

Metrics and drawbacks of large language model in medical systematic review

Lê Phước Trung^{1,3}

Trương Đăng Nghĩa^{2,3}

¹ 20522069, ² 20521658@gm.uit.edu.vn

³ Trường ĐH Công Nghệ Thông Tin

What ?

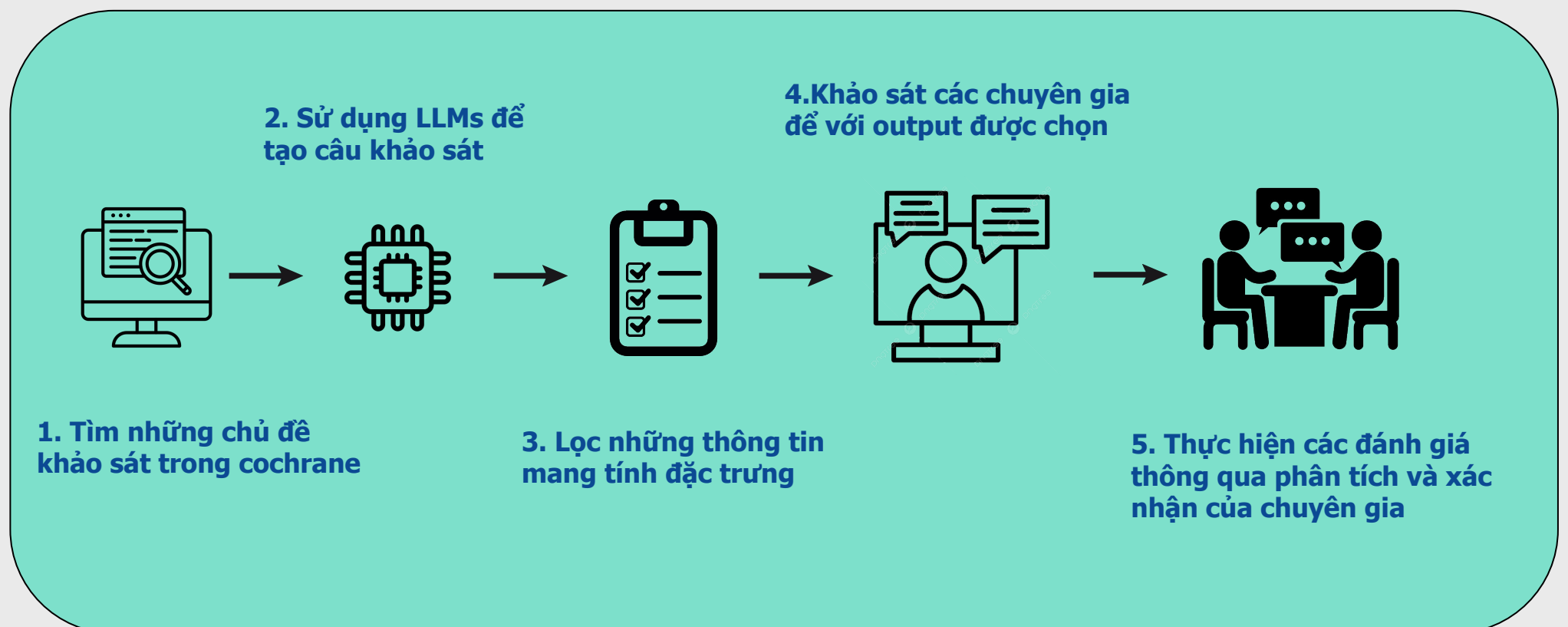
Chúng tôi giới thiệu quy trình để đánh giá tính chính xác của những câu trả lời mà mô hình LLMs đưa ra:

- Đánh giá tiềm năng và hạn chế của việc sử dụng mô hình LLMs trong y tế.
- Xác định chi tiết tính ứng dụng của LLMs trong lĩnh vực y tế, dựa trên kết quả đề xuất các biện pháp cải tiến để tối ưu hóa hiệu suất và giảm thiểu rủi ro khi triển khai LLMs.

Why ?

- Với sự phát triển vượt bậc của các mô hình ngôn ngữ lớn (LLMs) thì việc ứng dụng vào ngành y tế là hoàn toàn khả thi và có thể giúp tiết kiệm rất nhiều thời gian cho các quy trình xử lý trong y học. Tuy nhiên câu trả lời của mô hình có thể không chính xác và trong lĩnh vực y tế thì sự không chính xác có thể dẫn tới hậu quả tiêu cực về sau.

Overview



Description

1. Xây dựng bộ dữ liệu

- Sử dụng LLMs để tạo bộ dữ liệu ứng với mỗi chủ đề thu thập từ Cochrane Library (nơi đánh giá hệ thống trong lĩnh vực y học), chỉ triển khai prompt đơn giản nhằm mở rộng phạm vi của chủ đề và hướng tới người dùng không có kinh nghiệm trong y học.
- Tương tự, sử dụng LLMs để xây dựng bộ dữ liệu ứng với lĩnh vực chuyên sâu của mỗi chuyên gia bao gồm ít nhất một chủ đề.



2. Phân tích bộ dữ liệu

- Áp dụng phân tích định tính để xác định các loại lỗi và thuộc tính được coi là quan trọng bởi chuyên gia, xác định thủ công 6 mẫu output có nhiều đặc điểm được xác định trong quá trình phân tích dữ liệu.
- Chọn 2 trong 6 mẫu output trong phạm vi chuyên môn của các chuyên gia để tạo dữ liệu đánh giá.

3. Đánh giá độ hiệu quả

- Áp dụng phân tích chủ đề để mô tả các trường hợp cụ thể về tiềm năng và hạn chế của từng chủ đề trong bộ dữ liệu. Kết quả phân tích sẽ được thông qua các chuyên gia để xác nhận tính chính xác.

