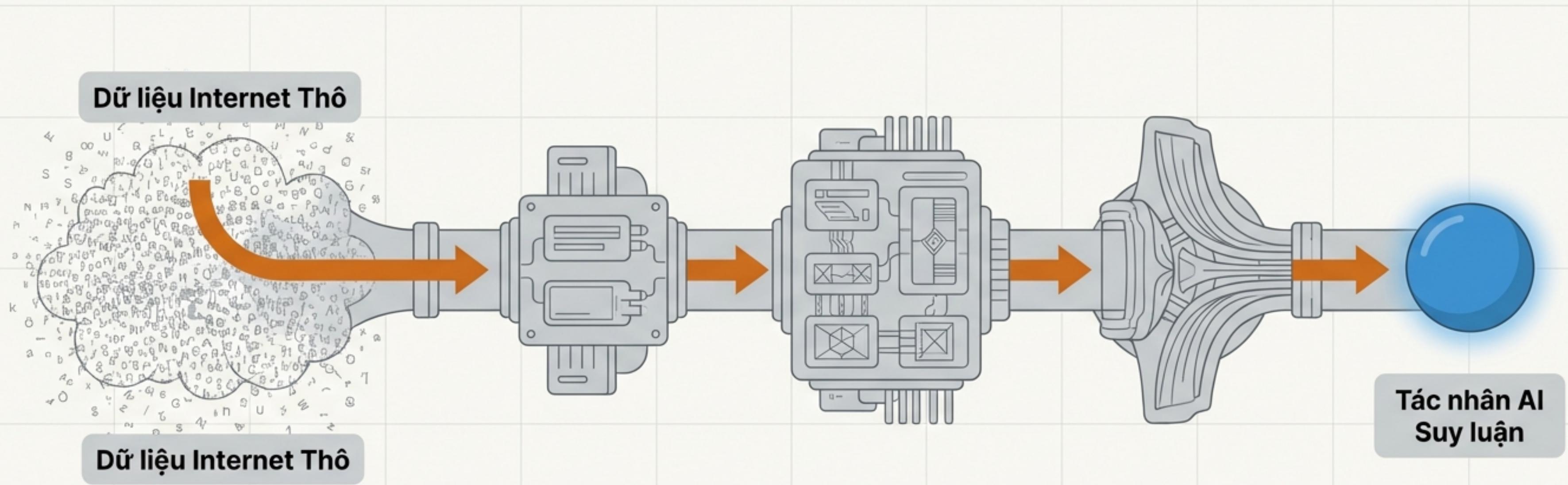


Từ Dữ Liệu Thô đến Trí Tuệ Suy Luận: Ba Giai Đoạn Hình Thành của một LLM

Một mô hình trực quan về cách xây dựng các hệ thống AI như ChatGPT, từ những khối văn bản thô trên Internet đến một tác nhân có khả năng suy luận phức tạp.

Chúng ta sẽ làm sáng tỏ quy trình ba giai đoạn để xây dựng một LLM: khám phá cách một mô hình học hỏi kiến thức, sau đó học cách trò chuyện, và cuối cùng là học cách suy nghĩ.

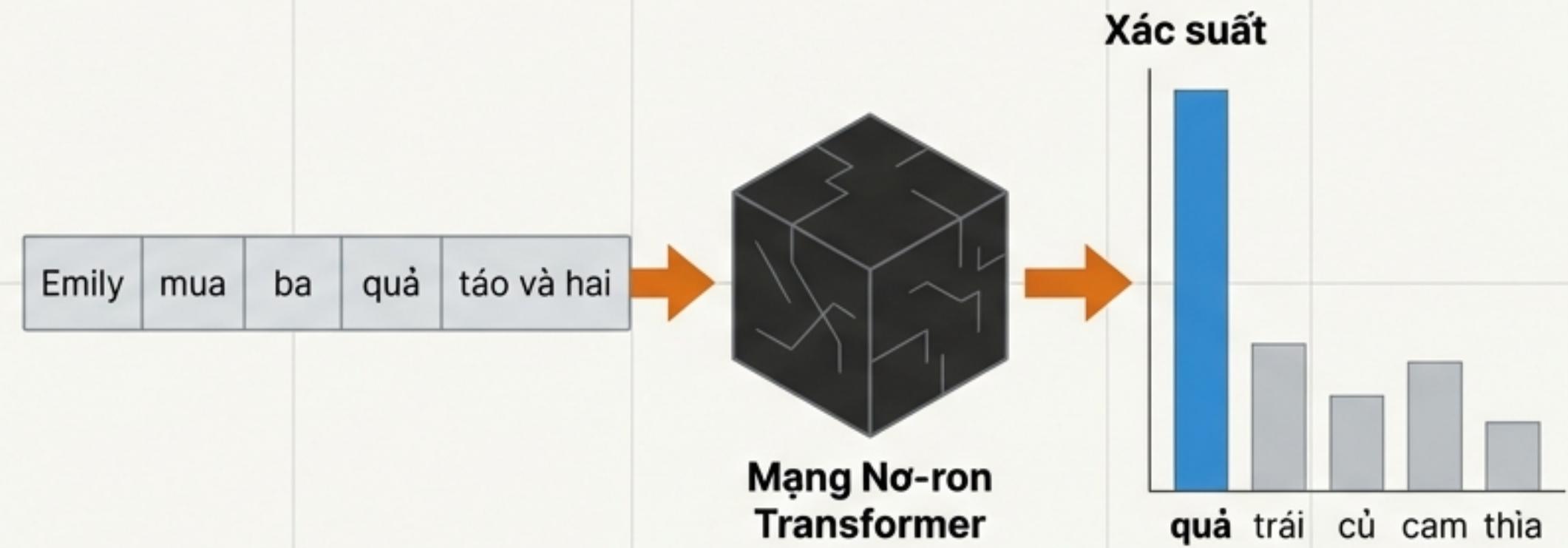


Nhiệm vụ cốt lõi của LLM: Dự đoán từ tiếp theo

Về cơ bản, một LLM là một “cỗ máy mô phỏng token” cực kỳ tinh vi. Nhiệm vụ duy nhất của nó là, với một chuỗi văn bản cho trước, dự đoán xác suất của từ (hay “token”) tiếp theo.

Toàn bộ khả năng—từ việc viết thơ đến giải toán—đều là kết quả phát sinh từ việc thực hiện nhiệm vụ đơn giản này ở một quy mô khổng lồ.

Mạng nơ-ron bên trong mô hình học các quy luật thống kê phức tạp từ dữ liệu để thực hiện dự đoán này. Trong suốt quá trình huấn luyện, chúng ta tinh chỉnh hàng tỷ “nút vặn” (tham số) của mạng để nó ngày càng dự đoán chính xác hơn.



