

Introduction to Large Language Models

LÊ ANH CƯỜNG

NLP-KD Lab, Faculty of Information Technology,

Ton Duc Thang University

8.2024

Nội dung

I. Hiểu về ChatGPT và LLMs

1. Insights about ChatGPT
2. LLM và machine learning

II. Một số kỹ thuật nâng cao trong huấn luyện LLMs

1. RLHF: PPO và DPO
2. Synthetic data
3. Distillation

III. Thảo luận về sự phát triển LLMs và Generative AI

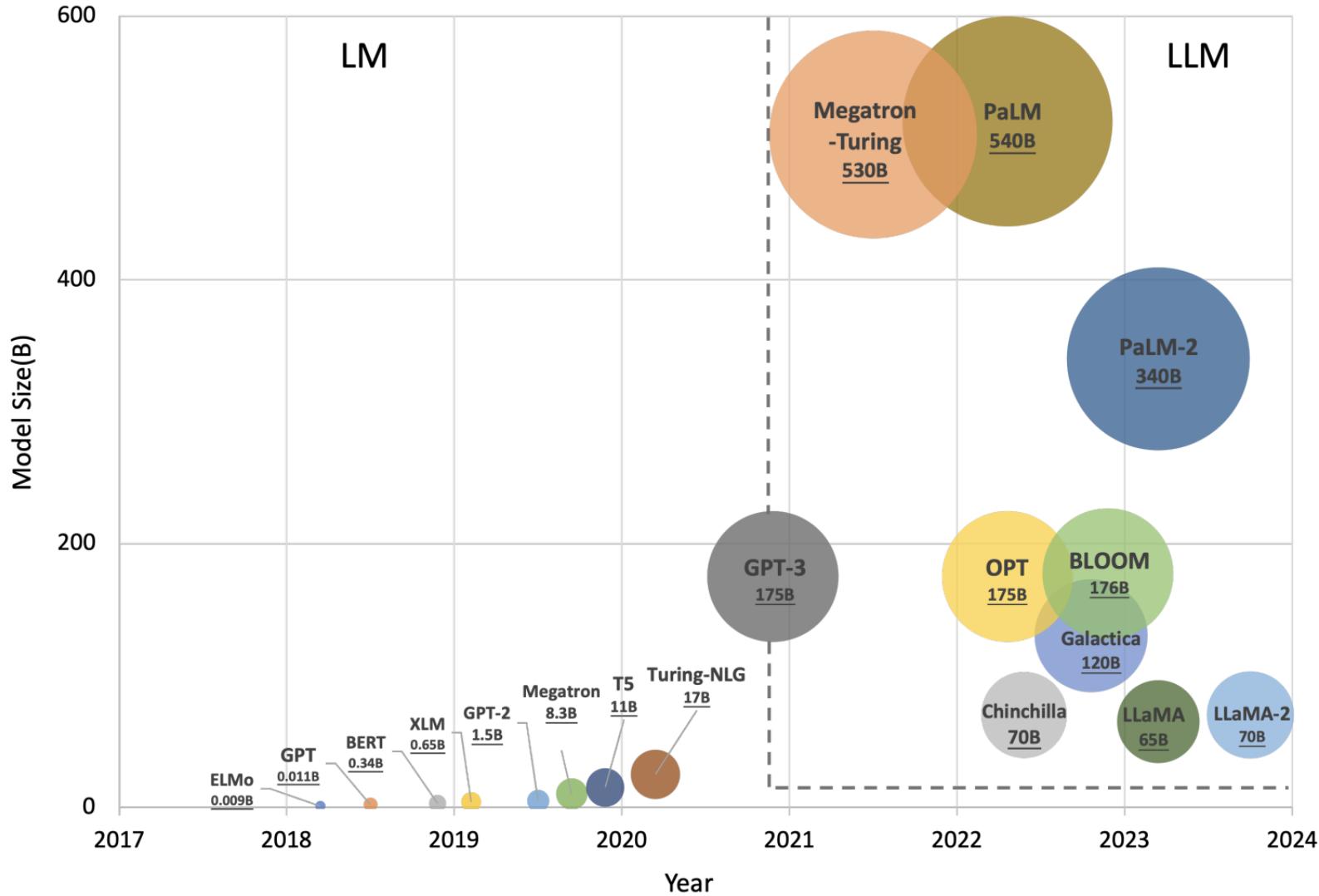
Các mô hình GPT của ứng dụng ChatGPT

Model	Parameters	Release date
GPT-1	117 million	2/2018
GPT-2	1.5 billion	2/2019
GPT-3	175 billion	6/2020
GPT-3.5	175 billion	11/2022
GPT-4	?	3/2023
GPT-4o	?	5/2024

Sự bùng nổ LLMs và ứng dụng

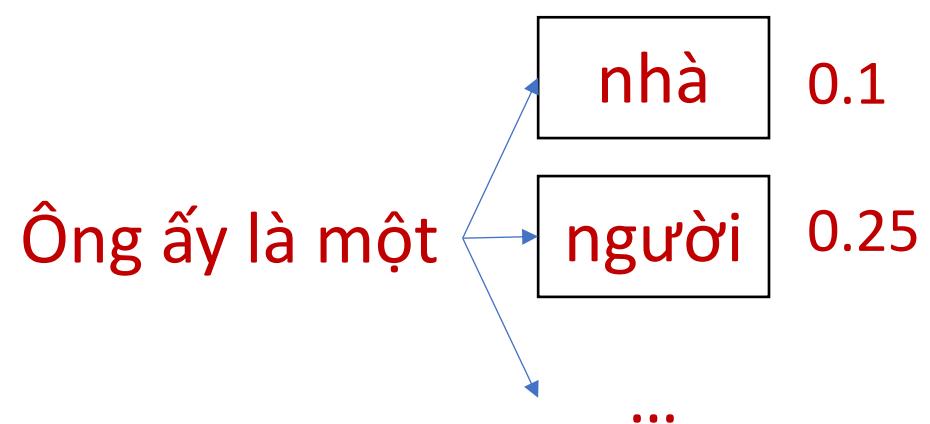
- Tất cả các công ty công nghệ lớn đều phát triển LLM như: Google ([PaLM](#), [Gemini](#), [Gemma](#)), Meta ([Llama-3](#)), Microsoft ([Phi-3](#)), Alibaba Cloud ([Qwen-2](#)), NVIDIA ([Megatron-Turing](#)).
- Các nhóm phát triển như, Perplexity AI (2022): [Perplexity](#); Anthropic (2021): [Claude 3.5 Sonnet](#); Mistral AI: [mixtral- 8x22b](#) (2023), dự án BigScience: [Bloom](#); ...

Kích thước LLMs

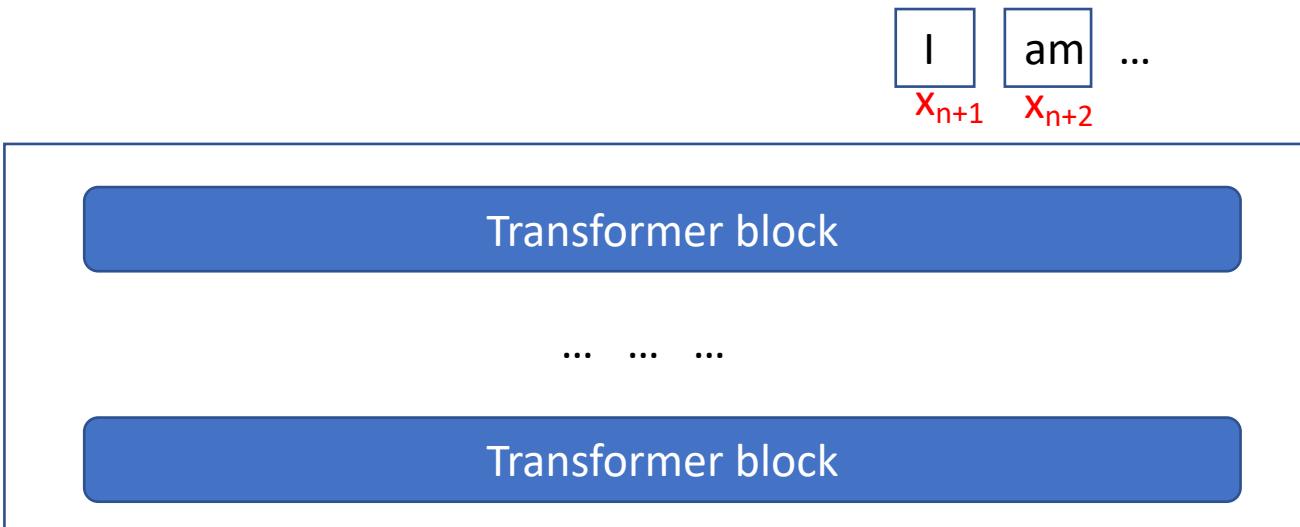


Mô hình ngôn ngữ (Language Model)

- ChatGPT là một mô hình ngôn ngữ lớn
- LM là mô hình cho phép tính xác suất xảy ra của một câu:
 $P(w_1 w_2 \dots w_n)$
- LM cho phép tính xác suất sinh ra một từ với điều kiện biết các từ trước nó:
 $P(w_k | w_1 \dots w_{k-1})$



LLM using GPT (Generative Pretrained Transformer)



“Hãy dịch sang tiếng Anh: Tôi đang ...”

