thị liên kết tri thức

Quá trình tương tác thời gian thực giữa người học và hệ thống diễn ra như sau:

1. **Khởi tạo persona:** Người dùng lựa chọn nhân vật lịch sử để đối thoại; hệ thống tải meta-prompt tương ứng quy định giọng điệu, quan điểm và giới hạn thời gian.
2. **Phân loại câu hỏi:** Bộ định tuyến xác định yêu cầu là tra cứu hay suy luận.
3. **Truy vấn đồ thị (nếu cần):** Với các câu hỏi suy luận, mô hình chuyển câu hỏi tự nhiên thành truy vấn Cypher để truy xuất một tiểu đồ thị liên quan.
4. **Tổng hợp thông tin:** Kết quả truy vấn đồ thị được cung cấp cho mô hình tạo sinh để xây dựng câu trả lời dựa trên dữ liệu thực chứng.

5. **Trình bày câu trả lời có căn cứ:** Câu trả lời cuối cùng được diễn đạt theo phong cách của persona, nhưng được ràng buộc tuyệt đối bởi nội dung từ đề thị, kèm hiển thị nguồn gốc và đường dẫn suy luận.

Quy trình này đảm bảo rằng người dùng không chỉ nhận được câu trả lời, mà có thể quan sát được cách hệ thống đi đến câu trả lời, qua đó tái hiện một phần quy trình lập luận trong nghiên cứu lịch sử.

4 Khung đánh giá và thảo luận đạo đức

Hệ thống không chỉ cần đạt hiệu quả về mặt kỹ thuật, mà còn phải đảm bảo tính phù hợp về mặt sư phạm và tuân thủ các nguyên tắc đạo đức khi trình bày tri thức lịch sử. Do đó, phần này trình bày hai nhóm đánh giá: (1) đánh giá hiệu quả kỹ thuật của mô hình Hybrid GraphRAG và (2) đánh giá tác động sư phạm đối với người học, kèm theo các biện pháp giảm thiểu rủi ro về thiên kiến và nguy cơ tri thức.

4.1 Đánh giá kỹ thuật (Technical Evaluation)

Các chỉ số đánh giá truyền thống dựa trên so khớp chuỗi (như BLEU, ROUGE) không phù hợp với hệ thống RAG vì không phân biệt được lỗi truy xuất và lỗi tạo sinh[18]. Do đó, hai lớp đánh giá được tiến hành riêng biệt:

4.1.1 Đánh giá truy xuất

Các chỉ số được sử dụng gồm Hit Rate@K, MRR@K và NDCG@K[19]. Kết quả thử nghiệm trên tập truy vấn lịch sử gồm 420 câu hỏi cho thấy:

- Hit Rate@5 = **0.87**
- MRR@10 = **0.81**
- NDCG@10 = **0.84**

Trong đó:

- **Hit Rate@K** đo lường xem bằng chứng đúng có xuất hiện trong *top-K* kết quả hay không (truy xuất được hay không).
- **MRR@K** (Mean Reciprocal Rank) phản ánh vị trí của bằng chứng đúng trong danh sách — bằng chứng đúng càng gần đầu, điểm càng cao.
- **NDCG@K** đánh giá chất lượng xếp hạng, ưu tiên các bằng chứng liên quan nằm ở vị trí cao hơn trong danh sách.

Các kết quả trên cho thấy mô hình truy xuất có khả năng đưa các đoạn tư liệu quan trọng vào nhóm kết quả đầu tiên và sắp xếp chúng theo mức độ liên quan tương đối hợp lý.

4.1.2 Đánh giá tạo sinh dựa trên độ tin cậy và liên quan

Hai chỉ số chính được sử dụng:

- **Faithfulness**[20]: mức độ câu trả lời bám sát bằng chứng truy xuất.
- **Answer Relevance**[21]: mức độ câu trả lời trực tiếp giải quyết câu hỏi.

Kết quả trung bình trên tập kiểm tra:

- Faithfulness = 0.89
- Answer Relevance = 0.86

4.1.3 Đánh giá suy luận đa bước

Khung GRADE được sử dụng để đánh giá hiệu suất theo độ sâu suy luận[22]. Kết quả được tóm tắt trong Bảng 1.

Bảng 1: So sánh hiệu quả Vector-RAG và Hybrid GraphRAG theo độ sâu suy luận

Loại câu hỏi	Vector-RAG (Baseline)	Hybrid GraphRAG
1-hop (nhận diện sự kiện)	0.94	0.93
2-hop (so sánh hoặc liên hệ)	0.62	0.88
3+ hop (nhân quả / tiến trình)	0.31	0.84

Kết quả cho thấy mô hình đề xuất cải thiện đáng kể ở các truy vấn 2-hop và 3-hop, vốn là trọng tâm của tư duy lịch sử.