Giai Đoạn 1: Huấn luyện Tiền kỲ - Xây dựng Nền tảng Tri thức

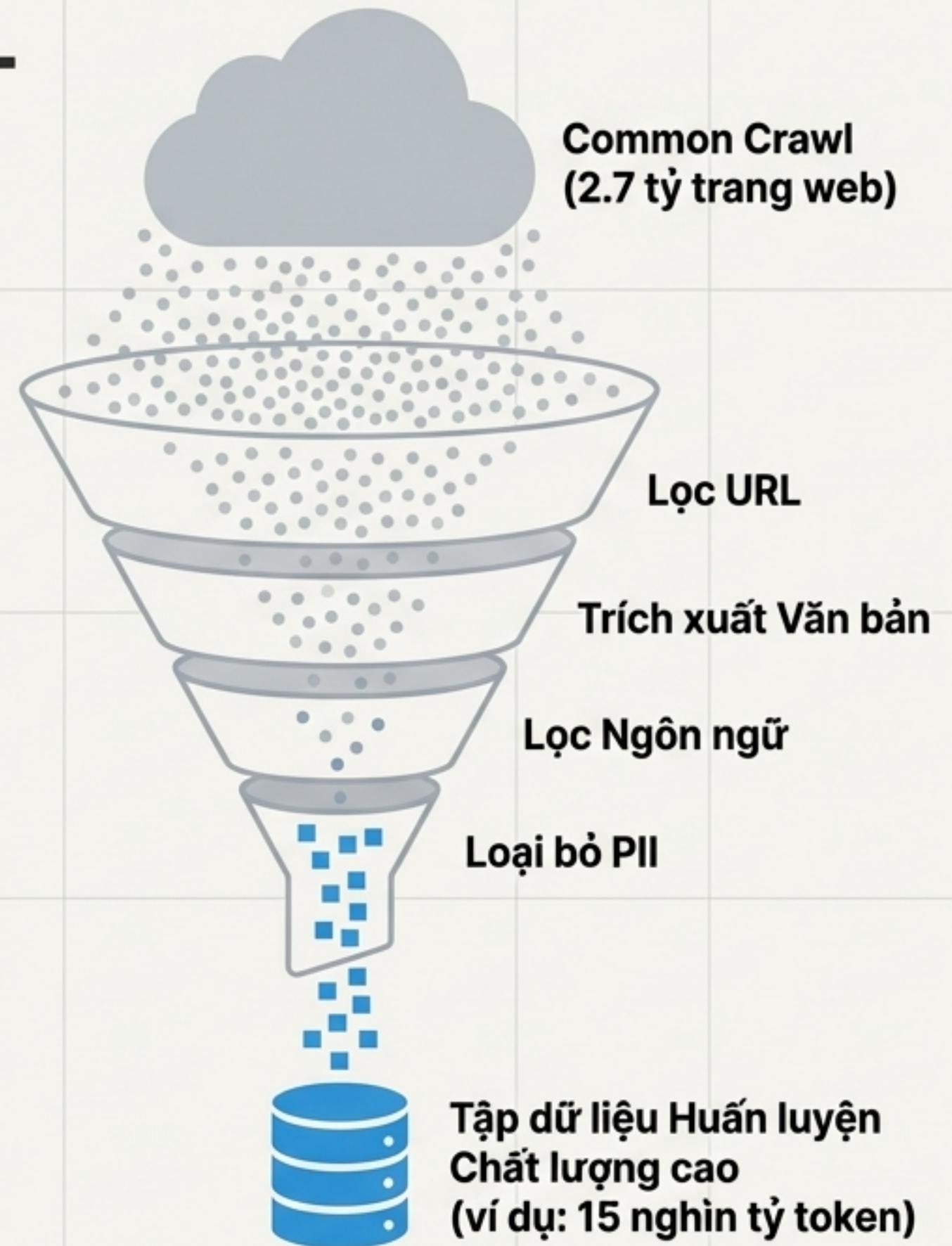
Thu thập Nguyên liệu từ Internet

Điểm khởi đầu là các kho dữ liệu web khổng lồ như Common Crawl, chứa hàng tỷ trang web. Mục tiêu là tạo ra một tập dữ liệu có **số lượng lớn, chất lượng rất cao**, và **đa dạng** về chủ đề. Ví dụ, tập dữ liệu FineWeb chứa khoảng 15 nghìn tỷ token (44 Terabytes), được chọn lọc kỹ lưỡng.

Quy trình Lọc và Làm sạch Chuyên sâu

Đây không phải là dữ liệu thô. Nó trải qua nhiều bước lọc nghiêm ngặt:

1. **Lọc URL**: Loại bỏ các trang web độc hại, spam, người lớn, và các tên miền trong danh sách đen.
2. **Trích xuất Văn bản**: Chỉ lấy nội dung văn bản chính, loại bỏ mã HTML, thanh điều hướng, quảng cáo.
3. **Lọc Ngôn ngữ**: Chọn lọc các tài liệu dựa trên ngôn ngữ (ví dụ: chỉ giữ lại các trang có >65% là tiếng Anh).
4. **Loại bỏ PII**: Phát hiện và xóa thông tin nhận dạng cá nhân như địa chỉ, số điện thoại.



Tokenization: Phân rã Ngôn ngữ thành các “Nguyên tử”

Mạng nơ-ron không xử lý ký tự hay từ, mà xử lý “token”. Token là các mẩu văn bản phổ biến.

Quá trình này được gọi là tokenization, chuyển đổi văn bản thô thành một chuỗi các ID số duy nhất. Mỗi ID đại diện cho một token cụ thể.

Ví dụ, GPT-4 sử dụng bộ từ vựng gồm 100.277 token duy nhất. Các token có thể là một từ, một phần của từ, hoặc thậm chí là một dấu câu có khoảng trắng.