$x_1 x_2 \dots$

x_{n+1}

x_n

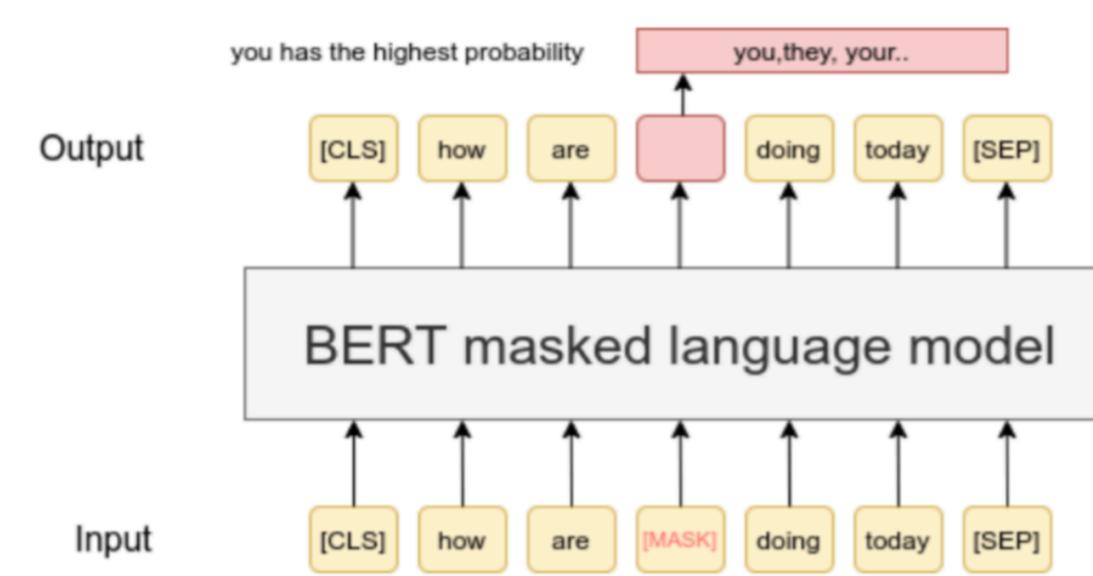
ChatGPT: Multi-Task (Đa nhiệm)



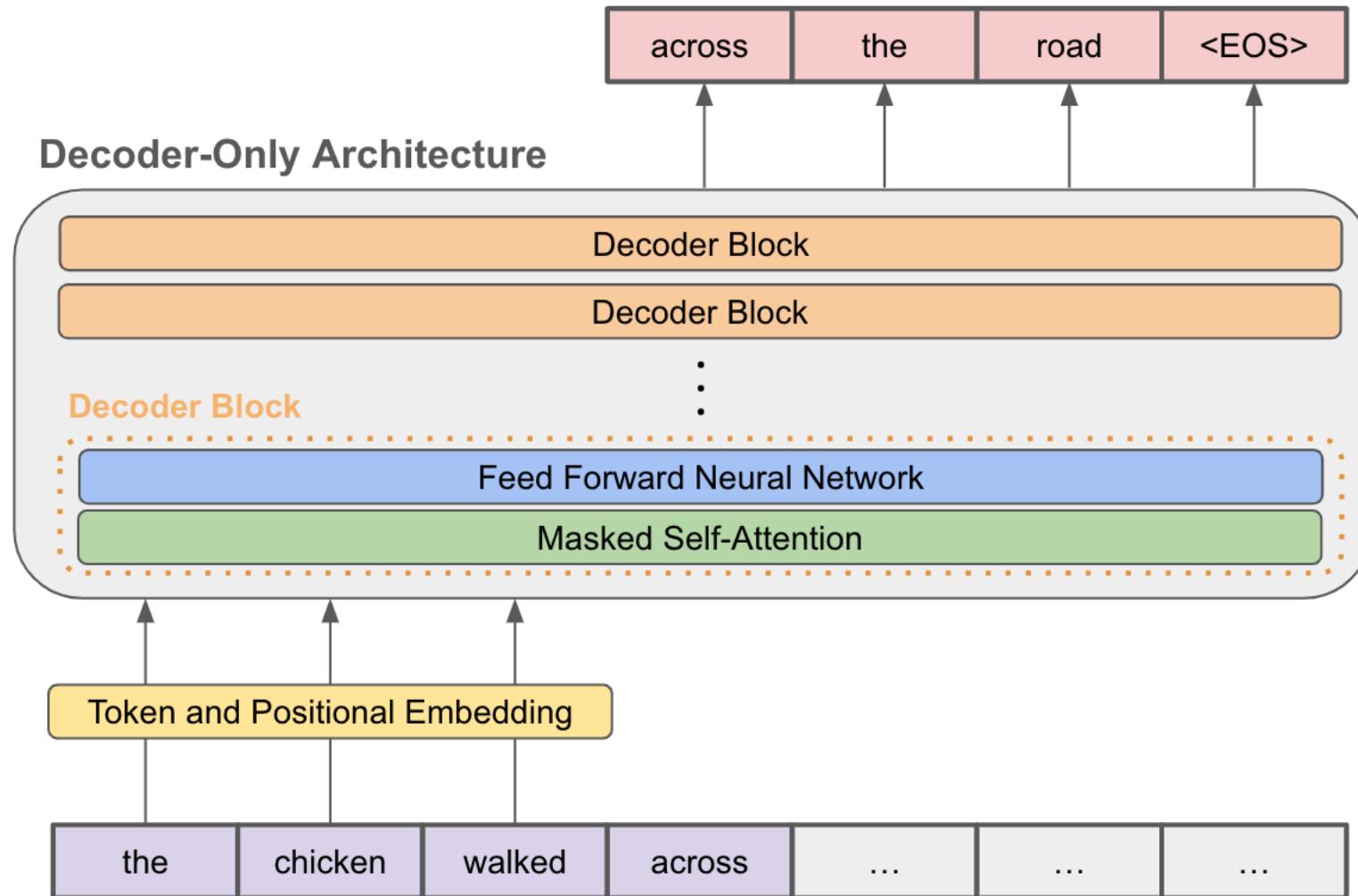
- Machine Translation
- Summarization
- Information Extraction
- Question Answering
- Coding
-

Bản chất của Huấn luyện mô hình ngôn ngữ

- Học biểu diễn từ (word representation)
- Học sinh từ dựa vào ngữ cảnh (conditional word generation)

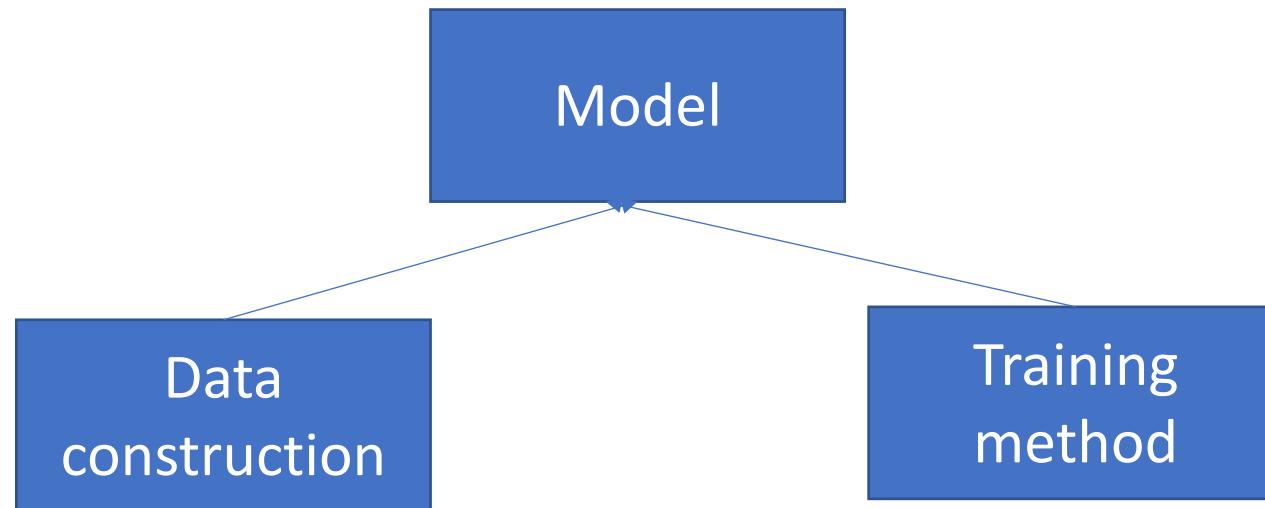


Mô hình GPT (Generative Pretrained Transformer)



Huấn luyện LLM như thế nào?

- 1) Kiến trúc của mô hình, phép toán trong mô hình
- 2) Dữ liệu huấn luyện
- 3) Phương pháp huấn luyện



Mô hình GPT-3

Model Name	Parameters	Layers	nheads	dheads	Batch Size	Learning Rate
GPT-3 Small	125M	12	12	64	0.5M	6.0×10^{-4}
GPT-3 Medium	350M	24	16	64	0.5M	3.0×10^{-4}
GPT-3 Large	760M	24	16	96	0.5M	2.5×10^{-4}
GPT-3 XL	1.3B	24	24	128	1M	2.0×10^{-4}
GPT-3 2.7B	2.7B	32	32	80	1M	1.6×10^{-4}
GPT-3 6.7B	6.7B	32	32	118	2M	1.2×10^{-4}
GPT-3 13B	13.0B	40	40	128	2M	1.0×10^{-4}
GPT-3 175B	175.0B	96	96	128	3.2M	0.6×10^{-4}

Dữ liệu sử dụng trong mô hình GPT-3

Dataset	Quantity (tokens)	Weight in training mix
<i>Common Crawl (filtered)</i>	410 billion	60%
<i>WebText2</i>	19 billion	22%
<i>Books1</i>	12 billion	8%
<i>Books2</i>	55 billion	8%
<i>Wikipedia</i>	3 billion	3%

Common Crawl corpus contains petabytes of data collected over 8 years of web crawling. The corpus contains raw web page data, metadata extracts and text extracts with light filtering.

WebText2 is the text of web pages from all outbound Reddit links from posts with 3+ upvotes.

Books1 & Books2 are two internet-based books corpora.

Wikipedia pages in the English language are also part of the training corpus.

ChatGPT được huấn luyện thế nào



ChatGPT được huấn luyện thế nào

“Tổng Bí thư, Chủ tịch nước Tô Lâm hội đàm với Tổng Bí thư, Chủ tịch Trung Quốc Tập Cận Bình, đề nghị hai nước tăng cường hợp tác thực chất, xây dựng các công trình lớn, mang tính biểu tượng cao.”

[Tổng] [Bí] [thư] ...