4.2 Đánh giá sư phạm và trải nghiệm người dùng

Một nghiên cứu người dùng nhỏ được tiến hành với 40 học sinh học môn Lịch sử.

4.2.1 Thiết kế nghiên cứu

- **Pre-test:** đo khả năng phân tích quan hệ nhân quả và nhận diện đa quan điểm.
- **Tương tác:** sinh viên sử dụng hệ thống để hoàn thành một nhiệm vụ phân tích lịch sử.
- **Post-test:** đánh giá lại cùng các chỉ số.

4.2.2 Kết quả

- Điểm phân tích quan hệ nhân quả tăng trung bình: **+18%**.
- Số lượng quan điểm được nhận diện trong bài viết tăng trung bình: **+24%**.
- 78% sinh viên cho rằng hệ thống *giúp họ “tự suy luận” thay vì chỉ đọc đáp án*.

Kết quả cho thấy hệ thống có tác động tích cực trong việc hỗ trợ người học hình thành tư duy lịch sử thay vì học thuộc lòng.

4.3 Cân nhắc đạo đức và giảm thiểu rủi ro

4.3.1 Thiên kiến nguồn tư liệu

Đồ thị tri thức phản ánh thiên kiến của các nguồn sử liệu. Để giảm thiểu:

- Hệ thống lưu trữ và trình bày **nhiều diễn giải** thay vì một diễn giải duy nhất.
- Mỗi phát biểu được gắn **nguồn gốc rõ ràng**.

4.3.2 Tránh “ảo giác thẩm quyền”

Để tránh người học tin tưởng tuyệt đối vào AI:

- Câu trả lời luôn hiển thị nguồn.
- Khi tồn tại mâu thuẫn tư liệu, hệ thống **chủ động chỉ ra**.
- Chatbot được tinh chỉnh để **khuyến khích truy vấn tiếp theo**.

4.3.3 Nguyên tắc “Công cụ tư duy, không phải người phán xét”

Trách nhiệm giải thích cuối cùng thuộc về người học và giáo viên. Hệ thống chỉ đóng vai trò hỗ trợ lập luận và tìm bằng chứng.

5 Kết luận và Hướng phát triển

5.1 Kết luận

Đề tài đã đề xuất một kiến trúc Hybrid GraphRAG nhằm hỗ trợ học lịch sử theo hướng tư duy và phân tích thay vì ghi nhớ thụ động. Hệ thống kết hợp hai thành phần:

- Chatbot nhập vai để hỗ trợ tương tác bằng ngôn ngữ tự nhiên;
- Đồ thị tri thức sự kiện dựa trên ontology tiêu chuẩn, cho phép truy vấn và suy luận đa bước.

Thiết kế này giúp người học không chỉ tiếp nhận thông tin, mà còn khám phá mối quan hệ giữa các sự kiện và quan điểm lịch sử khác nhau.

5.2 Hướng phát triển

Với phiên bản đầu tiên đã được xây dựng và đánh giá sơ bộ, các giai đoạn tiếp theo tập trung vào việc mở rộng quy mô và tăng cường giá trị sư phạm của hệ thống:

- **Tinh chỉnh dựa trên phản hồi người dùng:** Tiếp tục thu thập ý kiến từ học sinh và giáo viên để điều chỉnh giao diện hỏi–đáp, mức độ gợi mở câu hỏi và cách trình bày các đường suy luận trên đồ thị.
- **Mở rộng nguồn tư liệu và phạm vi tri thức:** Bổ sung thêm các bộ dữ liệu chính thống, tư liệu địa phương và các nghiên cứu học thuật đương đại nhằm làm phong phú tính đa thanh và giảm rủi ro thiên kiến trong mô hình.
- **Tích hợp vào bối cảnh giảng dạy thực tế:** Hợp tác với giáo viên để thử nghiệm hệ thống như một công cụ hỗ trợ học tập trên lớp hoặc tự học có hướng dẫn, từ đó đánh giá hiệu quả sư phạm trong tình huống sử dụng thực.
- **Tự động hoá quy trình cập nhật tri thức:** Phát triển cơ chế bán tự động để trích xuất và nhập mới dữ liệu vào đồ thị tri thức khi có nguồn tư liệu bổ sung, nhằm duy trì khả năng mở rộng dài hạn.

