BÁO CÁO: VIẾT PROMPT HIỆU QUẢ CHO CÁC TÁC VỤ HỌC TẬP

1. Giới thiệu

Mục tiêu của báo cáo này là thử nghiệm các cách viết prompt khác nhau để đánh giá mức độ ảnh hưởng của cấu trúc prompt đến chất lượng câu trả lời của mô hình ngôn ngữ lớn (LLM). Nghiên cứu tập trung vào ba tác vụ học tập phổ biến: tóm tắt tài liệu học thuật, giải thích khái niệm phức tạp và tạo bộ câu hỏi ôn tập.

2. Phân tích các tác vụ học tập

2.1 Tóm tắt tài liệu học thuật

- Mục tiêu: Rút gọn nội dung nhưng giữ lại các ý chính.

- Thách thức: AI có thể bỏ sót ý quan trọng hoặc tóm tắt quá dài/ngắn.

2.2 Giải thích khái niệm phức tạp

- Mục tiêu: Làm rõ các khái niệm khó hiểu.

- Thách thức: Câu trả lời có thể quá kỹ thuật hoặc thiếu ví dụ.

2.3 Tạo câu hỏi ôn tập

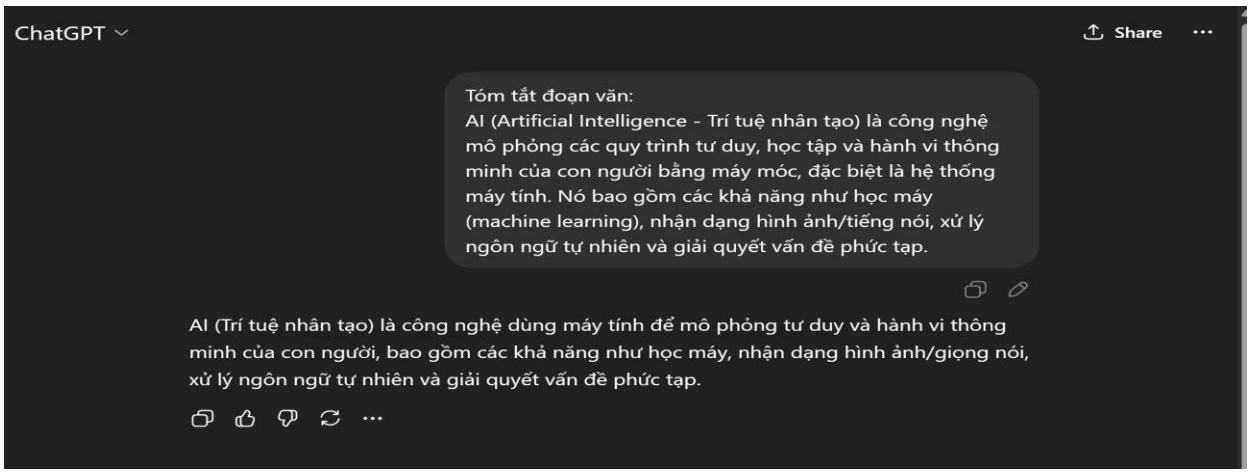
- Mục tiêu: Tạo câu hỏi giúp kiểm tra kiến thức.

- Thách thức: Câu hỏi có thể quá dễ hoặc thiếu cấu trúc.