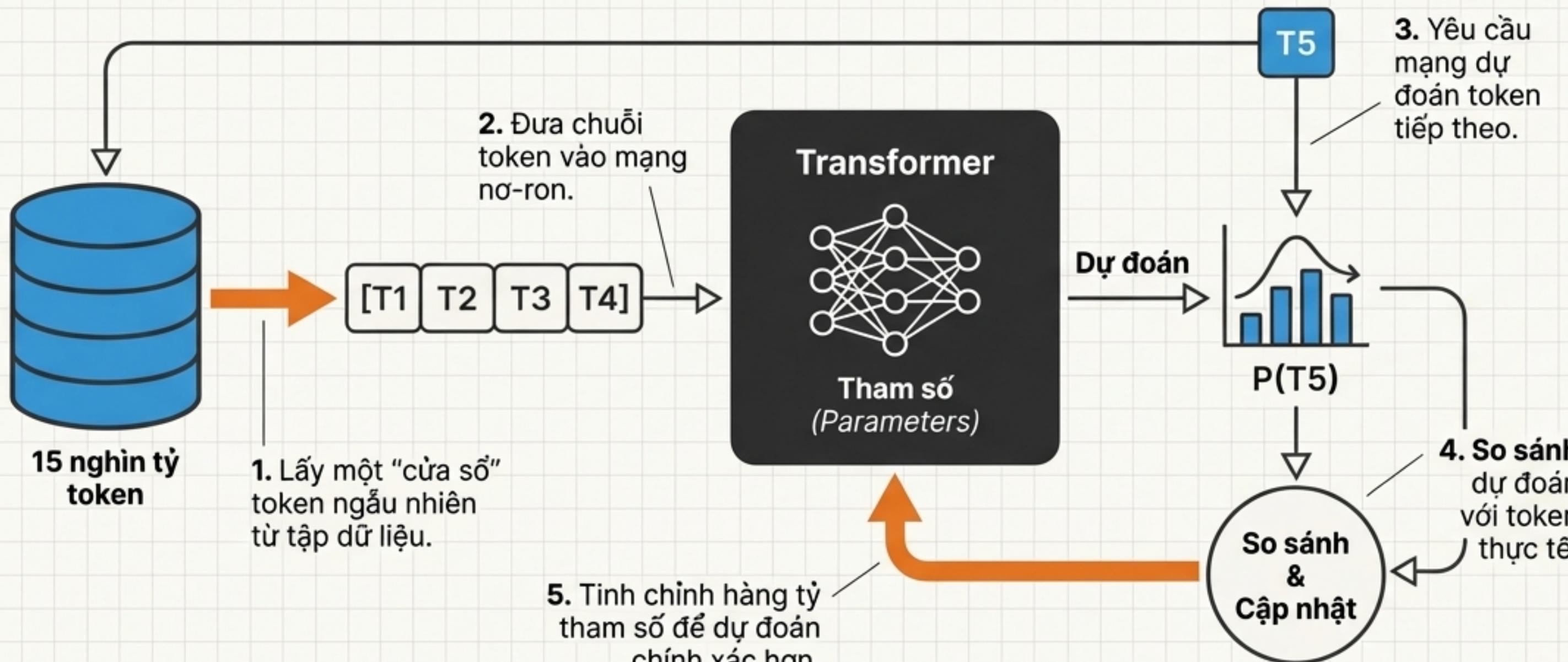
Cách văn bản được token hóa ảnh hưởng trực tiếp đến khả năng và “điểm mù” của mô hình.

`hello world` -> 2 token: `hello` (ID: 15339),
`world` (ID: 1917)
`Hello world.` -> 3 token: `Hello`, `world`, `.`

Mô hình ngôn ngữ lớn không thấy ký tự.

Mô_hình	ngôn_ngữ	lớn	không	thấy	ký_tự	.
34582	9812	1245	284	1356	56231	13

Huấn luyện Mô hình Nền tảng: Nén Toàn bộ Internet vào một Mạng Nơ-ron



Quy mô:

Lặp lại hàng nghìn tỷ lần. Đòi hỏi trung tâm dữ liệu khổng lồ với hàng chục nghìn GPU và tiêu tốn hàng triệu đô la.

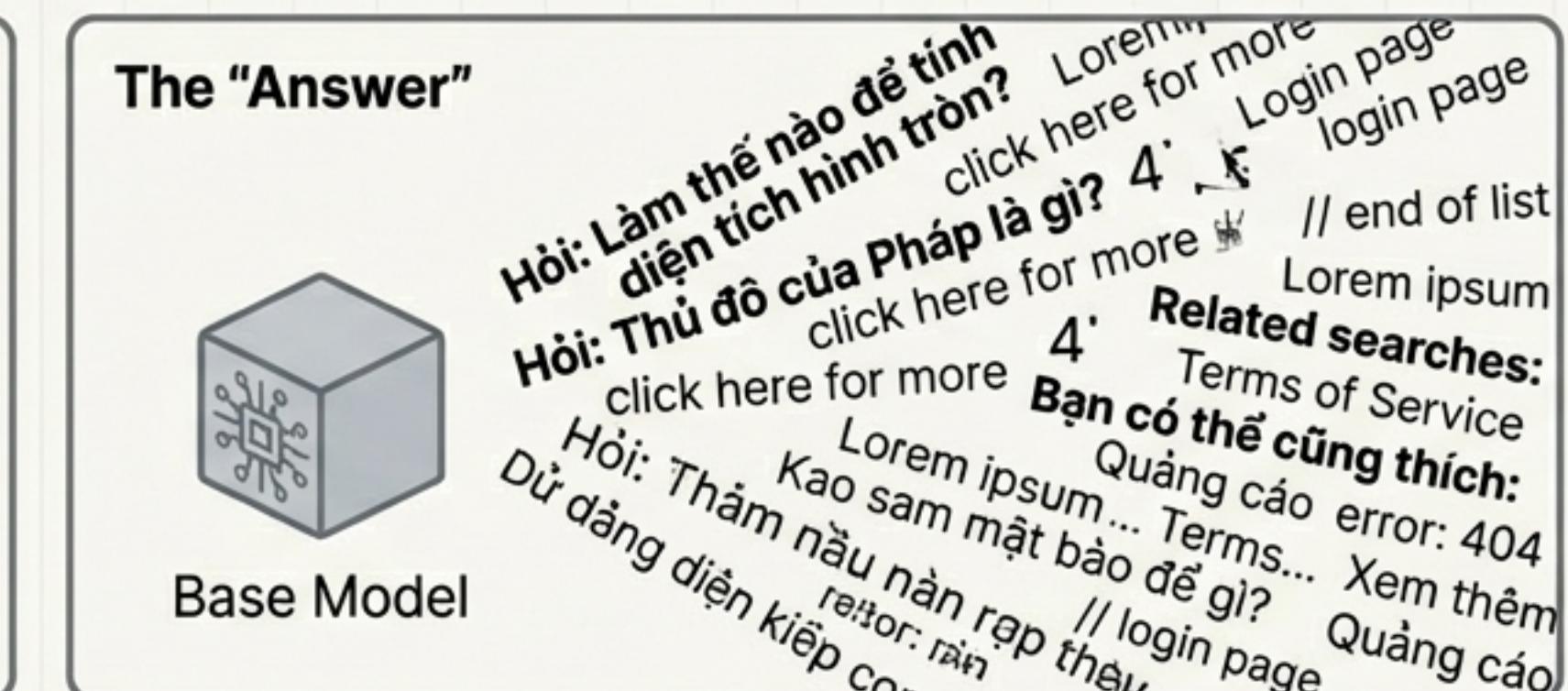
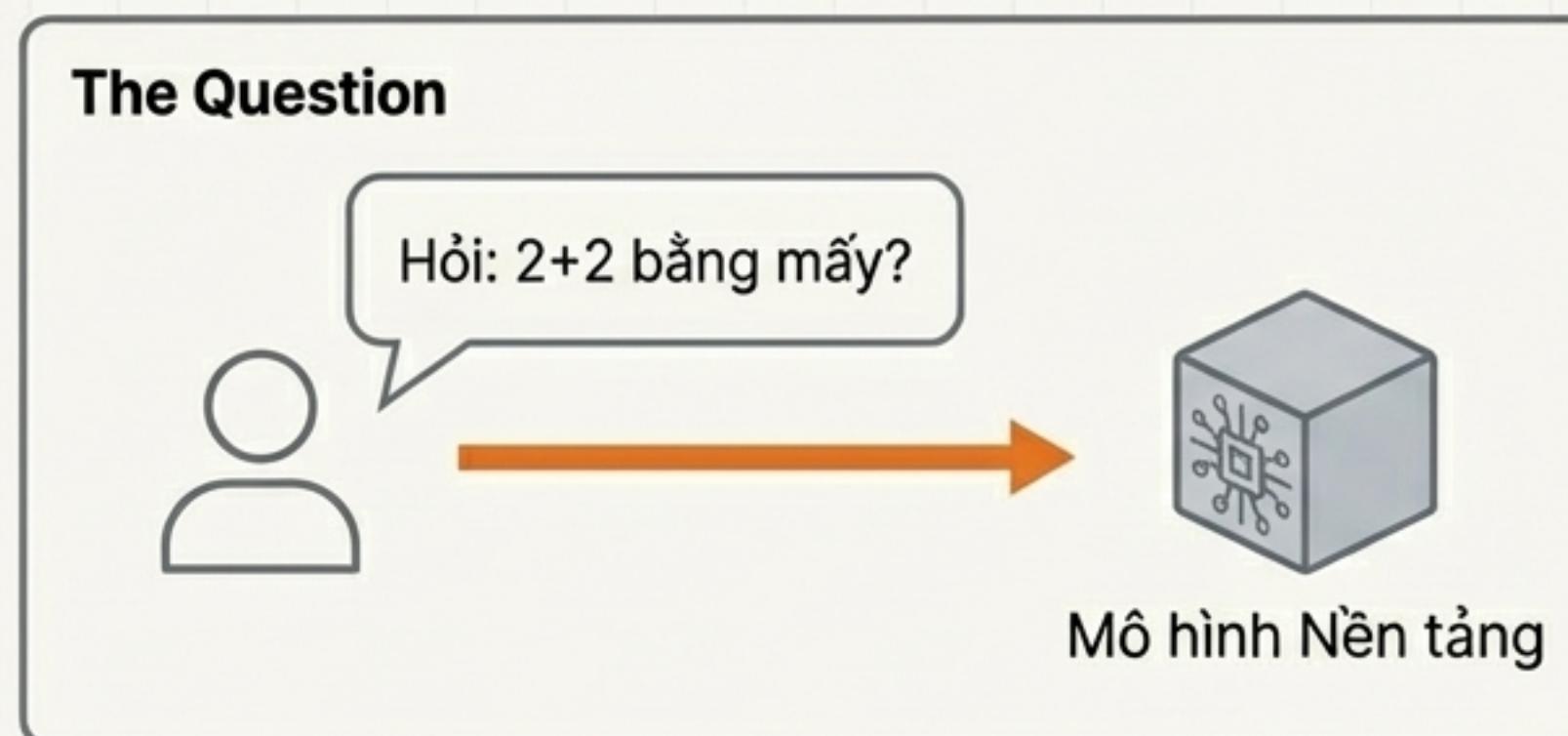
Kết quả:
Một **Mô hình Nền tảng (Base Model)**. Đây là một mạng nơ-ron có các tham số đã được "điều chỉnh" để chứa đựng kiến thức và các quy luật thống kê của ngôn ngữ.

