EVALUATING A MULTI-AGENT SYSTEM WITH LARGE LANGUAGE MODELS FOR PERSONALIZED GAMIFIED LEARNING

Truong Nguyen Yen Nhi[1]

[1] University of Information Technology
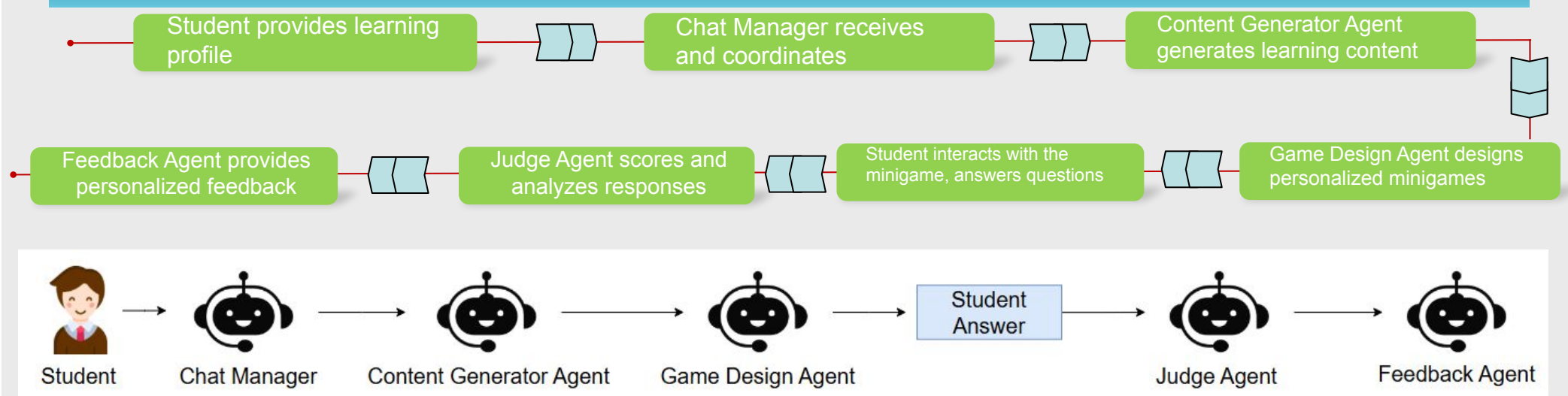HCMC, Vietnam

What ?

The study aims to implement a personalized learning system combining Multi-Agent and LLMs, specifically:
- Develop a system of Agents responsible for content, game design, assessment, and feedback.
- Integrate 4 LLMs (GPT-3.5, GPT-4, Claude, Llama 3) into each Agent.
- Develop educational games such as "4 Pics 1 Word" and "Coding Game".
- Conduct experimental evaluations with 50 university students.
- Compare the effectiveness of each LLM in different agent roles based on 5 quantitative criteria.

Why ?

- Personalized learning requires teachers to create content tailored to each learner, which often leads to overload and difficulty in maintaining teaching quality.
- The low teacher-to-student ratio poses challenges in implementing personalized learning in higher education.
- The proposed framework in [1] remains theoretical and has yet to be deployed and empirically validated.
- Evaluating the effectiveness of LLMs in different agent roles is essential to optimizing the learning system.

Overview

Student provides learning profile → Chat Manager receives and coordinates → Content Generator Agent generates learning content

Feedback Agent provides personalized feedback ← Judge Agent scores and analyzes responses ← Student interacts with the minigame, answers questions ← Game Design Agent designs personalized minigames

Description

1. User Profile

- Students input their profile, including: major, academic level, interests, and learning needs.
- This information is saved in the User Profile Database.

2. Chat Manager

- The Chat Manager receives the profile from the database and sends it to the Content Generator Agent.
- It also monitors the processing by the Game Design Agent.

3. Content Generator Agent

- Uses LLMs to generate lessons and question sets suitable for the learner.

4. Game Design Agent

- Receives content from the Content Generator Agent.
- Then designs personalized minigames according to the learner's interests.
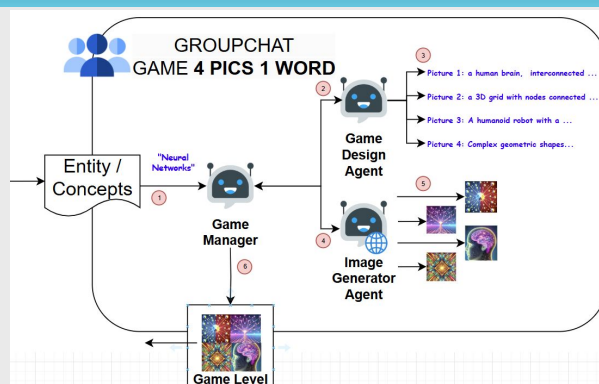
5. Student Interaction with Minigame

- Students access the minigame and complete assigned learning tasks.
- They answer questions and complete challenges.

6. Judge Agent Evaluates Answers

- Takes the student's answers as input.
- Uses LLM-based models to analyze: correct/incorrect, completeness, and logic.

7. Feedback Agent Generates Personalized Feedback
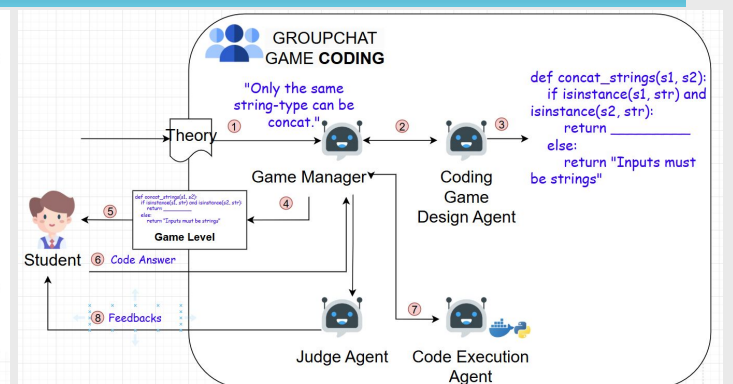
- Based on evaluations from the Judge Agent, the Feedback Agent generates feedback in natural language: Suggestions for improvement, error explanations, and motivational messages.

8. Evaluation

The system is evaluated using 5 criteria:
- Response time
- F1-score
- Appropriateness of questions
- Language quality of feedback
- Student satisfaction (Likert scale 1–5)

Statistical analysis includes t-test and ANOVA to identify the most effective LLM for each agent.

NII

Truong Nguyen Yen Nhi – University of Information Technology, HCMC
TEL : +84918***524 Email : nhitny.19@grad.uit.edu.vn