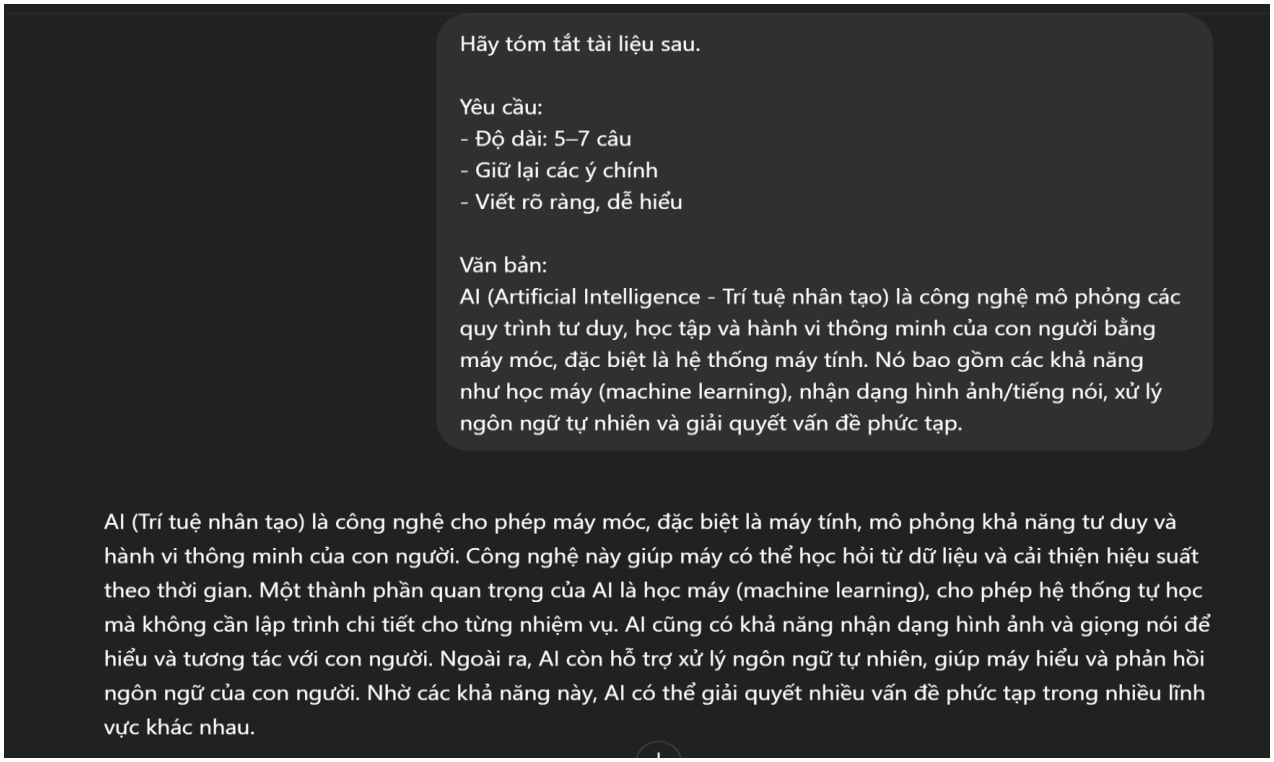
3. Thiết kế các phiên bản prompt

3.1 Tác vụ 1: Tóm tắt tài liệu

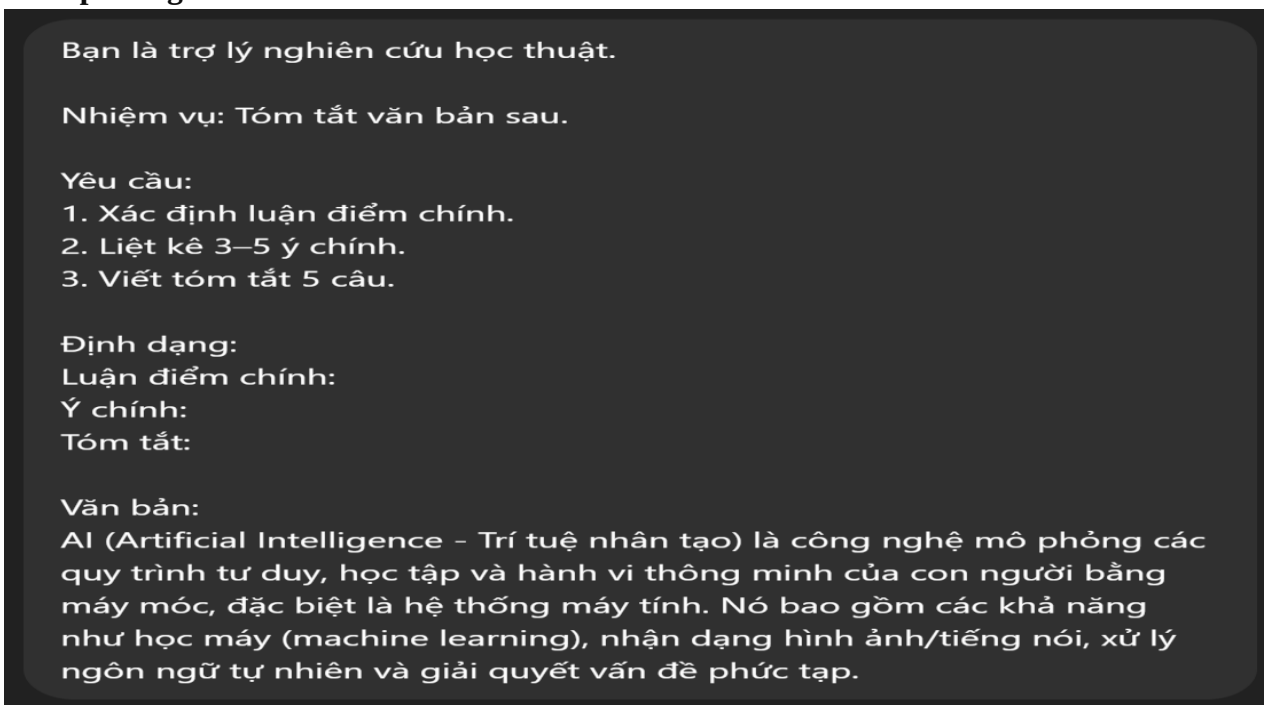
Prompt cơ bản:



Prompt cải tiến:



Prompt nâng cao:



Luận điểm chính:
AI là công nghệ cho phép máy móc, đặc biệt là máy tính, mô phỏng các quá trình tư duy, học tập và hành vi thông minh của con người.

Ý chính:

- AI mô phỏng khả năng tư duy và hành vi thông minh của con người bằng máy tính.
- Học máy (machine learning) cho phép hệ thống học từ dữ liệu và cải thiện hiệu suất.
- AI có khả năng nhận dạng hình ảnh và giọng nói.
- AI hỗ trợ xử lý ngôn ngữ tự nhiên để hiểu và tương tác với con người.
- AI có thể được sử dụng để giải quyết các vấn đề phức tạp.

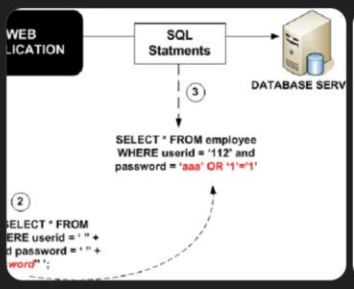
Tóm tắt:
AI là công nghệ cho phép máy tính mô phỏng các quá trình tư duy và hành vi thông minh của con người. Một thành phần quan trọng của AI là học máy, giúp hệ thống học từ dữ liệu và cải thiện hiệu suất theo thời gian. AI cũng có khả năng nhận dạng hình ảnh và giọng nói trong nhiều ứng dụng thực tế. Ngoài ra, công nghệ này hỗ trợ xử lý ngôn ngữ tự nhiên để máy hiểu và giao tiếp với con người. Nhờ những khả năng đó, AI có thể giải quyết nhiều vấn đề phức tạp trong nhiều lĩnh vực khác nhau.

Nhận xét: Prompt càng chi tiết thì câu trả lời nhận được sẽ càng rõ ràng và chính xác.

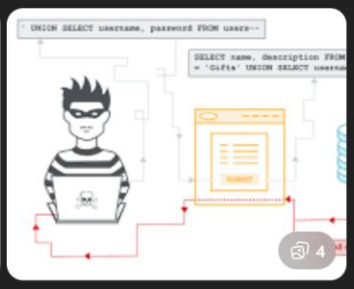
3.2 Tác vụ 2: Giải thích khái niệm

Prompt cơ bản:

Giải thích SQL Injection là gì.







SQL Injection là một kỹ thuật tấn công bảo mật web trong đó kẻ tấn công chèn (inject) các câu lệnh SQL độc hại vào dữ liệu đầu vào của ứng dụng. Khi ứng dụng không kiểm tra hoặc lọc dữ liệu đúng cách, các câu lệnh này được thực thi trực tiếp trên cơ sở dữ liệu.