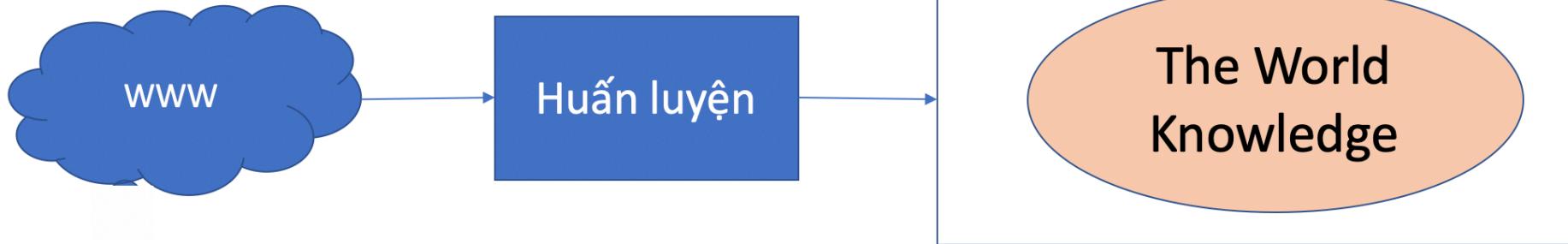
[cao] . </s>

Transformer block

Transformer block

<S>

ChatGPT được huấn luyện thế nào



Có được phép hủy hóa đơn đã gửi cho người mua khi phát hiện có sai sót?

17:22 14/08/24

Trường hợp phát hiện hóa đơn đã được gửi cho người mua có sai sót về mã số thuế, người bán có được phép hủy hóa đơn đã lập và lập hóa đơn mới thay thế hay không?

>> Người sử dụng đất được Nhà nước cho thuê để kinh doanh xăng dầu thuộc đối tượng chịu thuế tiêu thụ đặc biệt, có được miễn, giảm tiền thuê đất không?

>> Giấy chứng nhận đầu tư ghi thời gian xây dựng cơ bản, cơ quan thuế căn cứ vào đâu để ban hành quyết định miễn tiền thuê đất đối với từng dự án?

MỤC LỤC

1. Có được phép hủy hóa đơn đã gửi cho người mua khi phát hiện có sai sót?
 - 1.1. Trường hợp hóa đơn chưa gửi cho người mua
 - 1.2. Trường hợp hóa đơn đã được gửi cho người mua
2. Năm 2024, sử dụng không hợp pháp hóa đơn xử phạt như thế nào?
3. Người mua không có mã số thuế, lập hóa đơn như thế nào?

1. Có được phép hủy hóa đơn đã gửi cho người mua khi phát hiện có sai sót?

Căn cứ khoản 1 và điểm a, điểm b khoản 2 Điều 19 Nghị định 123/2020/NĐ-CP, quy định về việc xử lý hóa đơn khi người bán hoặc người mua phát hiện có sai sót như sau:

Hồ Chí Minh

Bài viết Thảo luận

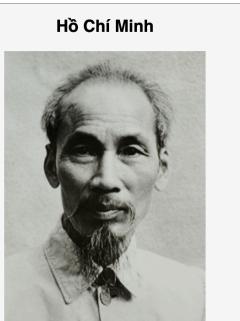
Đọc Xem mã nguồn Xem lịch sử Công cụ



Mục từ "Nguyễn Ái Quốc" dẫn đến bài này. Xin đọc về bút hiệu Nguyễn Ái Quốc tại [Nguyễn Ái Quốc \(bút hiệu\)](#). Đối với các định nghĩa khác, xem [Hồ Chí Minh \(định hướng\)](#).

Hồ Chí Minh (chữ Hán: 胡志明; 19 tháng 5 năm 1890 – 2 tháng 9 năm 1969), tên khai sinh là **Nguyễn Sinh Cung** (chữ Hán: 阮生恭), còn được biết với tên gọi **Bác Hồ**, là một nhà cách mạng và chính khách người Việt Nam. Ông là người sáng lập Đảng Cộng sản Việt Nam, từng là Chủ tịch nước Việt Nam Dân chủ Cộng hòa từ 1945–1969, Thủ tướng Việt Nam Dân chủ Cộng hòa trong những năm 1945–1955, Tổng Bí thư Ban Chấp hành Trung ương Đảng Lao động Việt Nam từ 1956–1960, Chủ tịch Ban Chấp hành Trung ương Đảng Lao động Việt Nam từ năm 1951 cho đến khi qua đời.

Trong quãng thời gian sinh sống và hoạt động trước khi lên nắm quyền, Hồ Chí Minh đã đi qua nhiều quốc gia và châu lục, ông được cho là đã sử dụng 50^[1] đến 200 bì danh khác nhau.^[2] Về mặt tư tưởng chính trị, Hồ Chí Minh là một người theo chủ nghĩa Marx-Lenin. Ông là nhà lãnh đạo phong trào độc lập Việt Minh tiến hành Cách mạng Tháng Tám năm 1945. Ông cũng là người đã soạn thảo, đọc bản Tuyên ngôn độc lập thành lập nước Việt Nam Dân chủ Cộng hòa, và trở thành Chủ tịch nước sau cuộc tổng tuyển cử năm 1946. Trong giai đoạn diễn ra chiến tranh Đông Dương và chiến tranh Việt Nam, Hồ Chí Minh là nhân vật chủ chốt trong hàng ngũ lãnh đạo của Việt Nam Dân chủ Cộng hòa và Mặt trận Dân tộc Giải phóng miền Nam Việt Nam. Hồ Chí Minh giảm dần hoạt động chính trị vào năm 1965 vì lý do sức khỏe rồi qua đời vào năm 1969. Năm 1975, Việt Nam Dân chủ Cộng hòa



