

**"How are LLMs affected by:"**

**- What about liability?:** Vấn đề trách nhiệm pháp lý khi LLMs tạo ra nội dung sai lệch hoặc gây hại vẫn đang được tranh luận và cần khung pháp lý rõ ràng.

**Hai mũi tên "Alignment?" trong sơ đồ:**

### 1. Từ "Max. probability of next word" đến "Goals" (của User):

**Vấn đề cốt lõi**: LLM hoạt động bằng cách chọn từ có xác suất cao nhất dựa trên dữ liệu huấn luyện

**Câu hỏi alignment**: Liệu cơ chế này có thực sự đáp ứng **mục tiêu thực tế của người dùng** không?

**Ví dụ**:

* **User muốn:** Thông tin chính xác về y tế
* **LLM làm:** Chọn câu trả lời "nghe có vẻ đúng" nhất từ internet (có thể sai)

**Vấn đề**: Xác suất cao ≠ Chính xác hoặc hữu ích cho user

### 2. Từ "Max. probability of next word" đến "Goals" (của Developers/owners):

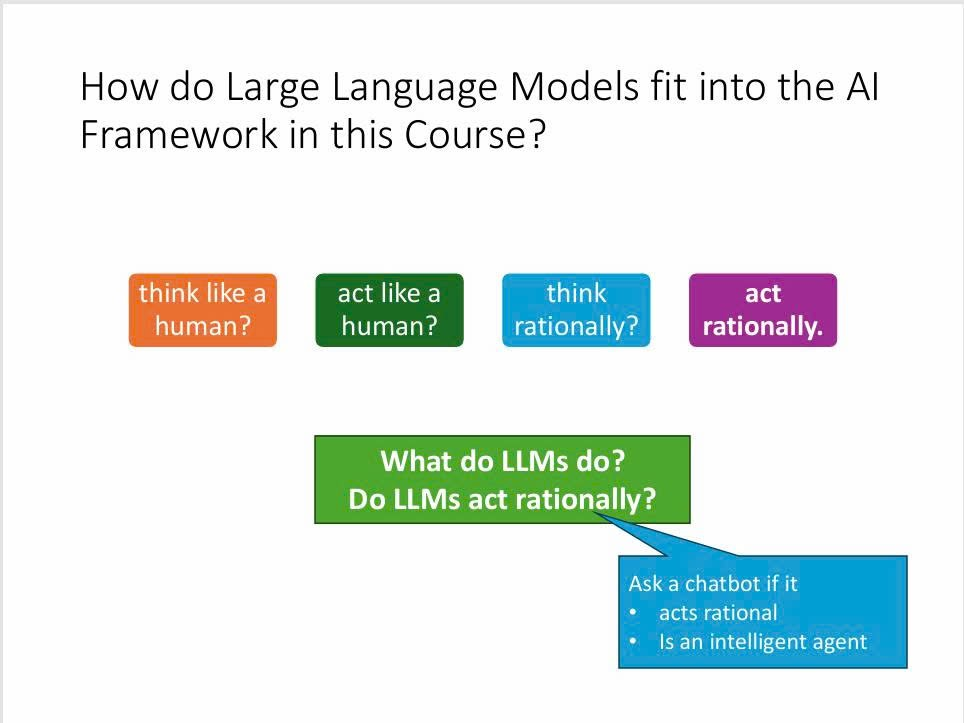
**Vấn đề cốt lõi**: Cùng cơ chế "chọn từ xác suất cao nhất"

**Câu hỏi alignment**: Liệu cách hoạt động này có phù hợp với **mục tiêu thiết kế của nhà phát triển** không?

**Ví dụ**:

* **Developer muốn:** AI an toàn, không tạo nội dung có hại
* **LLM làm:** Chọn từ dựa trên pattern trong data (có thể chứa nội dung có hại)

**Vấn đề**: Mục tiêu kỹ thuật (tối ưu xác suất) có thể xung đột với mục tiêu đạo đức/an toàn

****

**How do Large Language Models fit into the Al Framework in this Course?**

* **Think like a human?** - LLMs có thể mô phỏng cách suy nghĩ của con người thông qua việc xử lý ngôn ngữ tự nhiên, nhưng chúng không thực sự "suy nghĩ" như con người
* **Act like a human?** - LLMs có thể tương tác và phản hồi giống con người, tạo cảm giác như đang nói chuyện với người thật
* **Think rationally?** - LLMs có thể áp dụng logic và lý luận dựa trên dữ liệu đã học, nhưng không phải lúc nào cũng hoàn toàn hợp lý
* **Act rationally** - LLMs có thể đưa ra quyết định dựa trên thông tin có sẵn để đạt được mục tiêu tối ưu