Cơ chế hoạt động:

1. Ứng dụng web nhận dữ liệu từ người dùng (form login, URL, cookie, input field).
2. Ứng dụng ghép trực tiếp dữ liệu này vào câu truy vấn SQL.
3. Nếu dữ liệu chứa mã SQL độc hại, truy vấn bị thay đổi.
4. Cơ sở dữ liệu thực thi truy vấn đã bị sửa.

Ví dụ đơn giản

Ví dụ đơn giản

Truy vấn ban đầu:

```sql
SELECT * FROM users WHERE username = 'admin' AND password = '123';
```

Nếu ứng dụng không lọc dữ liệu, kẻ tấn công nhập:

```
username: admin
password: ' OR '1'='1
```

Truy vấn trở thành:

```sql
SELECT * FROM users
WHERE username = 'admin' AND password = '' OR '1'='1';
```

Điều kiện `'1'='1'` luôn đúng → đăng nhập thành công mà không cần mật khẩu.

Hậu quả

- Bypass login.
- Đọc dữ liệu nhạy cảm (user, password, credit ca
- Sửa hoặc xóa dữ liệu.

Prompt cải tiến:

Giải thích khái niệm SQL Injection cho sinh viên IT.

Yêu cầu:
- Giải thích đơn giản
- Có ví dụ minh họa
- Trình bày theo từng bước
- Không quá 200 từ

SQL Injection là một kỹ thuật tấn công vào ứng dụng web khi kẻ tấn công chèn câu lệnh SQL độc hại vào dữ liệu đầu vào. Nếu ứng dụng không kiểm tra dữ liệu đúng cách, câu lệnh đó sẽ được thực thi trên cơ sở dữ liệu.

Ví dụ:
Một trang đăng nhập tạo truy vấn SQL như sau:

```sql
SELECT * FROM users
WHERE username = 'admin' AND password = '123';
```

Các bước tấn công

Bước 1: Ứng dụng nhận dữ liệu từ người dùng
Người dùng nhập `username` và `password` vào form đăng nhập.

Các bước tấn công

Bước 1: Ứng dụng nhận dữ liệu từ người dùng
Người dùng nhập `username` và `password` vào form đăng nhập.

Bước 2: Ứng dụng ghép trực tiếp dữ liệu vào câu truy vấn SQL
Nếu lập trình viên nối chuỗi trực tiếp, truy vấn sẽ phụ thuộc vào dữ liệu người dùng nhập.