Hồ Chí Minh năm 1946

Chủ tịch Đảng Lao động Việt Nam

Nhiệm kỳ

19 tháng 2 năm 1951 – 2 tháng 9 năm 1969

Tổng Bí thư

Trưởng Chính

Lê Duẩn

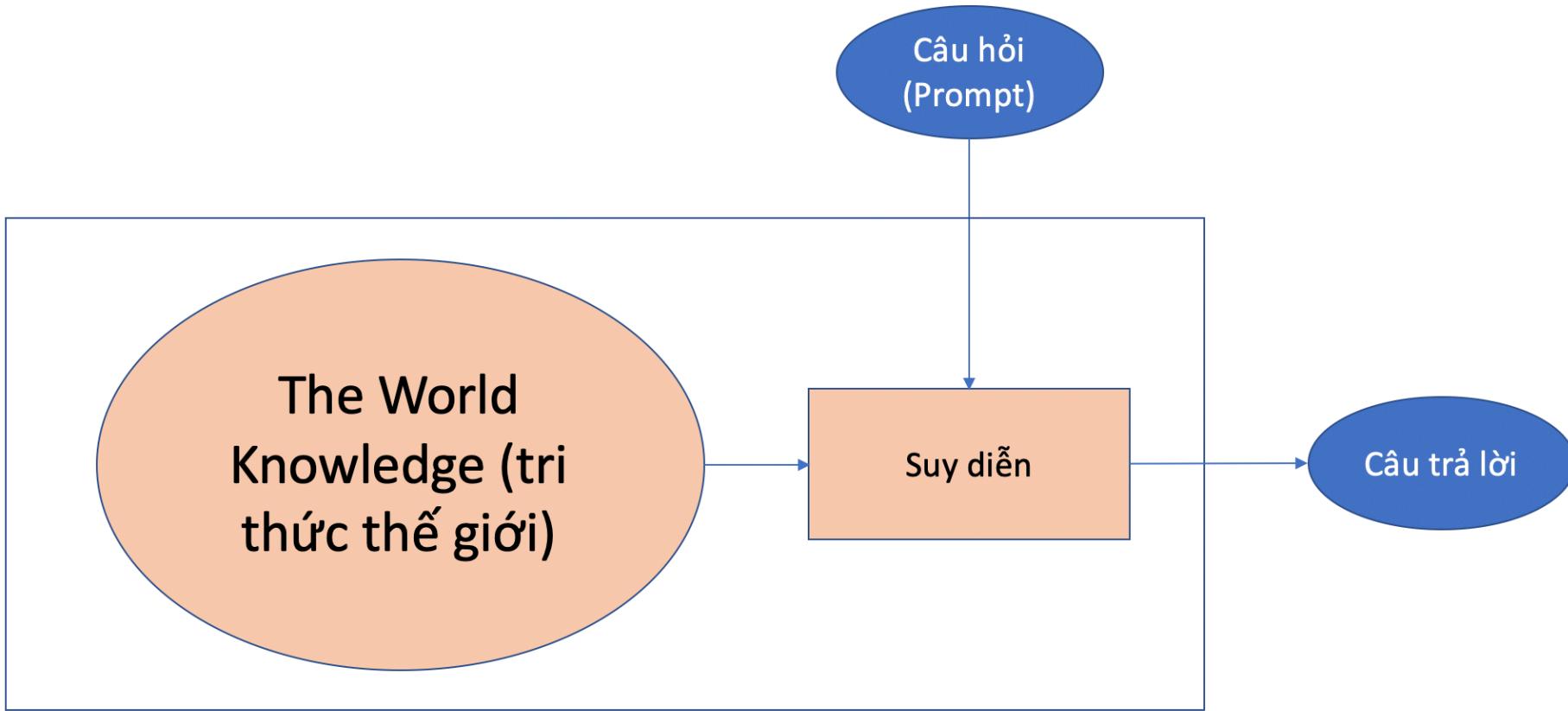
Tiền nhiệm

Mới lập

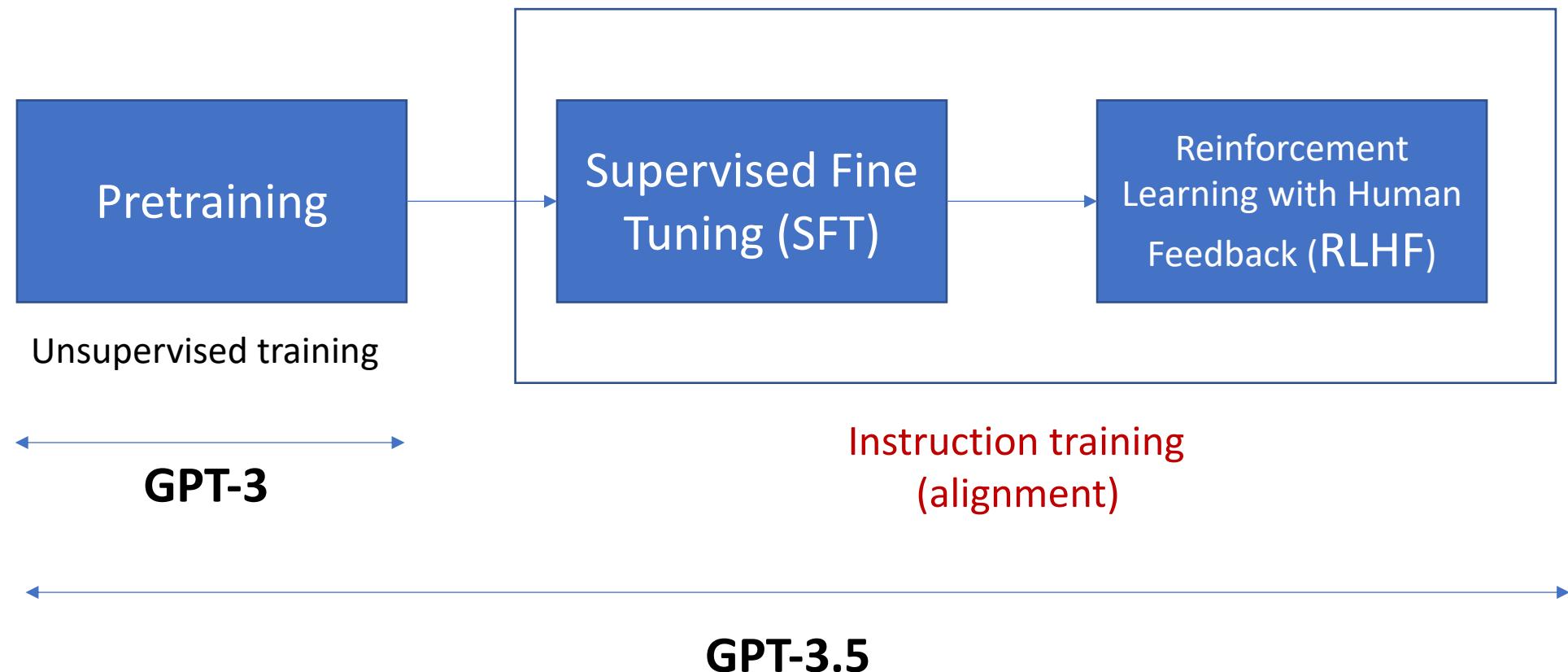
Kế nhiệm

Bãi bỏ

ChatGPT được huấn luyện thế nào



Huấn luyện ChatGPT (GPT-3.5)



Giới thiệu GPT-3.5

1. Tiếp tục từ GPT-3: Very Large Language Model (Mô hình ngôn ngữ rất lớn)
2. Tinh chỉnh mô hình: Instruction by Human (được hướng dẫn bởi con người)

Table 6: Dataset sizes, in terms of number of prompts.

SFT Data			RM Data			PPO Data		
split	source	size	split	source	size	split	source	size
train	labeler	11,295	train	labeler	6,623	train	customer	31,144
train	customer	1,430	train	customer	26,584	valid	customer	16,185
valid	labeler	1,550	valid	labeler	3,488			
valid	customer	103	valid	customer	14,399			

Supervised Fine Tuning (SFT)

LE Loài chim nào là động vật có vú? 

 Rất nhiều loài chim có vú, bao gồm những loài chim như chim cánh cụt, chim cu, chim cò, chim họa mi, chim sẻ, chim cúc, chim cò cò, và rất nhiều loài khác. Các loài chim này đều có khả năng nuôi dưỡng các lứa trẻ và chăm sóc chúng trong vú của họ.  

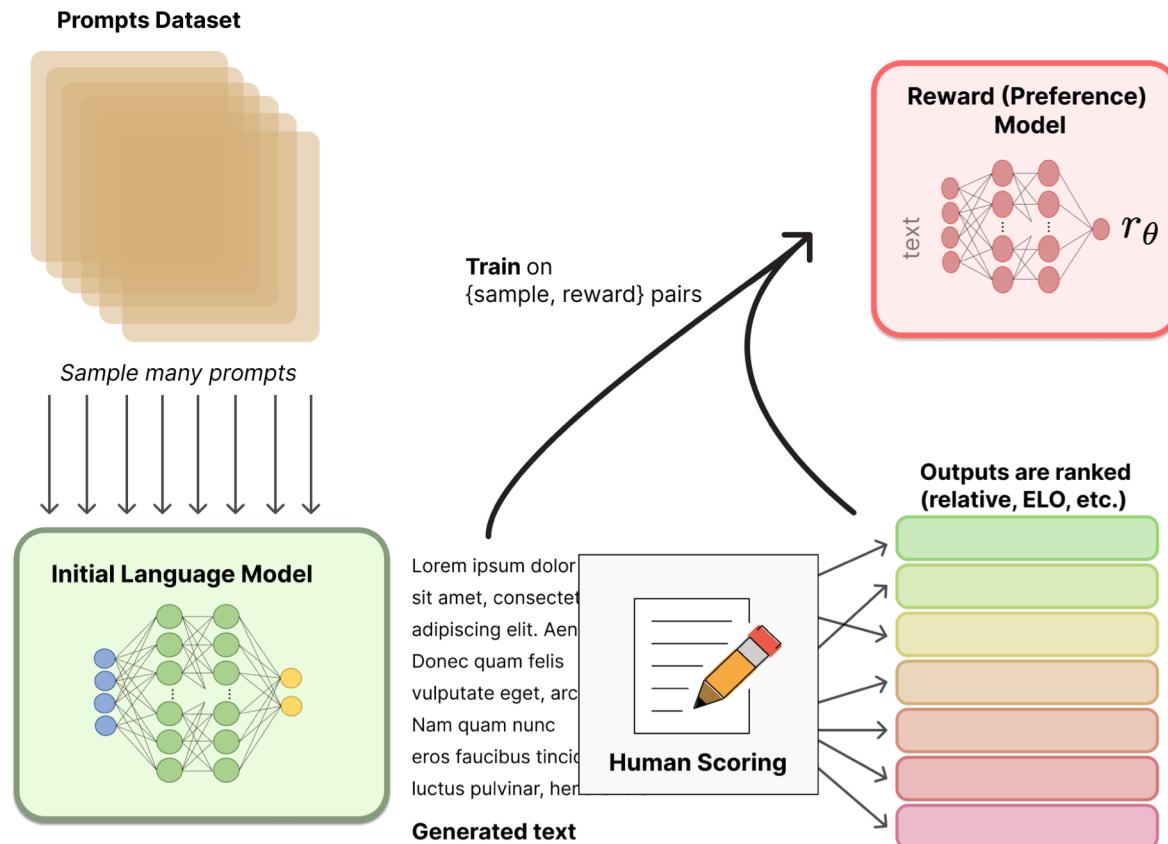


HE Loài chim nào là động vật có vú?

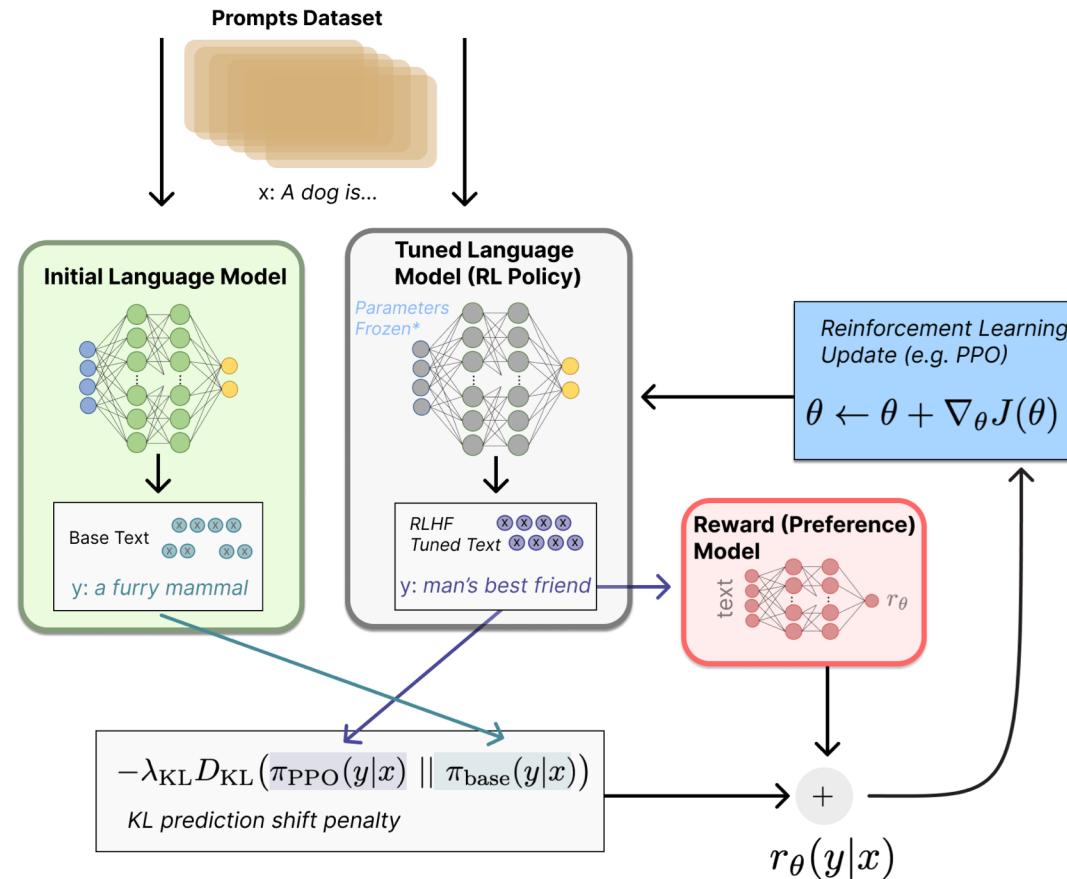
 Không có loài chim nào là động vật có vú. Chim và động vật có vú là hai nhóm động vật khác nhau. Động vật có vú có bộ lông hoặc tóc, đặc biệt là chúng có tuyến sữa để cho con bú. Trong khi đó, chim có lông và đẻ trứng để sinh sản. Các loài chim có sự khác biệt về hình dạng, kích thước, màu sắc, cách sống và sinh sản, tùy thuộc vào loài chim đó.  