Kết quả của Huấn luyện Tiền kỳ: Một “Cỗ máy Tự động Hoàn thành của Internet”

Vấn đề:

Mô hình Nền tảng không phải là một trợ lý. Nó chỉ là một cỗ máy mô phỏng văn bản trên Internet. Nó sẽ tiếp nối bất kỳ văn bản nào bạn đưa ra theo cách mà nó có khả năng xuất hiện nhất trên web.

Kết luận: Chúng ta có một mô hình rất uyên bác nhưng không hữu ích. Làm thế nào để dạy nó cách trò chuyện và tuân theo chỉ dẫn?



Giai Đoạn 2: Tinh chỉnh có Giám sát (SFT) - Dạy AI cách Trò chuyện

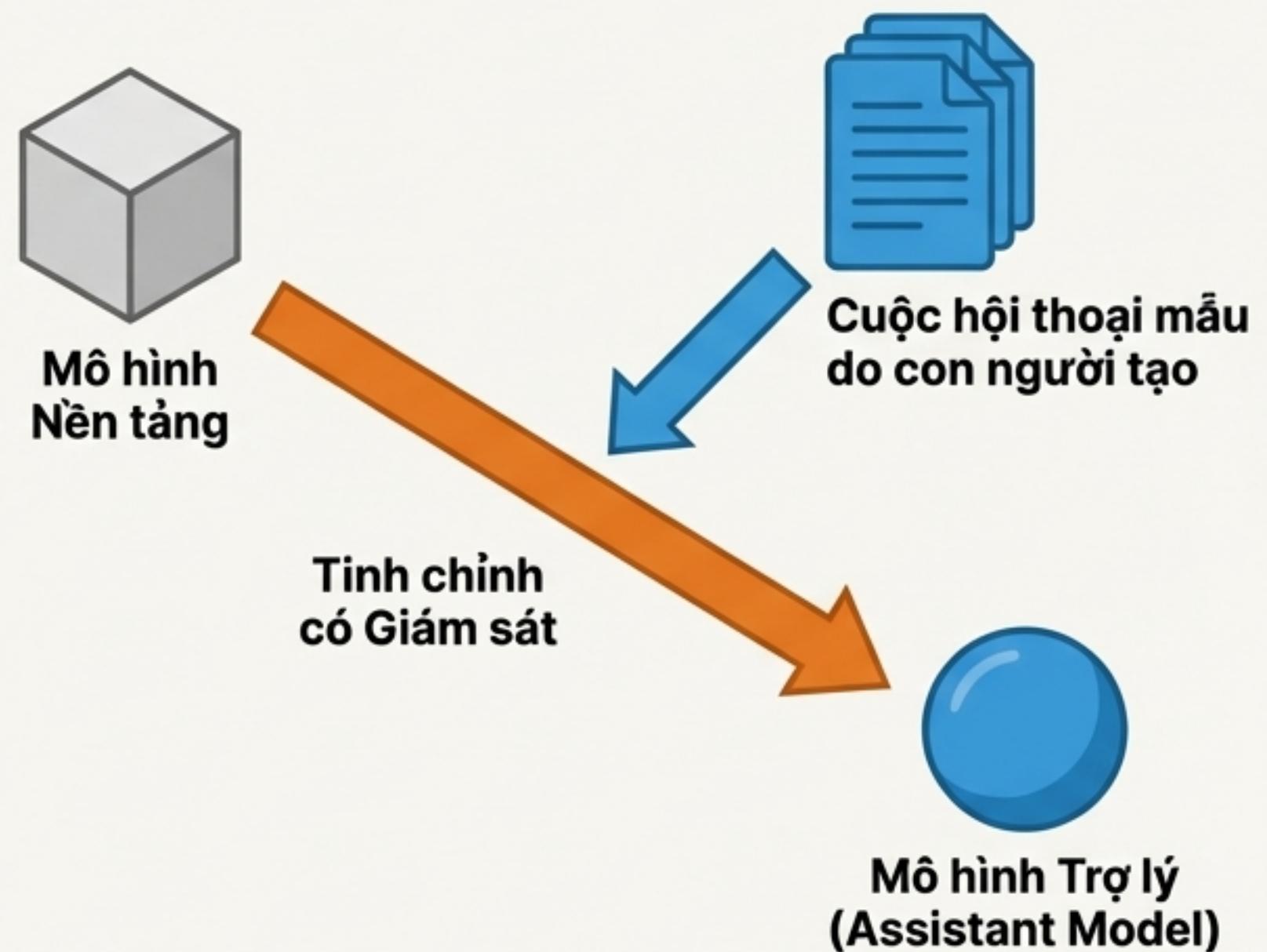
Giải pháp:

Chúng ta tiếp tục huấn luyện Mô hình Nền tảng, nhưng trên một tập dữ liệu hoàn toàn mới và khác biệt: một bộ sưu tập các cuộc hội thoại mẫu chất lượng cao.

Quy trình:

- Tạo Dữ liệu:** Các chuyên gia con người được thuê để viết ra hàng ngàn cuộc hội thoại. Với mỗi câu hỏi (prompt), họ viết ra câu trả lời "lý tưởng" của trợ lý.
- Chỉ dẫn:** Các chuyên gia tuân theo các chỉ dẫn chi tiết, ví dụ như câu trả lời phải "hữu ích, trung thực và vô hại".
- Tinh chỉnh (Fine-Tuning):** Mô hình Nền tảng được huấn luyện trên dữ liệu này, nhanh chóng học được khuôn mẫu của một trợ lý hữu ích.

Sự khác biệt về quy mô: Giai đoạn này nhanh hơn và rẻ hơn nhiều so với huấn luyện tiền kỳ (ví dụ: vài giờ so với vài tháng) vì tập dữ liệu nhỏ hơn đáng kể.



Kết quả của SFT: Một Trợ lý Hữu ích nhưng Dễ bị "Ảo giác"

Thành công:

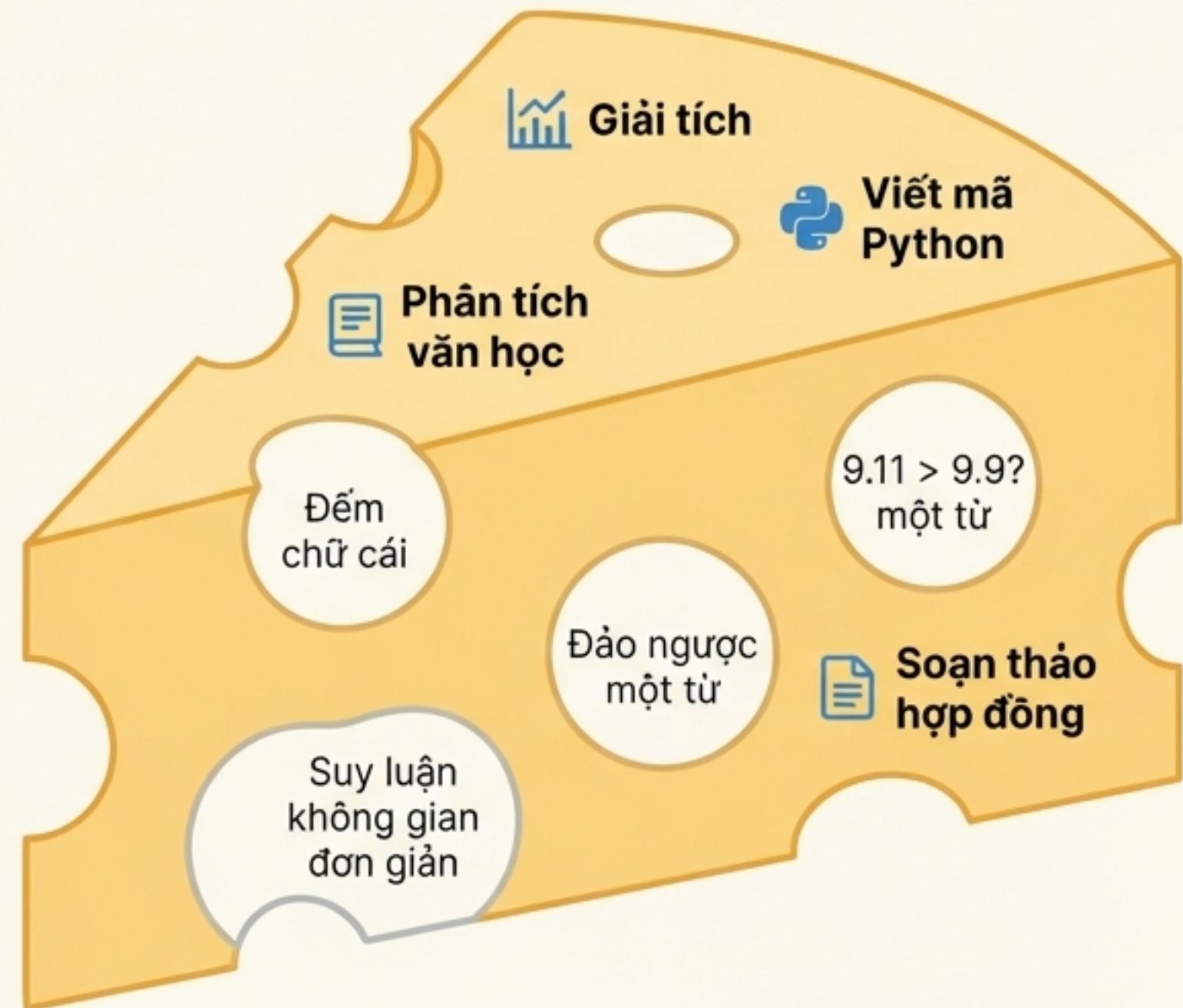
Mô hình giờ đây hoạt động như một trợ lý trò chuyện, trả lời các câu hỏi một cách hữu ích.

Vấn đề mới: Hiện tượng Ảo giác (Hallucination)

Khi được hỏi một câu mà nó không biết (ví dụ: Orson Kovats là ai?), mô hình sẽ không nói “Tôi không biết”. Thay vào đó, nó sẽ tự tin bịa ra một câu trả lời.

Tại sao? Vì nó đang bắt chước phong cách của dữ liệu huấn luyện. Trong dữ liệu SFT, các chuyên gia luôn trả lời một cách tự tin. Mô hình học theo phong cách này, ngay cả khi nó phải bịa ra thông tin. Nó là một “cỗ máy mô phỏng token”, không phải một cơ sở dữ liệu thực tế.

