**1. A Survey on LLM-based Multi-Agent AI Hospital**

Bài báo “**A Survey on LLM-based Multi-Agent AI Hospital”** là một bài đánh giá toàn diện phân tích **72 nghiên cứu** về các Agent y tế dựa trên mô hình ngôn ngữ lớn (LLM) được xuất bản từ năm 2023 đến 2025

Khảo sát được tổ chức một cách có hệ thống và phân loại nghiên cứu bệnh viện AI thành ba lĩnh vực chính:

A diagram of a family tree

AI-generated content may be incorrect.

* **Các Thành phần Cốt lõi (Core Components)**:
* **Vai trò Agent (Agent Roles)**: Phân tích các vai trò cơ bản của Agent trong bệnh viện AI. Các Agent được phân loại thành nhiều nhóm...:
* **Agent tập trung vào bệnh nhân (Patient-Centered Agents)**: Bao gồm **Agent bệnh nhân (Patient Agents)** mô phỏng các đặc điểm bệnh nhân đa dạng, **Agent bệnh nhân tâm lý (Psychological Patient Agents - PPA)** mô phỏng các tình trạng sức khỏe tâm thần, và **Agent nội trú (Resident Agents - RAs)** mô phỏng quá trình bệnh tiến triển và hành vi tìm kiếm dịch vụ chăm sóc sức khỏe.
* **Agent chuyên gia y tế (Medical Professional Agents)**: Đóng vai trò quan trọng trong mô phỏng bệnh viện AI, bao gồm **Agent bác sĩ đa khoa (General Doctor Agent)**, **Agent chuyên khoa (Specialist Agent)**, **Agent trị liệu (Therapist Agent)**, **Agent y tá (Nurse Agent)**, **Agent kỹ thuật viên y tế (Medical Technician Agents)**, và **Agent sinh viên y khoa & giám khảo (Medical Student & Examiner Agents)**.
* **Agent làm việc nhóm AI y tế (Medical AI Teamwork Agents)**: Tập trung vào giải quyết các nhiệm vụ phức tạp, bao gồm **Agent suy luận theo mục tiêu (Goal-Driven Reasoning Agent)**, **Agent thẩm định lâm sàng (Clinical Judge Agent)**, **Agent phê bình (Critic Agent)**, **Agent lập kế hoạch (Planning Agent)**, **Agent ra quyết định (Decision Agent)**, và **Agent ghi âm (Recording Agent)**.
* **Agent nghiên cứu hỗ trợ AI (AI-Assisted Research Agents)**: Tối ưu hóa việc khám phá tri thức mới và hỗ trợ nghiên cứu khoa học, bao gồm **Agent lập kế hoạch nghiên cứu (Research Planning Agent)**, **Agent thực thi nghiên cứu (Research Executor Agent)**, **Agent phê bình khoa học (Scientific Critic Agent)**, và **Agent cơ sở dữ liệu (Database Agent)**.
* **Mô hình tương tác (Interaction Patterns)**: Bệnh viện AI sử dụng các mô hình khác nhau để tăng cường hiệu quả và độ tin cậy. Các mô hình này bao gồm **Hợp tác tập trung vào nhiệm vụ (Task-Focused Collaboration)**, **Ra quyết định có hướng dẫn từ chuyên gia (Expert-Guided Decision-Making)**, **Tối ưu hóa vấn đề lặp lại (Iterative Problem Optimization - IPO)**, **Tích hợp tri thức tự động (Automated Knowledge Integration - AKI)**, **Phối hợp dựa trên vai trò (Role-Based Coordination)**, và **Tranh luận tương tác đa vòng (Multi-Round Interactive Debate)**.
* **Tích hợp công cụ (Tool Integration)**: Các Agent trong bệnh viện AI sử dụng nhiều công cụ đa dạng để tăng cường hiệu quả và độ chính xác. Các công cụ được khảo sát bao gồm **Hệ thống truy xuất (Retrieval Systems)**, **Đồ thị tri thức (Knowledge Graphs)**, **Cây quyết định y tế (Medical Decision Trees)**, **LLM-as-KB (LLM như là cơ sở tri thức)**, **Xử lý đa phương thức (Multi-Modality Processing)**, **Suy luận tính toán (Computational Reasoning)**, và các công cụ hỗ trợ ra quyết định lâm sàng hoặc nghiên cứu y sinh khác.
* **Quản lý bộ nhớ (Memory Management)**: Bệnh viện AI tận dụng quản lý bộ nhớ có cấu trúc để học tập và ra quyết định thích ứng. Bao gồm **Bộ nhớ dài hạn (Long-Term Memory - LTM)** (với bộ nhớ nội bộ và bộ nhớ ngoài có khả năng lưu trữ tĩnh và cập nhật động)..., **Bộ nhớ ngắn hạn (Short-Term Memory - STM)**, và **Bộ nhớ làm việc chung đa Agent (Multi-Agent Shared Working Memory - WM)**.