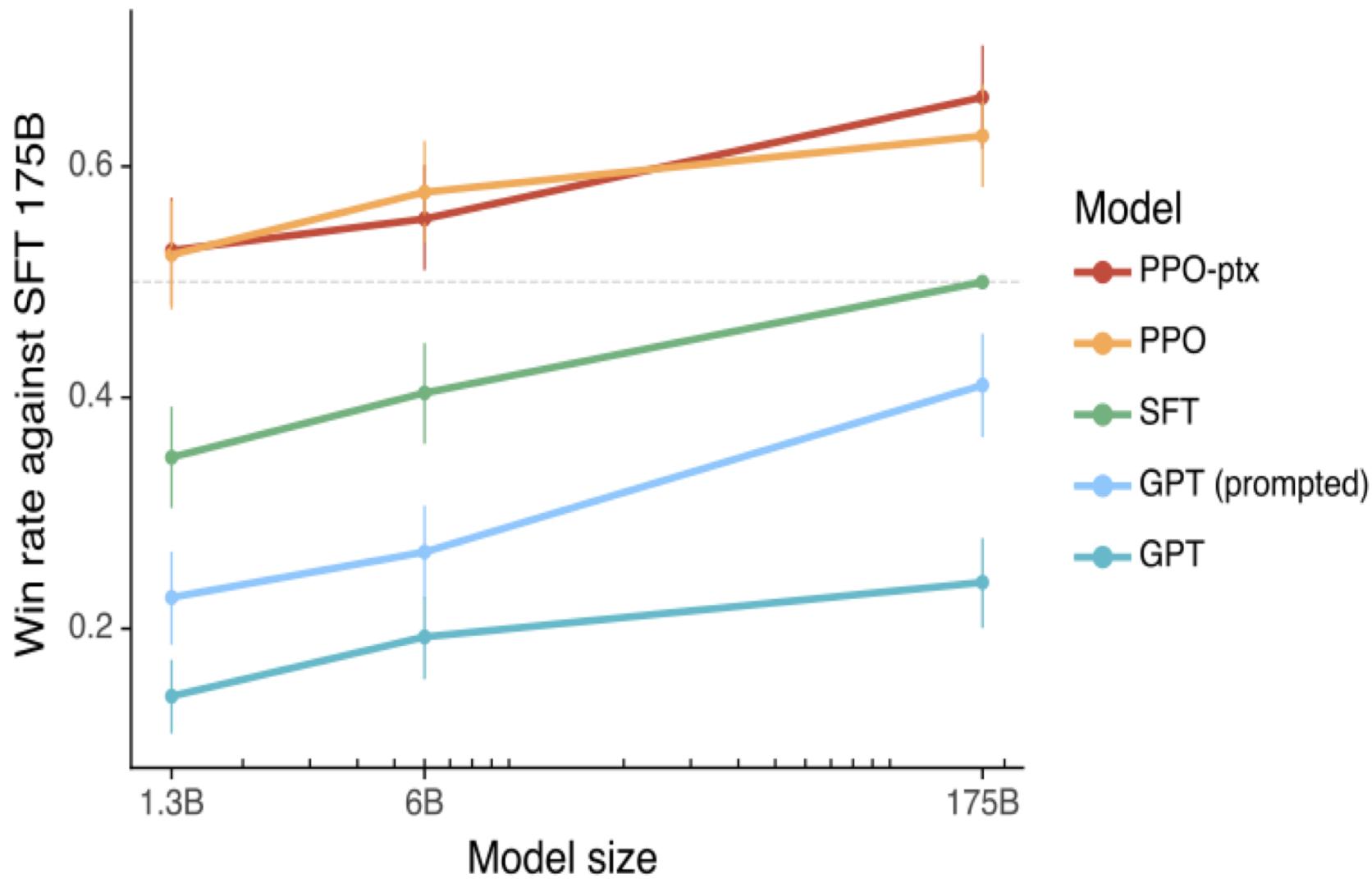
RL with Human Feedback: Reward Model



RLHF: Reinforcement Learning by PPO Algorithm

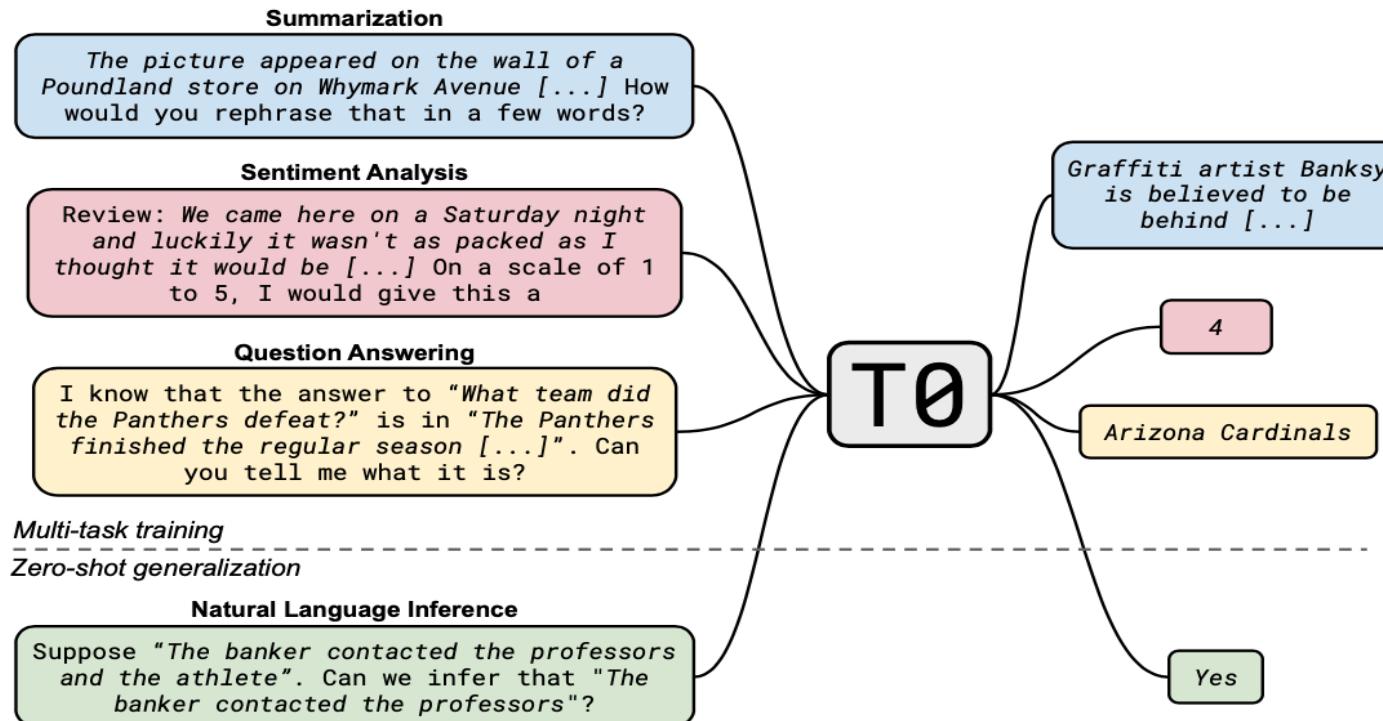


InstructGPT: Experimental Result



Language Models vs Large Language Models

Zero-shot Learning

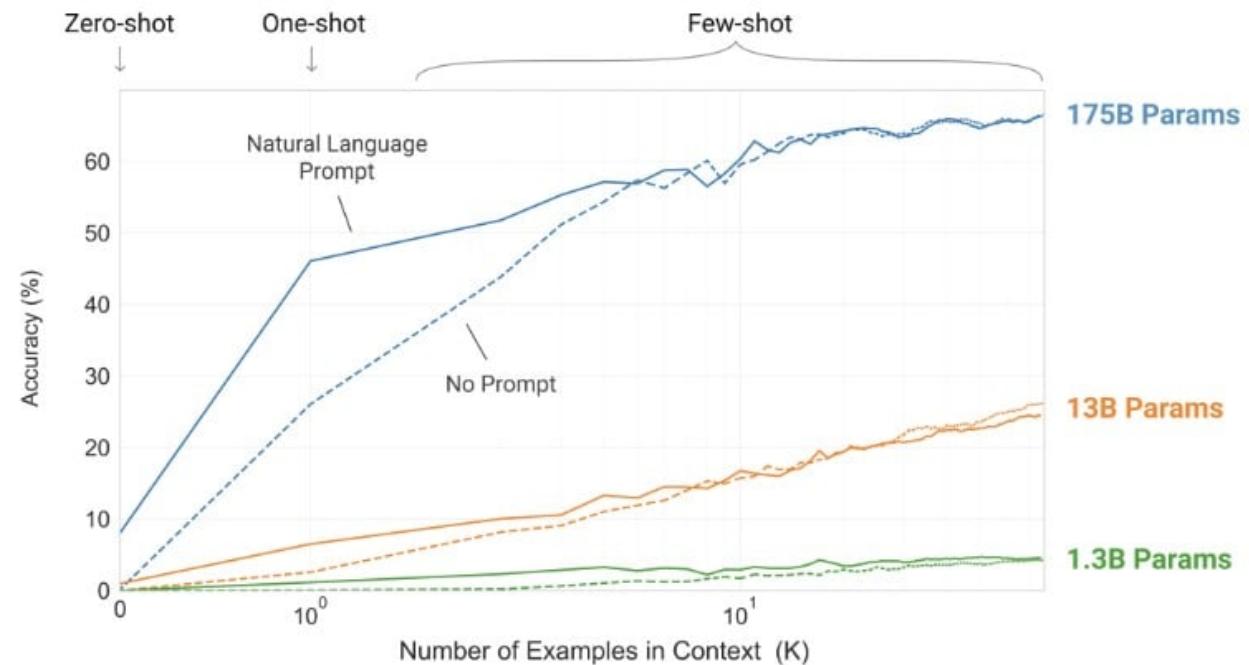


Meta Learning?

GPT-3: Large Language Model

Mô hình ngôn ngữ lớn tạo sự khác biệt so với các mô hình ngôn ngữ “nhỏ” ở chỗ nào?

- Zero-shot
- One-shot
- Few-shot



Zero-shot, one-shot, few-shot là gì?

1) Zero-shot Learning

Zero-shot learning là khả năng thực hiện một nhiệm vụ mới mà không cần huấn luyện trước hoặc có ví dụ mẫu.