Mô hình tư duy: Kiến thức Lỗ chỗ như Phô mai Thụy Sĩ



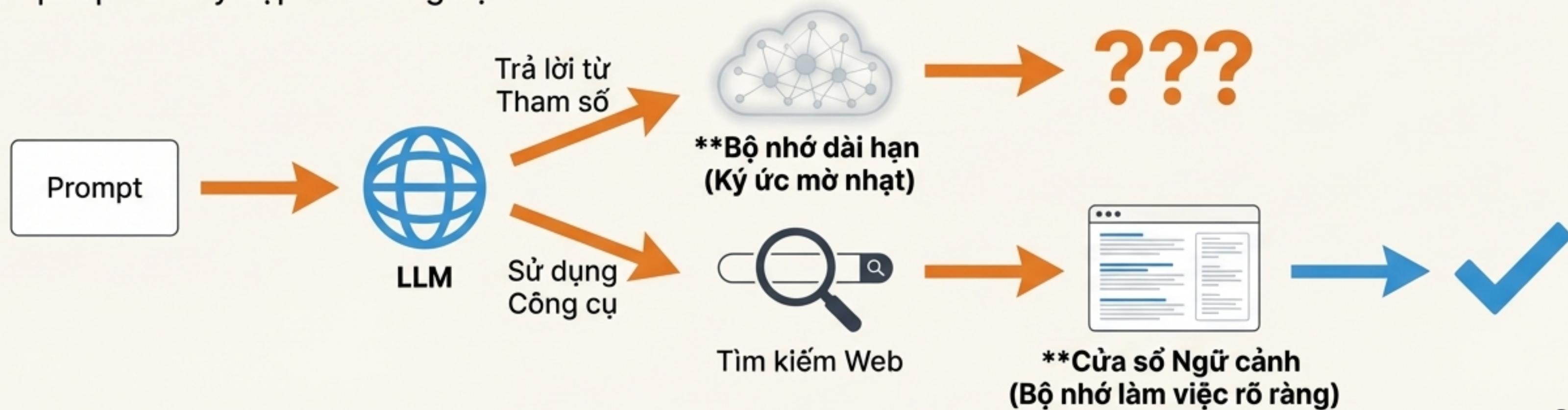
Các biện pháp Giảm thiểu: Dạy AI nói "Tôi không biết" và Cung cấp Công cụ

Biện pháp 1: Dạy mô hình nhận biết giới hạn của mình

Chúng ta có thể tự động "thăm dò" kiến thức của mô hình. Với những câu hỏi mà nó liên tục trả lời sai, chúng ta tạo ra các ví dụ huấn luyện mới nơi câu trả lời đúng là "Tôi xin lỗi, tôi không có thông tin về điều đó." Điều này dạy mô hình cách liên kết sự không chắc chắn bên trong với việc thể hiện ra bên ngoài.

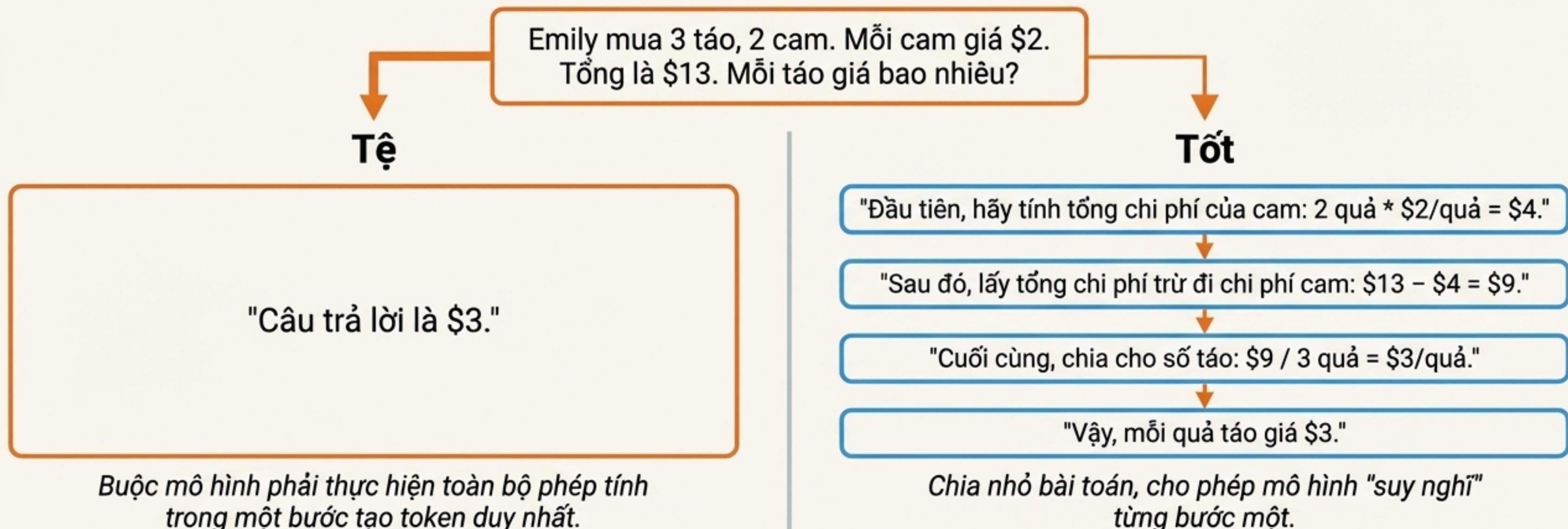
Biện pháp 2: Cung cấp cho mô hình các công cụ (Tools)

Con người khi không biết sẽ tìm kiếm trên Google. Chúng ta có thể cho LLM khả năng tương tự bằng cách cho phép nó truy cập các công cụ.



Một Hạn chế Sâu sắc hơn: Mô hình cần Token để "Suy nghĩ"

Việc tính toán trong một LLM diễn ra trên mỗi token được tạo ra. Mỗi token chỉ có một lượng tính toán hữu hạn. Ép mô hình trả lời ngay lập tức sẽ dẫn đến sai sót, giống như cố gắng giải một bài toán phức tạp trong đầu chỉ trong một giây.



Kết luận: Mô hình SFT có thể được dạy để bắt chước chuỗi suy luận của con người, nhưng nó có thể khám phá ra các chiến lược suy luận hiệu quả hơn cho chính nó không?