* **Cơ chế suy luận (Reasoning Mechanisms)**: Các cơ chế suy luận khác nhau tăng cường khả năng ra quyết định23. Bao gồm **Suy luận trực tiếp (Direct Reasoning)** (gồm suy luận đơn đường và đa đường) và **Suy luận dựa trên phản hồi (Feedback-Based Reasoning)** (gồm phản hồi bên ngoài và tự thân).
* **Các Ứng dụng (Applications)**:
* **Mô phỏng các kịch bản cụ thể (Simulating Specific Scenarios)**: Bao gồm **Mô phỏng quy trình làm việc lâm sàng (Clinical Workflow Simulation)**, **Mô phỏng nhóm y tế đa ngành (Multi-Disciplinary Medical Team Simulation - MDT)**, **Bệnh nhân mô phỏng cho giáo dục và đào tạo y tế (Simulated Patients for Medical Education & Training)**, **Tư vấn tâm lý và mô phỏng tương tác sức khỏe tâm thần (Psychological Counseling and Mental Health Interaction Simulation)**, và **Tối ưu hóa quy trình y tế khác và mô phỏng đa ngành (Other Medical Process Optimization and Cross-Disciplinary Simulation)**.
* **Giải quyết các nhiệm vụ phức tạp (Solving Complex Tasks)**: Bao gồm **Ra quyết định lâm sàng chính xác (Precision Clinical Decision-Making)**, **Tối ưu hóa thử nghiệm lâm sàng và hệ thống phân loại (Clinical Trial Optimization and Triage Systems)**, **Nhiệm vụ chuyên sâu tri thức và khoa học dữ liệu y tế (Knowledge-Intensive Tasks and Medical Data Science)**, và **Khám phá khoa học và tự động hóa nghiên cứu (Scientific Discovery and Research Automation)**.
* **Đánh giá Agent (Evaluating Agents)**: Khảo sát cách đánh giá các Agent AI trong môi trường lâm sàng đã phát triển từ các tiêu chuẩn tĩnh sang mô phỏng multi Agent động.
* **Tổng hợp dữ liệu để đào tạo (Synthesizing Data for Training)**: Đóng vai trò quan trọng trong việc tạo ra các bộ dữ liệu đào tạo đa dạng và thực tế về mặt lâm sàng, khắc phục vấn đề riêng tư của dữ liệu bệnh nhân thực.
* **Các Thách thức Chính & Hướng đi Tương lai (Key Challenges & Future Directions)**: Khảo sát thảo luận về các thách thức hiện có trong từng thành phần cốt lõi và ứng dụng, cũng như cách phát triển bệnh viện AI thành một nền tảng nghiên cứu và thực hành có ý nghĩa hơn. Các thách thức chính bao gồm...:
* **Trong vai trò Agent**: Đảm bảo tính nhất quán về hành vi, quản lý thông tin bất đối xứng, và đảm bảo sự đa dạng của Agent bệnh nhân....
* **Trong mô hình tương tác**: Xác định rõ vai trò con người trong hệ thống và mô hình hóa sự không chắc chắn trong ra quyết định chiến lược.
* **Trong tích hợp công cụ**: Công cụ thường được xử lý như các tiện ích tĩnh, thiếu khung đánh giá dựa trên tác động thực tế.
* **Trong quản lý bộ nhớ**: Hạn chế của hồ sơ sức khỏe điện tử (EHR) tĩnh và thiếu cơ chế truy cập bộ nhớ động.
* **Trong cơ chế suy luận**: Hạn chế của suy luận trực tiếp đơn đường và khả năng xử lý sự không chắc chắn.
* **Trong mô phỏng kịch bản & giải quyết nhiệm vụ phức tạp**: Độ chính xác và toàn diện của mô phỏng y tế, ảnh hưởng của các yếu tố môi trường bên ngoài, tính mạnh mẽ và độ tin cậy của hệ thống, và quản lý rủi ro....
* **Trong đánh giá Agent**: Hạn chế của các phương pháp đánh giá hiện có và chưa xem xét đủ chi phí tính toán.
* **Trong tổng hợp dữ liệu đào tạo**: Hạn chế của học tăng cường (RL) trong môi trường y tế và thiếu sự đa dạng và công bằng trong dữ liệu tổng hợp.
* **Quản trị, đạo đức và vai trò của nhà nghiên cứu AI & người thực hành y tế**: Các vấn đề về trách nhiệm giải trình, minh bạch, thiên vị, tích hợp lâm sàng và thiếu sự hợp tác liên ngành.