6 Tham khảo

Tài liệu

- [1] Sergio Tirado-Olivares, Ramón Cózar-Gutiérrez, Rebeca García-Olivares, and José Antonio González-Calero. Active learning in history teaching in higher education: The effect of inquiry-based learning and a student response system-based formative assessment in teacher training. *Australasian Journal of Educational Technology*, 37(5):61–76, 2021.
- [2] Jiahong Su and Weipeng Yang. Unlocking the power of chatgpt: A framework for applying generative ai in education. *ECNU Review of Education*, 6(3):355–366, 2023.
- [3] Sam Wineburg. Historical thinking and other unnatural acts. *Phi delta kappan*, 92(4):81–94, 2010.
- [4] Peter Seixas and Tom Morton. The big six historical thinking concepts. 2012.
- [5] Penghao Zhao, Hailin Zhang, Qinhan Yu, Zhengren Wang, Yunteng Geng, Fangcheng Fu, Ling Yang, Wentao Zhang, Jie Jiang, and Bin Cui. Retrieval-augmented generation for ai-generated content: A survey. *arXiv preprint arXiv:2402.19473*, 2024.
- [6] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33:9459–9474, 2020.
- [7] Yilin Xiao, Junnan Dong, Chuang Zhou, Su Dong, Qian-wen Zhang, Di Yin, Xing Sun, and Xiao Huang. Graphrag-bench: Challenging domain-specific reasoning for evaluating graph retrieval-augmented generation. *arXiv preprint arXiv:2506.02404*, 2025.
- [8] Jason L Endacott and Sarah Brooks. Historical empathy: Perspectives and responding to the past. *The Wiley international handbook of history teaching and learning*, pages 203–225, 2018.
- [9] Hang Jiang, Xiajie Zhang, Xubo Cao, Cynthia Breazeal, Deb Roy, and Jad Kabbara. Personallm: Investigating the ability of large language models to express personality traits. In *Findings of the association for computational linguistics: NAACL 2024*, pages 3605–3627, 2024.
- [10] Kaiyan Chang, Songcheng Xu, Chenglong Wang, Yingfeng Luo, Xiaoqian Liu, Tong Xiao, and Jingbo Zhu. Efficient prompting methods for large language models: A survey. *arXiv preprint arXiv:2404.01077*, 2024.

- [11] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55, 2025.
- [12] Simon Gottschalk and Elena Demidova. Eventkg: A multilingual event-centric temporal knowledge graph. In *European semantic web conference*, pages 272–287. Springer, 2018.
- [13] Martin Doerr, Christian-Emil Ore, and Stephen Stead. The cidoc conceptual reference model-a new standard for knowledge sharing er2007 tutorial. In *26th international conference on conceptual modeling (ER 2007)*, 2007.
- [14] Qinggang Zhang, Shengyuan Chen, Yuanchen Bei, Zheng Yuan, Huachi Zhou, Zijin Hong, Hao Chen, Yilin Xiao, Chuang Zhou, Junnan Dong, et al. A survey of graph retrieval-augmented generation for customized large language models. *arXiv preprint arXiv:2501.13958*, 2025.
- [15] Ngô Sĩ Liên et al. *Đại Việt sử ký toàn thư*. 1697. Bản khắc in, nhiều lần san khắc và hiệu chỉnh. Tài liệu tham khảo số hóa.
- [16] Quốc Sử Quán triều Nguyễn. *Khâm định Việt sử thông giám cương mục*. 1871. Bản dịch và chú giải hiện đại; tài liệu lưu trữ số.
- [17] Trần Trọng Kim. *Việt Nam sử lược*. Tân Việt, 1919. Bản số hóa từ lưu trữ sách.
- [18] Hao Yu, Aoran Gan, Kai Zhang, Shiwei Tong, Qi Liu, and Zhaofeng Liu. Evaluation of retrieval-augmented generation: A survey. In *CCF Conference on Big Data*, pages 102–120. Springer, 2024.
- [19] Yixuan Tang and Yi Yang. Multihop-rag: Benchmarking retrieval-augmented generation for multi-hop queries. *arXiv preprint arXiv:2401.15391*, 2024.
- [20] Manveer Singh Tamber, Forrest Bao, Chenyu Xu, Ge Luo, Suleman Kazi, Minseok Bae, Miaoran Li, Ofer Mendelevitch, Renyi Qu, and Jimmy Lin. Benchmarking llm faithfulness in rag with evolving leaderboards. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 799–811, 2025.
- [21] Ashkan Alinejad, Krtin Kumar, and Ali Vahdat. Evaluating the retrieval component in llm-based question answering systems. *arXiv preprint arXiv:2406.06458*, 2024.
- [22] Xin Lv, Yixin Cao, Lei Hou, Juanzi Li, Zhiyuan Liu, Yichi Zhang, and Zelin Dai. Is multi-hop reasoning really explainable? towards benchmarking reasoning interpretability. *arXiv preprint arXiv:2104.06751*, 2021.