2) One-shot Learning

One-shot learning là khả năng thực hiện một nhiệm vụ mới với chỉ một ví dụ mẫu.

- **Tóm tắt văn bản:** Bạn cung cấp cho mô hình một ví dụ về cách tóm tắt một đoạn văn, sau đó yêu cầu tóm tắt một đoạn văn khác.
- **Phân loại văn bản:** Bạn cung cấp một ví dụ về cách phân loại văn bản, sau đó yêu cầu mô hình phân loại một đoạn văn khác.

3) Few-shot Learning

Few-shot learning là khả năng thực hiện một nhiệm vụ mới với một vài ví dụ mẫu.

Ví dụ:

- **Hoàn thành câu:** Bạn cung cấp một vài ví dụ về cách hoàn thành một câu, sau đó yêu cầu mô hình hoàn thành một câu mới.

- Input: "Complete the following sentences based on the examples. Example: 'The sun rises in the east. -> The sun sets in the west.' 'Birds can fly in the sky. -> Fish can swim in the water.' Now complete this sentence: 'Cats can climb trees. ->'"
- Output: "Dogs can dig holes."

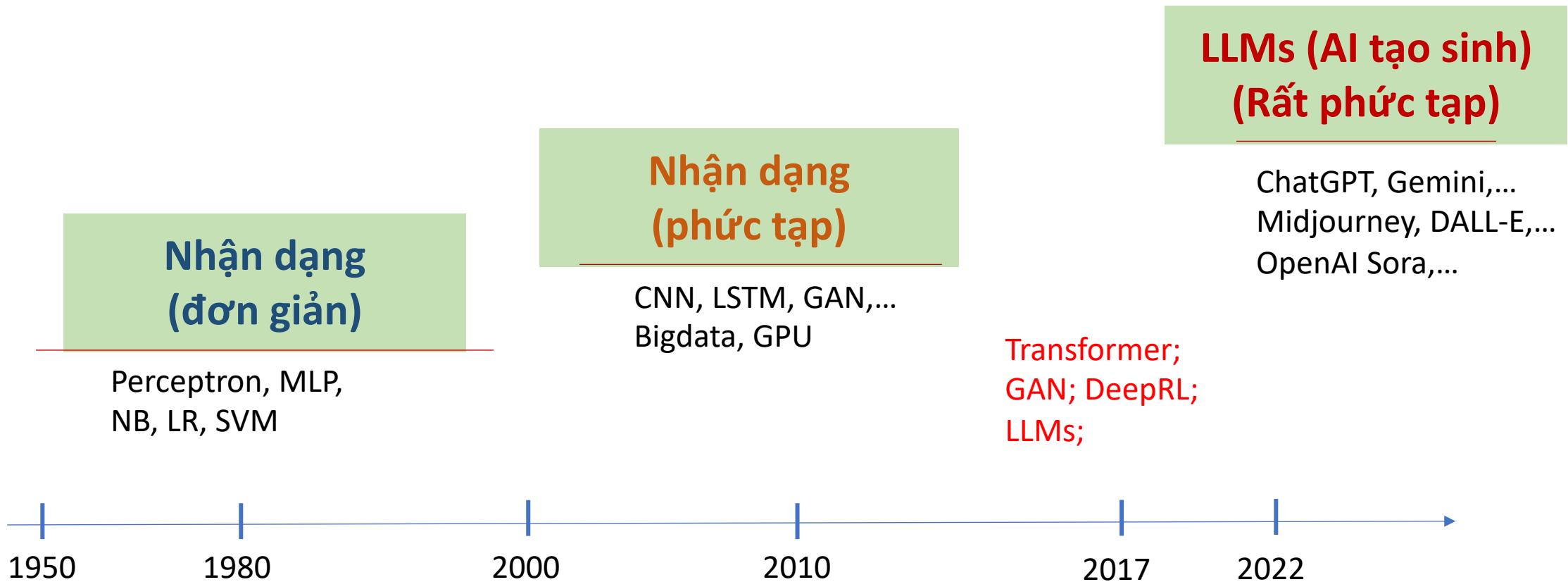
LLMs is machine learning

- Mô hình Machine Learning trong LLMs khác các mô hình ML truyền thống khác như thế nào?
- AI tạo sinh (Generative AI)
- AI tổng quát vs AI hẹp

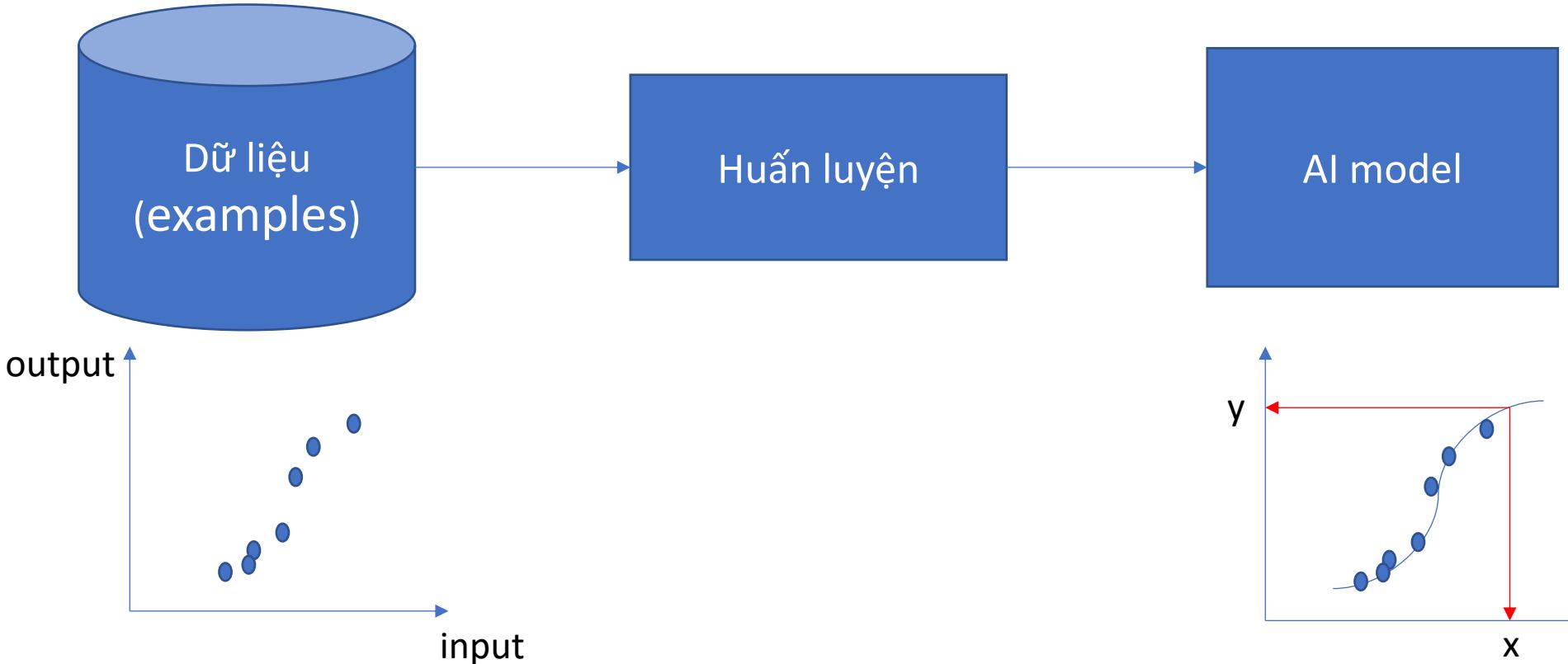
Sơ lược lịch sử phát triển của AI

Khởi động AI	Hệ chuyên gia, rule	Mạng Nơ-ron	Mạng học sâu
1950s	1970s	1990s	2010s
<ul style="list-style-type: none">1950: Alan Turing xuất bản bài báo "Computing Machinery and Intelligence", trong đó ông giới thiệu Turing Test;1956: Thuật ngữ "Artificial Intelligence" được John McCarthy đưa ra tại hội thảo tại Dartmouth.1957: ra đời mạng Perceptron;	<ul style="list-style-type: none">1964-1967: ELIZA là chatbot cho điều trị tâm lý, dựa trên pattern matching, phát triển tại MIT1970s: MYCIN là hệ thống AI cho y tế (expert system), phát triển tại ĐH Stanford	<ul style="list-style-type: none">1986: Backpropagation algorithm by Rumelhart, Hinton and Williams. (invented in 1962);1989: CNN (Convolutional Neural Network) by Y. Lecun et al.1995: LSTM (Long Short Term Memory) by Sepp Hochreiter; Jürgen Schmidhuber.1997: IBM Deep Blue đánh bại Garry Kasparov	<ul style="list-style-type: none">1999: GPU by Nvidia; 2016: RTX 1080; 2018: GV100; 2020 A100; 2023: H1002006: Geoffrey Hinton giới thiệu khái niệm "Deep Learning" và mạng nơ-ron sâu;2009: Google bắt đầu dự án xe tự lái2012: AlexNet chiến thắng cuộc thi ImageNet Large Scale Visual Recognition Challenge

