

ĐÁNH GIÁ HỆ THỐNG MULTI-AGENT KẾT HỢP MÔ HÌNH NGÔN NGỮ LỚN CHO HỌC TẬP CÁ NHÂN HÓA QUA TRÒ CHƠI

Trương Nguyễn Yến Nhi¹

¹ Trường ĐH Công Nghệ Thông Tin (UIT)

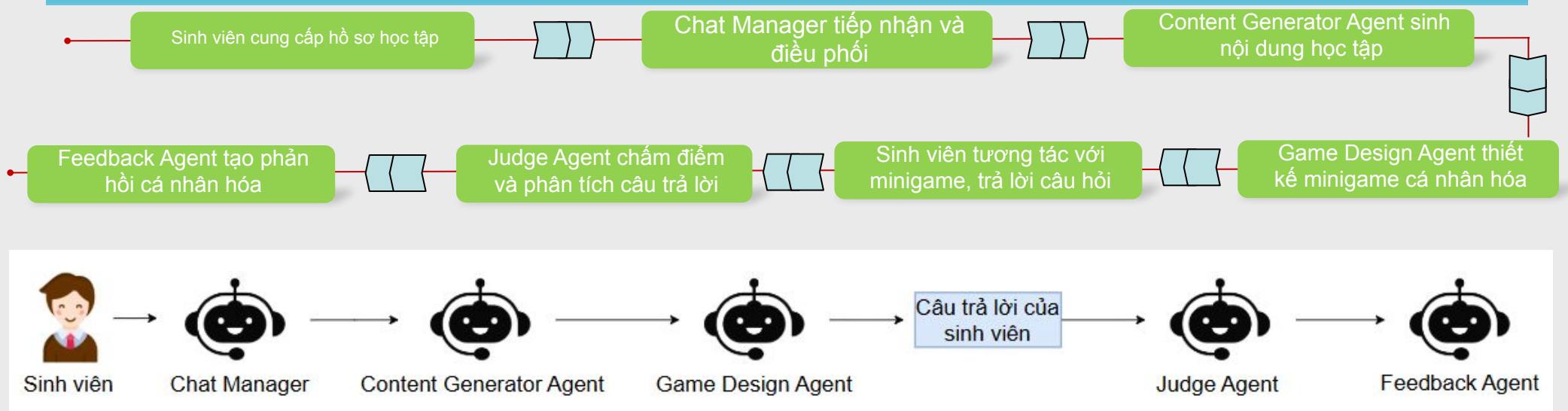
What ?

- Nghiên cứu hiện thực hóa hệ thống học tập cá nhân hóa kết hợp Multi-Agent và LLMs, cụ thể:
- Xây dựng hệ thống gồm các Agent: tạo nội dung, thiết kế trò chơi, đánh giá, phản hồi.
 - Tích hợp 4 mô hình LLM (GPT-3.5, GPT-4, Claude, Llama 3) vào từng Agent.
 - Phát triển hai loại trò chơi học tập: "4 Pics 1 Word" và "Coding Game".
 - Tổ chức đánh giá thực nghiệm với 50 sinh viên đại học.
 - So sánh hiệu quả từng LLM theo vai trò tác tử thông qua 5 tiêu chí định lượng.

Why ?

- Học tập cá nhân hóa yêu cầu giáo viên phải tạo nội dung phù hợp với từng người học, điều này dẫn đến tình trạng quá tải và khó duy trì chất lượng giảng dạy.
- Tỷ lệ giảng viên/sinh viên còn thấp khiến việc áp dụng cá nhân hóa trong thực tiễn giáo dục đại học gặp nhiều hạn chế.
- Framework được đề xuất trong nghiên cứu [1] hiện chỉ dừng ở mức lý thuyết, chưa được triển khai và kiểm chứng thực nghiệm.
- Việc đánh giá định lượng hiệu quả của các mô hình LLM trong từng vai trò agent là cần thiết để tối ưu hệ thống học tập.

Overview



Description

1. Hồ sơ người học

- Sinh viên nhập hồ sơ gồm: ngành học, trình độ, sở thích, nhu cầu học tập.
- Thông tin được lưu vào User Profile Database.

2. Chat Manager

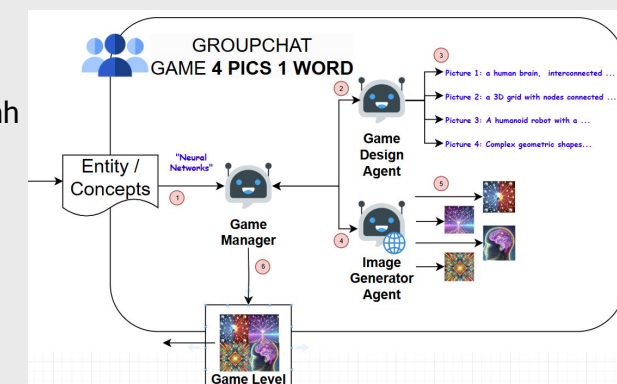
- Chat Manager nhận hồ sơ từ cơ sở dữ liệu sau đó: gửi thông tin đến Content Generator Agent. Đồng thời theo dõi quá trình xử lý của Game Design Agent.

3. Content Generator Agent

- Sử dụng LLM để tạo bài học, câu hỏi gợi ý phù hợp với người học.

4. Game Design Agent

- Nhận nội dung từ Content Generator Agent. Sau đó thiết kế minigame cá nhân hóa theo sở thích của người học.

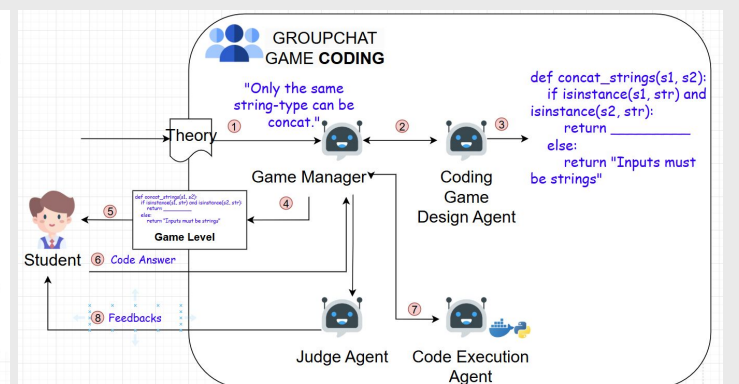


5. Sinh viên tương tác với minigame

- Sinh viên truy cập minigame, thực hiện các nhiệm vụ học tập.
- Trả lời câu hỏi, hoàn thành thử thách.

6. Judge Agent đánh giá câu trả lời

- Nhận input là câu trả lời của sinh viên.
- Phân tích bằng mô hình LLM: đúng/sai, độ đầy đủ, mức độ logic.



7. Feedback Agent tạo phản hồi cá nhân hóa

- Dựa trên đánh giá từ Judge Agent. Feedback Agent tạo phản hồi ngữ cảnh: Gợi ý cải thiện, giải thích lỗi sai, khuyến khích học tiếp

Kết quả

Hệ thống được đánh giá qua 5 tiêu chí: thời gian phản hồi, F1-score, mức độ phù hợp, phản hồi đúng ngữ cảnh và mức độ hài lòng (Likert 1–5).

Kết quả được phân tích bằng t-test và ANOVA để xác định LLM tối ưu cho từng Agent.