**2. Multi-agents system for medical diagnosis**

**Mục tiêu chính của bài báo:**

* Đề xuất một hệ thống multi Agent (M.A.S) để phân phối quá trình chẩn đoán cho ba Agent chuyên gia (tai mũi họng, tiêu hóa và tim mạch).
* Cải thiện mức độ chăm sóc y tế bằng cách cung cấp các chẩn đoán chính xác hơn và các lựa chọn điều trị hiệu quả hơn.
* Tự động hóa và tối ưu hóa các quy trình y tế, giảm thiểu lỗi chẩn đoán và điều trị, đồng thời cải thiện thời gian tìm kiếm nguồn lực y tế.

**Bối cảnh và Vấn đề trong Y học:**

* Lĩnh vực y học là một môi trường rộng lớn, đặc trưng bởi **khía cạnh ra quyết định được chia sẻ và phân tán**.
* Nó đòi hỏi sự giao tiếp và phối hợp phức tạp giữa các khoa y tế khác nhau, bác sĩ và bệnh nhân.
* Các công nghệ mới tạo ra lượng dữ liệu khổng lồ về gen và protein, cần có các hệ thống xử lý thông tin mạnh mẽ để khai thác hiệu quả.

**Giải pháp đề xuất: Hệ thống Multi Agent (M.A.S) trong Chẩn đoán Y tế:**

* M.A.S tạo điều kiện thuận lợi cho việc quản lý các quyết định và hành động, đảm bảo giao tiếp và phối hợp bằng cách giảm lỗi chẩn đoán và điều trị24.
* Hệ thống mô phỏng một tập hợp các **bác sĩ chuyên khoa**, mỗi bác sĩ được mô hình hóa bởi một Agent có khả năng ra quyết định và giao tiếp với các Agent khác.
* Mục tiêu của hệ thống là đạt được một **chẩn đoán y tế cuối cùng** dựa trên các thông tin ban đầu thông qua sự hợp tác, phối hợp và suy luận của các Agent6.

**Các thành phần chính của hệ thống:**

A diagram of a system

AI-generated content may be incorrect.

1.**Một tập hợp các Agent**:

* Mô hình hóa các bác sĩ chuyên khoa.
* Mỗi Agent có **sự tự chủ** (kiểm soát hành vi riêng), **hành vi có chủ ý** (có khả năng suy luận, hợp tác, giao tiếp dựa trên niềm tin và mục tiêu), và **khả năng phản ứng kịp thời** (nhận thức môi trường và đưa ra phản hồi kịp thời).
* Kiến thức riêng của mỗi Agent (lĩnh vực y tế) được tích hợp.
* Kiến trúc nội bộ bao gồm: chuyên môn, niềm tin, lập luận, module giao tiếp.