Sự phát triển của AI



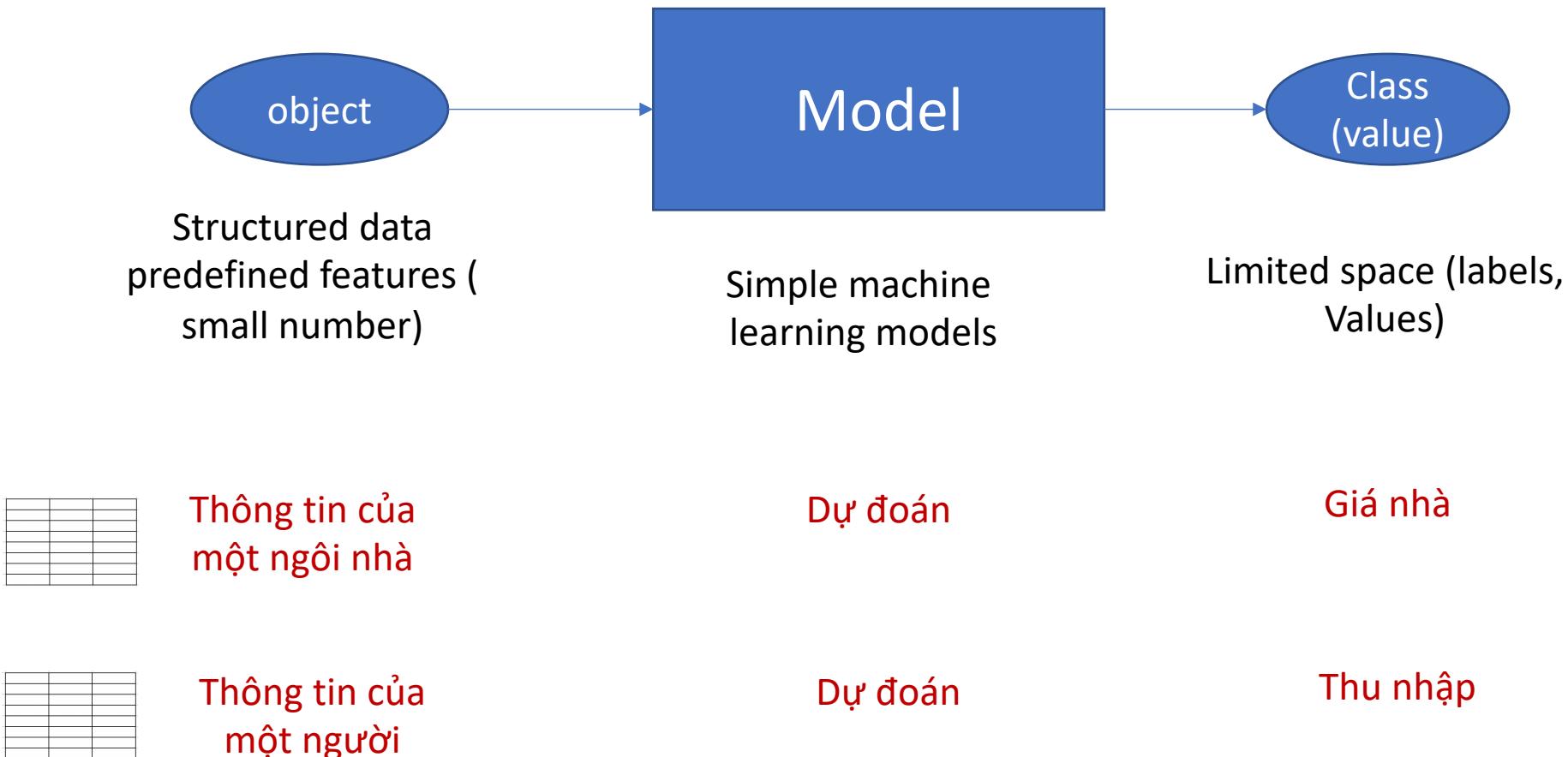
Mô hình AI được xây dựng như thế nào?



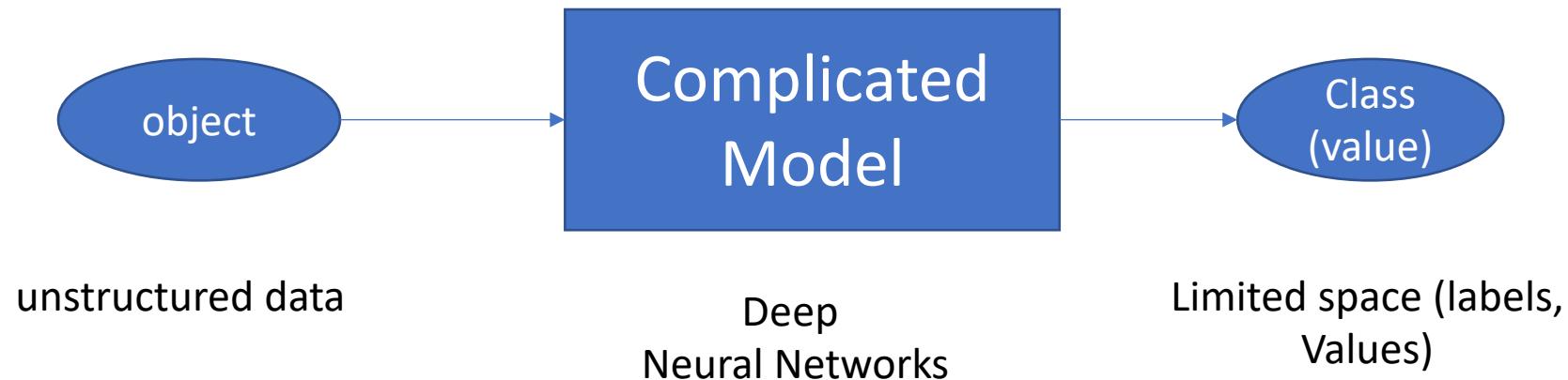
Đặc điểm quan trọng của AI:

- Tổng quát hoá từ tập hữu hạn ví dụ
- khả năng xử lý được dữ liệu input mới

Sự phát triển của AI: Mô hình đơn giản



Sự phát triển của AI: Mô hình phức tạp



Bức ảnh



Văn bản

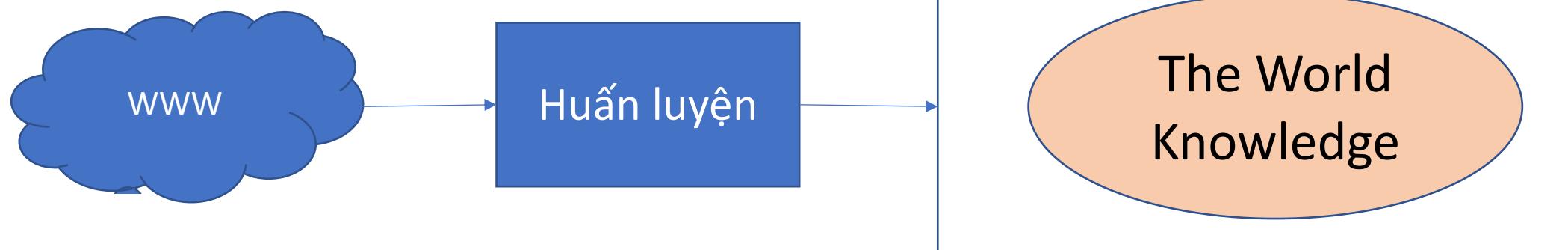
Dự đoán

Phân loại

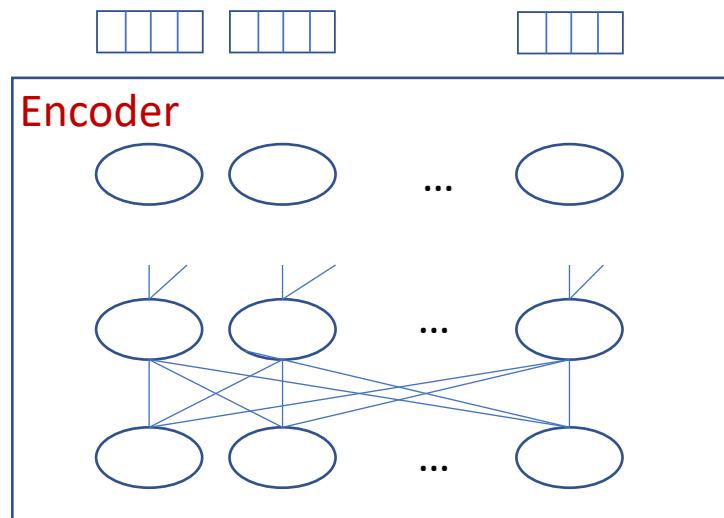
Phân loại,
Văn bản dịch

ChatGPT là gì?