Giai Đoạn 3: Học tăng cường (RL) - Dạy AI cách *Suy nghĩ*

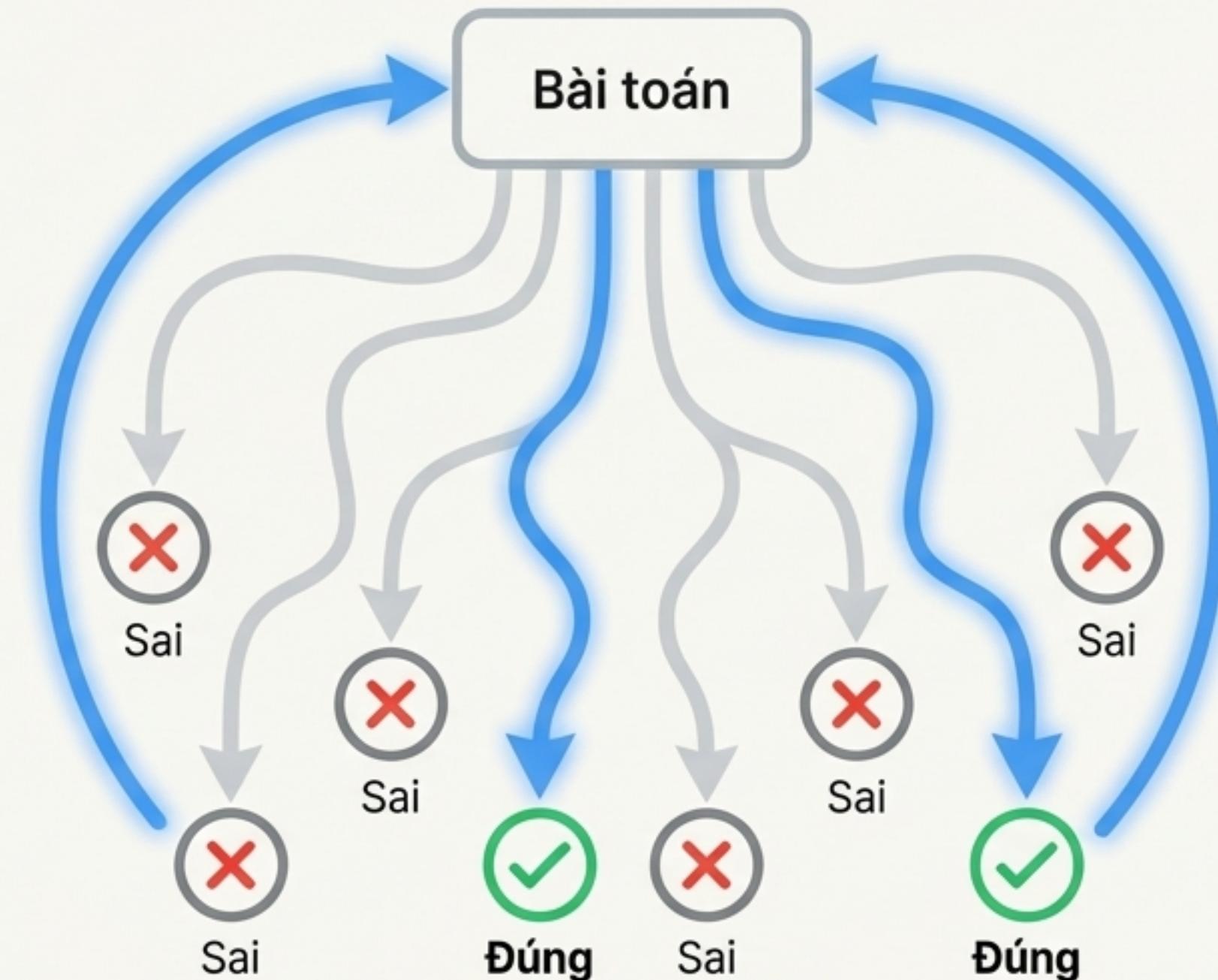
Phép ẩn dụ:

Sau khi đọc sách giáo khoa (Huấn luyện Tiền kỳ) và xem các ví dụ có lời giải (SFT), bây giờ là lúc làm các bài tập thực hành.

Quy trình:

- Thử nghiệm:** Với một vấn đề có đáp án có thể kiểm chứng, mô hình sẽ tạo ra hàng ngàn phương án giải quyết khác nhau.
- Kiểm tra:** Một chương trình tự động sẽ kiểm tra xem phương án nào dẫn đến đáp án đúng.
- Tăng cường:** Các chuỗi token (tức là "quá trình suy nghĩ") dẫn đến câu trả lời đúng sẽ được "thưởng". Mô hình được cập nhật để có nhiều khả năng tạo ra các chuỗi suy nghĩ tương tự trong tương lai.

Điểm cốt lõi: Mô hình không bắt chước giải pháp của con người. Nó tự khám phá ra những con đường suy luận hiệu quả nhất cho chính nó.



Sự trỗi dậy của Suy luận: "Độc thoại Nội tâm" của AI

Huấn luyện RL không chỉ giúp mô hình chính xác hơn, nó còn thay đổi cách mô hình giải quyết vấn đề. Mô hình RL thể hiện một “độc thoại nội tâm”, tự kiểm tra và tự sửa lỗi—một chiến lược nhận thức được phát hiện một cách tự phát để tối đa hóa sự chính xác.

Mô hình Trợ lý (SFT) - Bắt chước chuyên gia

Để giải quyết vấn đề này, chúng ta sẽ thực hiện các bước sau:

1. Tính tổng chi phí của cam:

$$2 \text{ quả cam} \times \$2/\text{quả} = \$4$$

2. Tìm tổng chi phí của táo:

$$\$13 \text{ (tổng)} - \$4 \text{ (cam)} = \$9$$

3. Tính chi phí mỗi quả táo:

$$\$9 / 3 \text{ quả táo} = \$3/\text{quả}$$

Câu trả lời là mỗi quả táo có giá \$3.

Mô hình Suy luận (RL) - Tự khám phá

Được rồi, để tôi nghĩ xem nào...

Chi phí cam là $2 * \$2 = \4 .

Tổng là \$13, vậy chi phí táo là $\$13 - \$4 = \$9$.

Có 3 quả táo, vậy mỗi quả là $\$9 / 3 = \3 .

Khoan đã, để tôi kiểm tra lại. $3 * \$3 + 2 * \$2 = \$9 + \$4 = \$13$. Đúng rồi.

Kết quả có vẻ chính xác.

Câu trả lời cuối cùng là \$3.

Vượt ra ngoài Sự bắt chước của Con người: Bài học từ AlphaGo

Sự tương đồng:

Sức mạnh của RL đã được chứng minh trong lĩnh vực cờ vây.