2.**Bảng đen (Blackboard)**:

* Một cấu trúc phần mềm đảm bảo giao tiếp giữa nhiều Agent nhận thức hoạt động đồng thời bằng cách quản lý các giải pháp bộ phận theo thời gian thực dưới sự kiểm soát.
* Trong hệ thống này, Blackboard được chia thành các cấp độ, mỗi cấp độ liên kết với một Agent (ví dụ: một chuyên khoa).

3.**Cơ chế kiểm soát (Control Mechanism)**:

A diagram of a person's process

AI-generated content may be incorrect.

* Đóng vai trò quan trọng trong việc quản lý sự hợp tác và phối hợp giữa các Agent.
* Sử dụng **kiểm soát thủ tục (procedural control)**, được giải quyết bởi một chương trình duy nhất bao gồm: Monitor, Diary, và Ordonnansor.
* Chu trình hoạt động: Monitor đặt các Agent được kích hoạt vào Diary, sau đó Ordonnansor lựa chọn Agent theo các chiến lược khác nhau. Agent được chọn sẽ bắt đầu hoạt động, thường là sửa đổi Blackboard bằng cách thêm, xóa hoặc sửa đổi các thông tin.

4.**Module thu thập dữ liệu (Data acquisition Module)**:

* Cho phép cập nhật cơ sở kiến thức của mỗi Agent.

5.**Giao diện (Interface)**:

* Đóng vai trò trung gian giữa người dùng và hệ thống để thu thập dữ liệu.
* Người dùng có thể nhập các dấu hiệu lâm sàng từ một danh sách có sẵn hoặc một bệnh nghi ngờ.

**Cách thức hoạt động của hệ thống:**

* **Thu thập dữ liệu:** Người dùng (bác sĩ) cung cấp các dấu hiệu hoặc một bệnh nghi ngờ của bệnh nhân. Hệ thống sẽ hỏi thêm nếu cần để làm rõ các dấu hiệu.
* **Kích hoạt cơ chế kiểm soát:** Sau khi tham vấn Blackboard, bộ điều khiển phát hiện các Agent quan tâm đến những thay đổi và kích hoạt Agent ưu tiên.
* **Kích hoạt Agent:** Agent nhận dữ liệu từ Blackboard và áp dụng chiến lược suy luận của riêng mình để đưa ra chẩn đoán12.
* **Suy luận:** Công cụ suy luận, với kiến thức cụ thể của Agent (cơ sở quy tắc) và trạng thái Blackboard, có thể dẫn đến chẩn đoán bộ phận hoặc cuối cùng.
* **Sửa đổi Blackboard:** Blackboard được sửa đổi sau mỗi lần suy luận của một Agent (chèn các thông tin mới) cho chu trình tiếp theo.
* **Giải thích kết quả:** Khi Blackboard không còn chứa điều kiện kích hoạt cho các Agent khác, bộ điều khiển sẽ giải thích kết quả và đưa ra chẩn đoán y tế cuối cùng.
* Để cập nhật dữ liệu mới (cơ sở quy tắc cho mỗi Agent), hệ thống có chế độ cập nhật yêu cầu mật khẩu để đảm bảo tính bảo mật.

**Thách thức và Hướng đi tương lai:**

* **Giao tiếp:** Cần có tiêu chuẩn giao tiếp, từ vựng và ontology chung do sự phức tạp và phân tán của các hệ thống khoa y tế.
* **Bảo vệ dữ liệu y tế:** Thông tin y tế phải được bảo mật (ví dụ: sử dụng phương pháp mã hóa).
* Cần có khả năng điều chỉnh để hệ thống **linh hoạt hơn** cho việc thêm hoặc xóa Agent.
* Cải thiện giao diện để cho phép xác định các dấu hiệu hoặc triệu chứng bằng **văn bản tự do (free text)**.
* Khả năng thích ứng hệ thống để sử dụng trên Internet