Bước 3: Kẻ tấn công chèn mã SQL
Ví dụ nhập mật khẩu:

```
' OR '1'='1
```

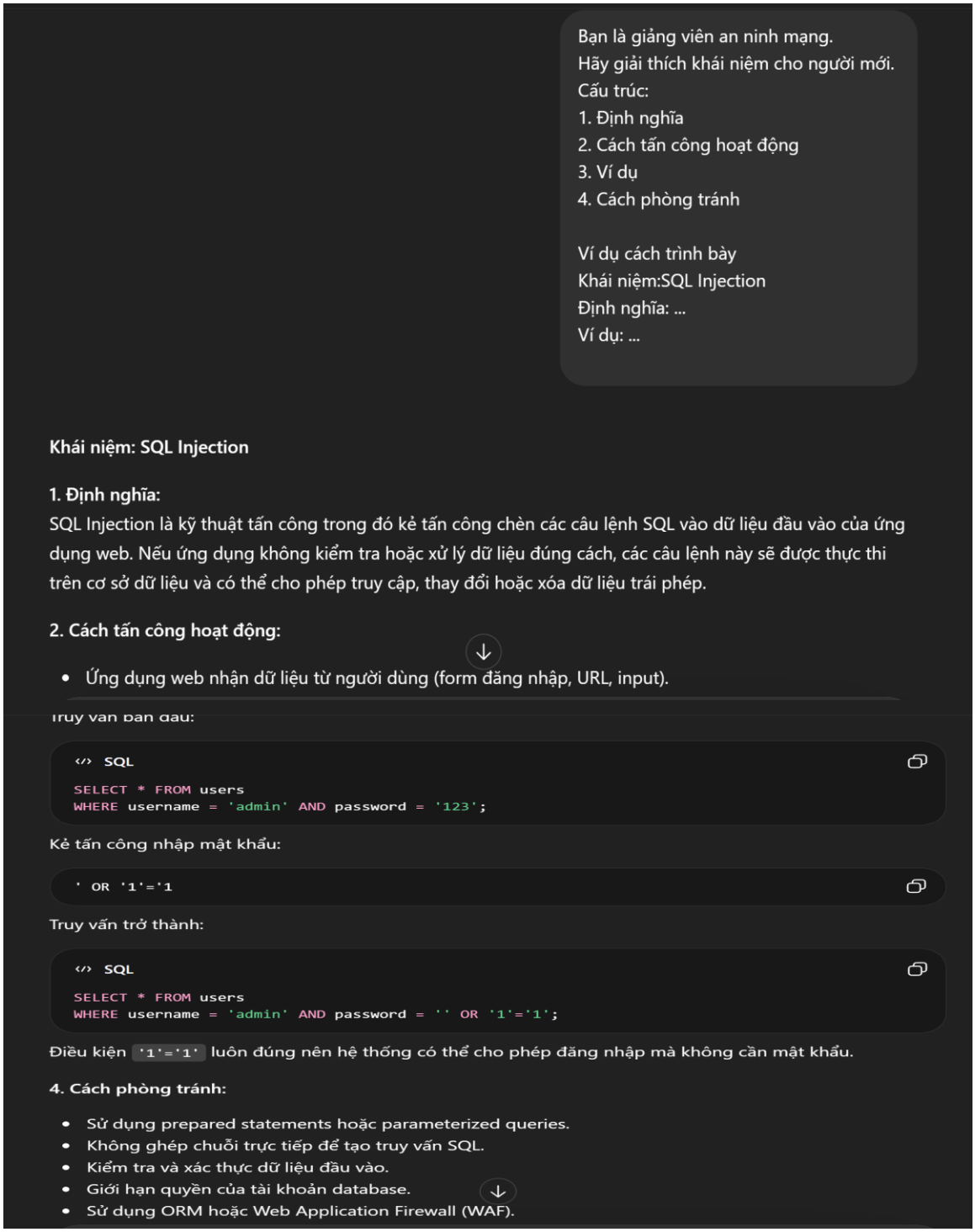
Bước 4: Truy vấn bị thay đổi

```sql
SELECT * FROM users
WHERE username = 'admin' AND password = '' OR '1'='1';
```

Điều kiện `'1'='1'` luôn đúng nên hệ thống có thể cho đăng nhập mà không cần mật khẩu.

Kết quả:
Kẻ tấn công có thể truy cập trái phép hoặc đọc, sửa, dữ liệu trong cơ sở dữ liệu.

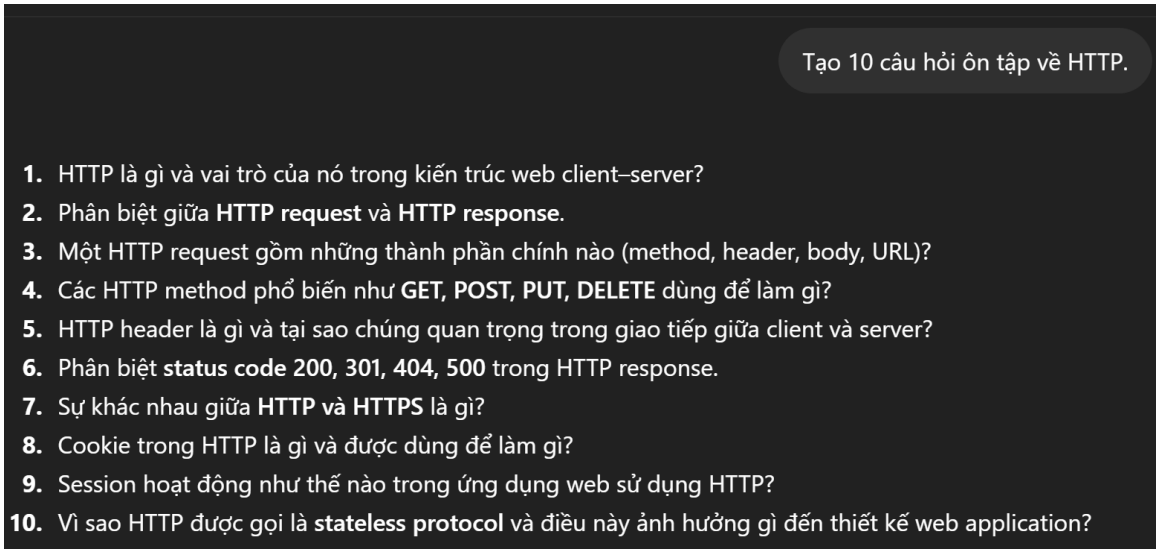
Prompt nâng cao:



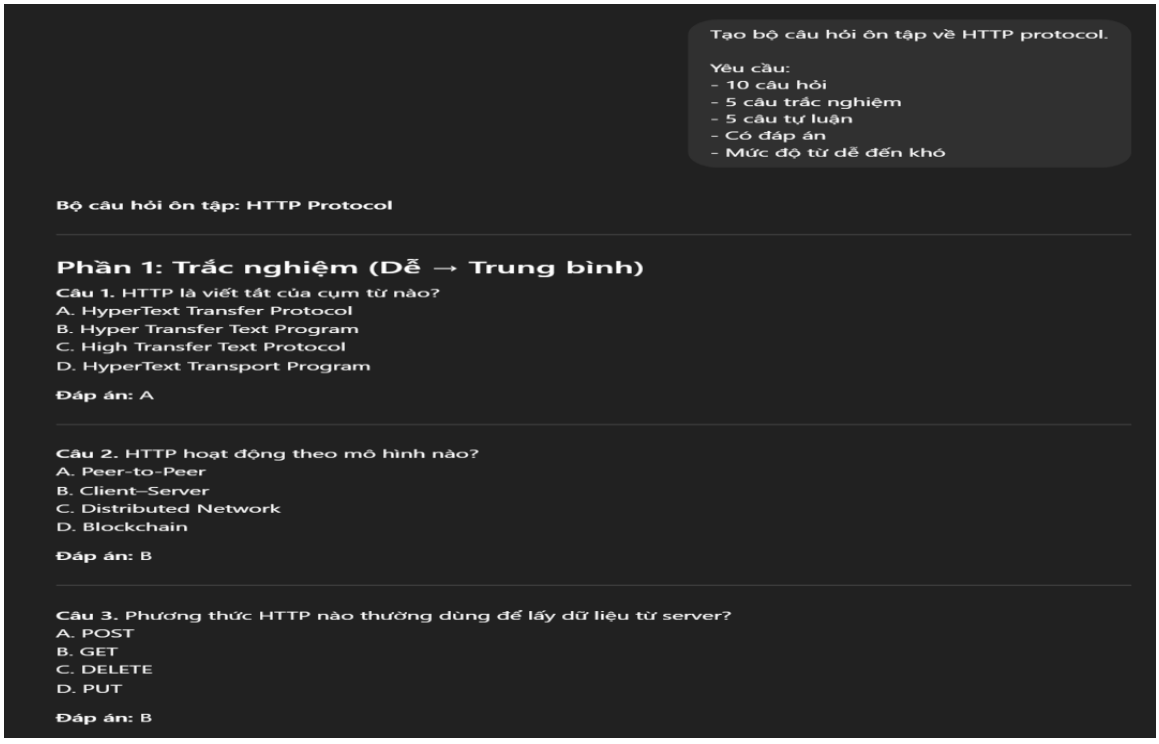
Nhận xét: Prompt càng chi tiết thì câu trả lời nhận được sẽ càng mạch lạc và dễ hiểu.

3.3 Tác vụ 3: Tạo câu hỏi ôn tập

Prompt cơ bản:



Prompt cải tiến:



Câu 4. Mã trạng thái HTTP nào biểu thị yêu cầu thành công?
A. 404
B. 200
C. 500
D. 301

Đáp án: B

Câu 5. HTTP được gọi là giao thức **stateless** vì:
A. Không sử dụng TCP
B. Không lưu trạng thái giữa các request
C. Không hỗ trợ header
D. Không có response

Đáp án: B

Phần 2: Tự luận (Trung bình → Khó)

Câu 6. Trình bày các thành phần chính của một HTTP Request.