**"What do LLMs do?"**

LLMs thực hiện các chức năng chính sau:

* **Xử lý và hiểu ngôn ngữ tự nhiên**
* **Tạo ra văn bản có ý nghĩa và mạch lạc**
* **Trả lời câu hỏi dựa trên kiến thức đã học**
* **Dịch thuật, tóm tắt, và phân tích văn bản**
* **Hỗ trợ các tác vụ sáng tạo như viết code, thơ, văn**

**"Do LLMs act rationally?"**

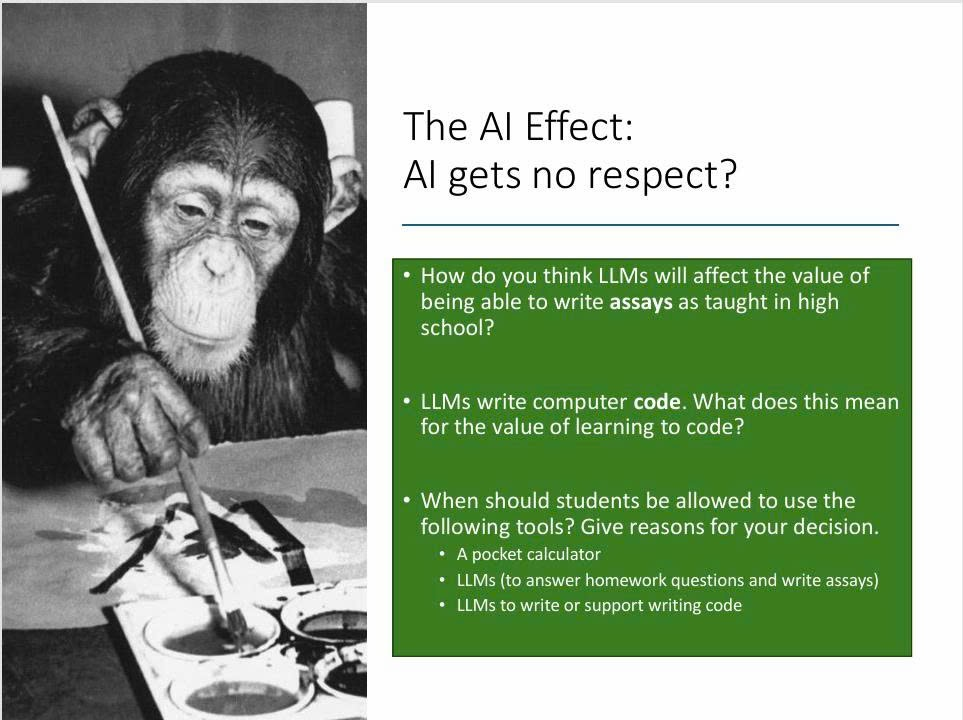
Đây là câu hỏi phức tạp:

**Có thể nói LLMs hành động hợp lý khi:**

* Chúng đưa ra câu trả lời dựa trên logic và dữ liệu
* Chúng cố gắng tối ưu hóa để đạt được mục tiêu được đặt ra
* Chúng có thể lý luận qua nhiều bước

**Nhưng LLMs không hoàn toàn hợp lý vì:**

* Chúng có thể tạo ra thông tin sai lệch (hallucination)
* Chúng bị ảnh hưởng bởi bias trong dữ liệu huấn luyện
* Chúng không có khả năng kiểm chứng thông tin thực tế
* Chúng có thể mâu thuẫn trong các phản hồi khác nhau



**"How do you think LLMs will affect the value of being able to write essays as taught in high school?"**

LLMs có thể làm giảm giá trị của việc viết luận truyền thống, nhưng sẽ tăng tầm quan trọng của:

* Tư duy phản biện
* Khả năng đánh giá và chỉnh sửa nội dung
* Sáng tạo và ý tưởng độc đáo
* Kỹ năng prompt engineering

**"LLMs write computer code. What does this mean for the value of learning to code?"**

Việc học lập trình vẫn quan trọng nhưng sẽ chuyển trọng tâm sang:

* Hiểu logic và kiến trúc hệ thống
* Debug và tối ưu hóa code
* Thiết kế thuật toán phức tạp
* Quản lý dự án và làm việc nhóm

**"When should students be allowed to use the following tools?"**

* **Máy tính bỏ túi:** Khi học các khái niệm toán học cấp cao, không phải tính toán cơ bản
* **LLMs (để trả lời bài tập và viết luận):** Như công cụ hỗ trợ brainstorming và chỉnh sửa, không thay thế hoàn toàn tư duy
* **LLMs để viết code**: Khi đã hiểu cơ bản về lập trình và có thể đánh giá chất lượng code