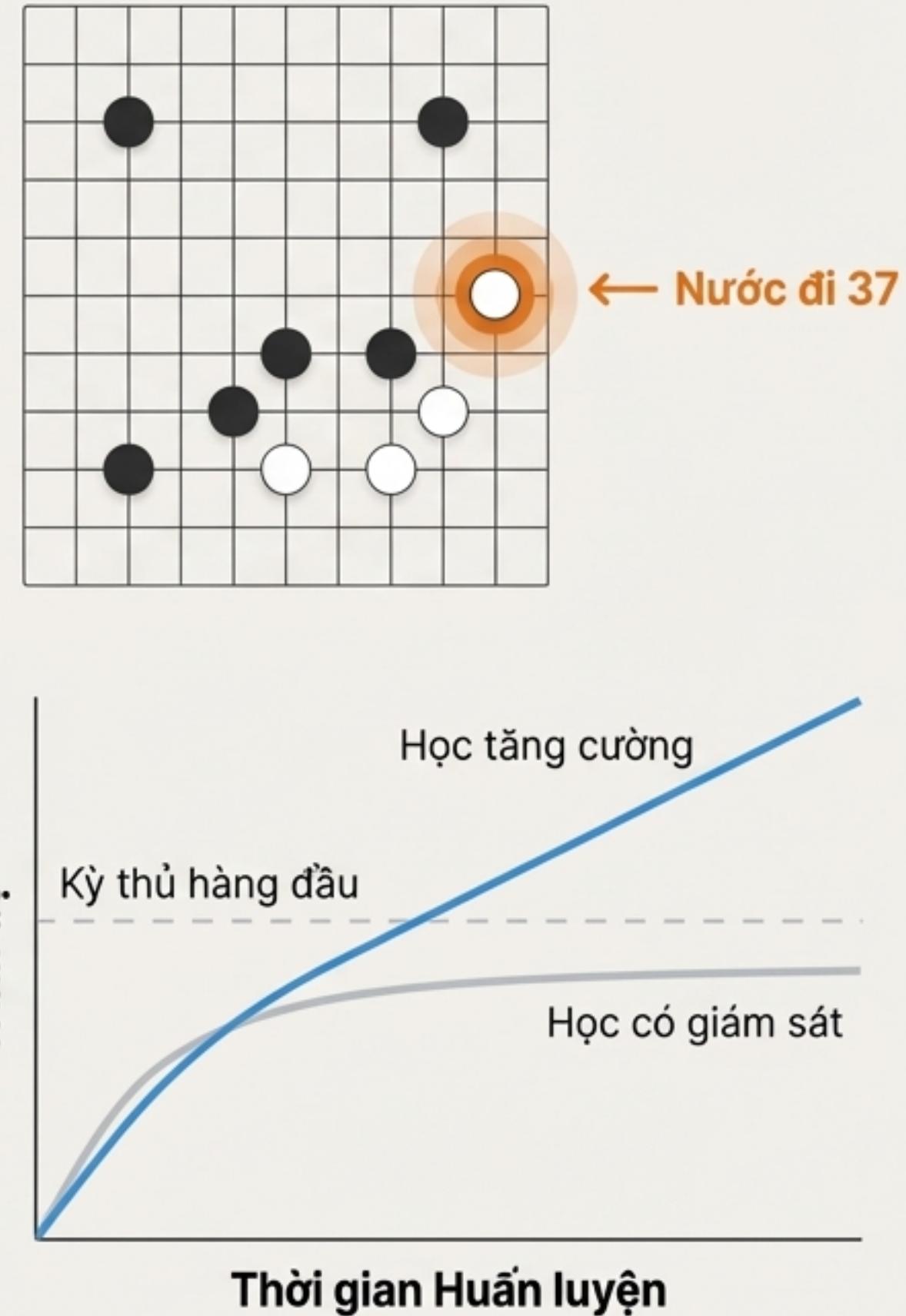
- ***Học có giám sát (SFT)**: Một phiên bản AlphaGo được huấn luyện để bắt chước các nước đi của những kỳ thủ hàng đầu. Nó trở nên rất giỏi, nhưng không bao giờ có thể vượt qua những người giỏi nhất.
- ***Học tăng cường (RL)**: AlphaGo chơi hàng triệu ván cờ với chính nó, tự khám phá các chiến lược dẫn đến chiến thắng. Nó không bị giới hạn bởi tư duy của con người.

Nước đi 37 (Move 37):

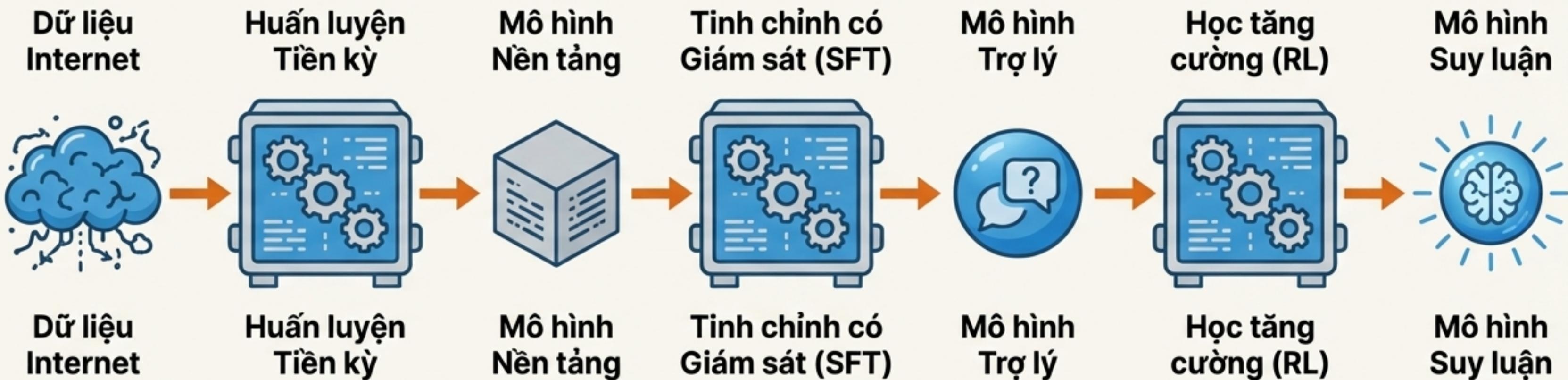
Trong trận đấu lịch sử với Lee Sedol, AlphaGo đã thực hiện một nước đi mà các chuyên gia con người cho là sai lầm (xác suất 1/10.000). Nhưng cuối cùng, đó lại là một nước đi thiên tài, mang tính quyết định.

Hàm ý cho LLM:

Tương tự, RL có khả năng cho phép LLM khám phá ra những “nước đi” trong suy luận—những chiến lược giải quyết vấn đề mà con người chưa từng nghĩ đến. Đây là con đường dẫn đến trí tuệ thực sự mới mẻ.



Tóm tắt: Ba Giai đoạn của một Mô hình Ngôn ngữ Lớn



Giai đoạn 1: Mô hình Nền tảng (Base Model)

- Quá trình: Tiếp thu kiến thức bằng cách "đọc" một phần lớn Internet.
- Kết quả: Một cỗ máy mô phỏng Internet uyên bác nhưng không hữu ích, không tuân theo chỉ dẫn.

Giai đoạn 2: Mô hình Trợ lý (Assistant Model)

- Quá trình: Học cách trò chuyện bằng cách bắt chước các cuộc hội thoại mẫu do chuyên gia con người tạo ra (SFT).
- Kết quả: Một trợ lý hữu ích, có khả năng tuân theo chỉ dẫn, nhưng có kiến thức "lõi chở" và suy luận nông.

Giai đoạn 3: Mô hình Suy luận (Reasoning Model)

- Quá trình: Học cách *suy nghĩ* bằng cách tự giải quyết các vấn đề và khám phá các chiến lược hiệu quả (RL).
- Kết quả: Một tác nhân suy luận mạnh mẽ, có khả năng thể hiện các quá trình tư duy phức tạp và đạt được độ chính xác siêu phàm.