Đáp án:
HTTP Request gồm:

- Request line (method, URL, HTTP version)
- Headers
- Blank line
- Body (tùy chọn, thường dùng với POST/PUT)

Câu 7. Giải thích vai trò của HTTP headers trong giao tiếp giữa client và server.

Đáp án:
HTTP headers chứa thông tin bổ sung về request hoặc response như loại nội dung (Content-Type), cookie, user-agent, authorization, cache control. Chúng giúp client và server hiểu cách xử lý dữ liệu.

Câu 8. Phân biệt GET và POST.

Đáp án:

- GET: dùng để lấy dữ liệu, tham số nằm trong URL, thường không thay đổi trạng thái server.
- POST: dùng để gửi dữ liệu lên server, dữ liệu nằm trong body, thường dùng để tạo hoặc thay đổi dữ liệu.

Prompt nâng cao:

Bạn là giảng viên đại học.

Hãy tạo một bộ câu hỏi ôn tập về giao thức HTTP với các yêu cầu sau:

Tổng cộng 12 câu hỏi
Chia thành 3 mức độ: dễ, trung bình, khó
Kết hợp câu hỏi trắc nghiệm và câu hỏi tự luận
Mỗi câu phải có đáp án

Định dạng đầu ra:

Bộ câu hỏi ôn tập: HTTP Protocol

Mức độ dễ

Câu 1 (Trắc nghiệm)
HTTP là viết tắt của cụm từ nào?
A. HyperText Transfer Protocol
B. Hyper Transfer Text Program
C. High Text Transmission Protocol
D. HyperText Transport Program

D. Port

Đáp án: B

Câu 11 (Tự luận)
HTTP response gồm những thành phần nào?

Đáp án:
HTTP response gồm:

- Status line (HTTP version + status code)
- Response headers
- Dòng trống
- Body chứa dữ liệu trả về (HTML, JSON, file...)

Câu 12 (Tự luận)
Giải thích sự khác nhau giữa HTTP và HTTPS.

Đáp án:
HTTP truyền dữ liệu dưới dạng văn bản thuần nên có thể bị nghe lén hoặc chỉnh sửa. HTTPS sử dụng TLS/SSL để mã hóa dữ liệu giữa client và server, giúp bảo vệ tính bảo mật, toàn vẹn và xác thực của dữ liệu khi truyền qua mạng.

4. Bảng so sánh kết quả

Prompt	Chất lượng nội dung	Cấu trúc câu trả lời	Độ rõ ràng
Cơ bản	Trung bình	Khó theo dõi	Trung bình
Cải tiến	Khá tốt	Có cấu trúc	Khá
Nâng cao	Rất tốt	Rõ ràng	Cao

5. Phân tích hiệu quả của prompt

- Prompt cơ bản: yêu cầu chung chung, thiếu ngữ cảnh và định dạng → kết quả dễ thiếu chi tiết hoặc không ổn định.

- Prompt cải tiến: có ràng buộc rõ ràng (độ dài, cấu trúc, đối tượng) → câu trả lời rõ ràng và có tổ chức hơn.

- Prompt nâng cao: sử dụng kỹ thuật như role prompting và cấu trúc từng bước → mô hình hiểu nhiệm vụ tốt hơn và tạo kết quả chính xác, logic hơn.

- Prompt có định dạng đầu ra giúp câu trả lời dễ đọc, nhất quán.

- Prompt càng cụ thể về mục tiêu và yêu cầu thì chất lượng kết quả càng cao.

6. Nguyên tắc viết prompt hiệu quả

- Cung cấp ngữ cảnh rõ ràng.

- Chỉ định format đầu ra.

- Đặt ràng buộc cụ thể.

- Sử dụng role prompting.

- Sử dụng structured instructions.

7. Kết luận

Tóm tắt kết quả thử nghiệm và khẳng định tầm quan trọng của việc viết prompt có cấu trúc trong việc khai thác hiệu quả các mô hình AI.