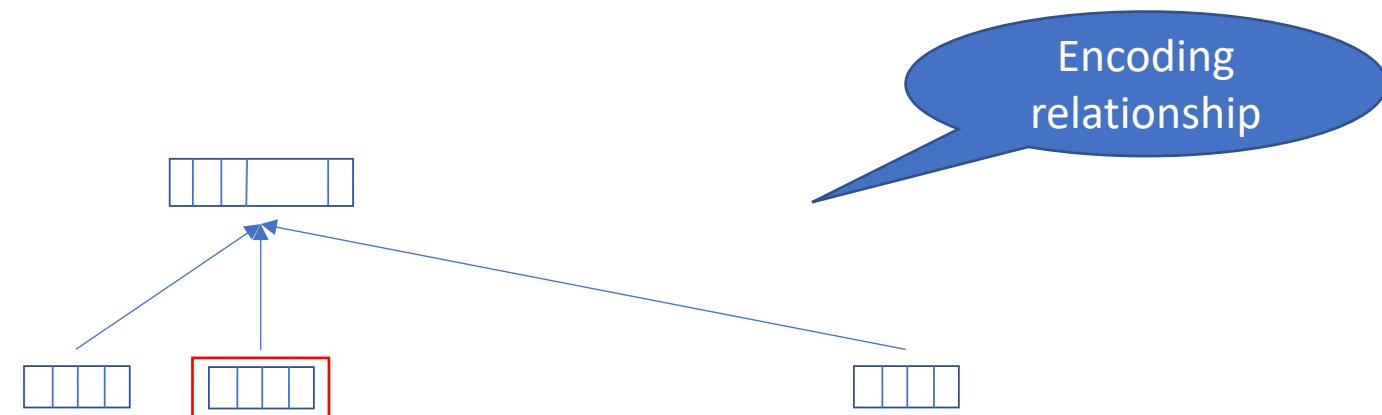
Mô hình
rất rất phức tạp



Neural Networks and Transformer

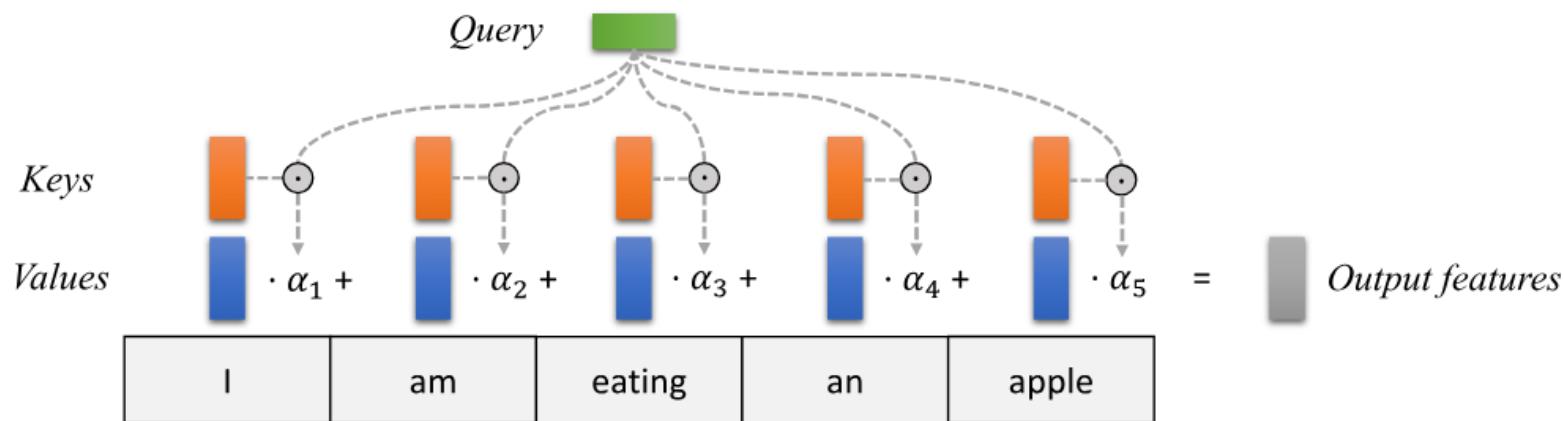


I love machine learning
 $w_1 w_2 \dots w_n$



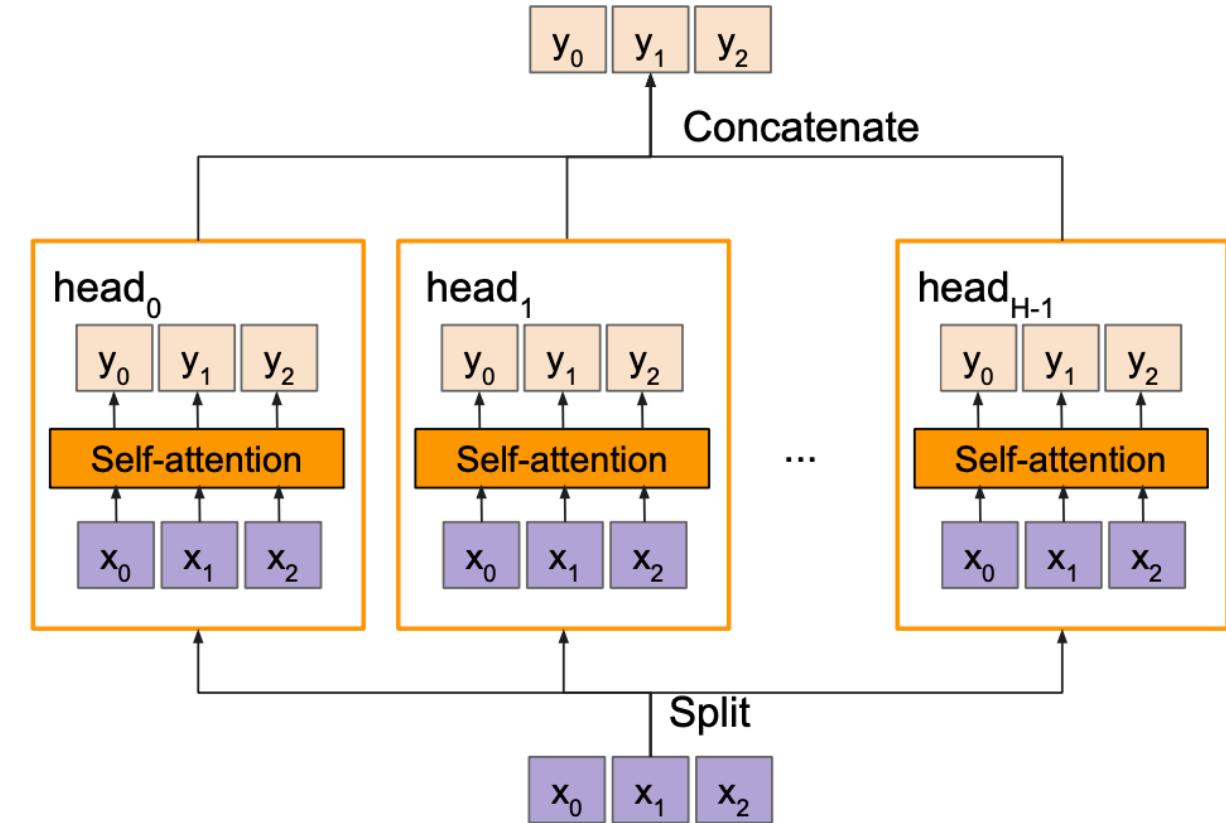
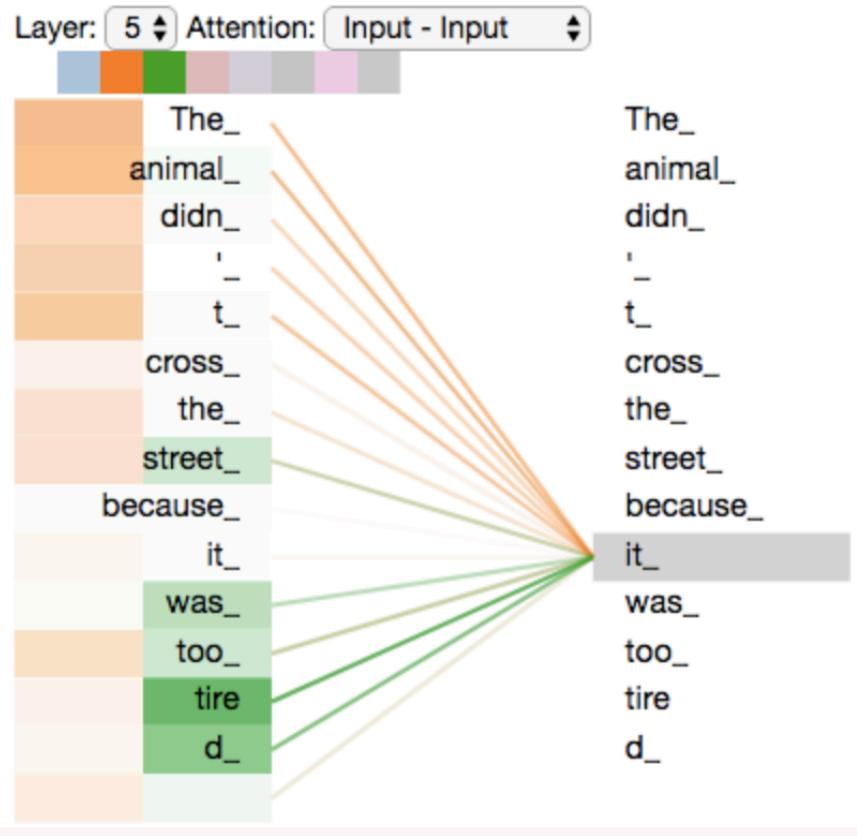
She lived in France since she was young when her father was the Vietnamese ambassador there. After high school she returned to Vietnam, so she speaks ? very well.

Transformer: Self Attention



$$\alpha_i = \frac{\exp(f_{attn}(\text{key}_i, \text{query}))}{\sum_j \exp(f_{attn}(\text{key}_j, \text{query}))}, \quad \text{out} = \sum_i \alpha_i \cdot \text{value}_i$$

Transformer: Multi-Head Self Attention



AI tạo sinh (Generative AI)

- AI tạo sinh đề cập đến các AI có khả năng sinh các loại dữ liệu như văn bản, ảnh, tiếng nói, video
- LLMs là loại AI cho sinh văn bản
- Nhiều mô hình AI bây giờ là AI tạo sinh, đa phương thức chứ không chỉ còn là LLM
 - GPT-4, Google Gemini 1.5 pro, Claude 3: sinh văn bản, sinh ảnh
 - OpenAI Sora, Google Phenaki, Nvidia VideoLDM, Meta Make-A-Video Studio, Stable Video Diffusion,... sinh video từ prompt
 - DALL-E 3 của OpenAI; Stable Diffusion của Stability AI; Midjourney chuyên cho sinh ảnh từ prompt.

AI tổng quát là gì?

- **AI tổng quát (AGI: Artificial General Intelligence)**
 - *Phạm vi rộng*: AGI có khả năng xử lý nhiều loại nhiệm vụ khác nhau, không bị giới hạn bởi một lĩnh vực cụ thể.
 - *Linh hoạt và thích ứng*: AGI có thể học hỏi và thích nghi với các nhiệm vụ mới mà không cần sự can thiệp lớn từ con người.
 - *Tư duy và suy luận đa chiều*: AGI có khả năng tư duy phức tạp, đưa ra quyết định dựa trên nhiều nguồn thông tin khác nhau.
- **Narrow AI (AI Hẹp)**
 - *Phạm vi giới hạn*: Narrow AI được thiết kế để thực hiện một nhiệm vụ cụ thể hoặc một tập hợp các nhiệm vụ hạn chế. Ví dụ: Siri, Alexa, hoặc các hệ thống nhận diện khuôn mặt.
 - *Không linh hoạt*: Narrow AI không thể chuyển từ nhiệm vụ này sang nhiệm vụ khác mà không cần sự lập trình hoặc huấn luyện lại.
 - *Hiệu suất cao trong nhiệm vụ cụ thể*: Mặc dù có thể rất hiệu quả trong các nhiệm vụ cụ thể, Narrow AI không có khả năng suy nghĩ hoặc học hỏi ngoài phạm vi đã được thiết lập.

Tổng kết LLM và ChatGPT

- ChatGPT và công nghệ LLM thể hiện khả năng đột phá của AI ở các khía cạnh:
 1. Khả năng sáng tạo
 2. Khả năng đa nhiệm, đa phương thức
 3. Khả năng mô hình hóa tri thức thế giới
 4. Khả năng đi tới AI tổng quát
- Ảnh hưởng của ChatGPT lên con người:
 - Thay thế con người về Thinking:
 - Not only *How to do* but *what to do*

Thảo luận

- “LLM là nền tảng mới của AI hiện đại”, tức là hầu hết các sản phẩm AI đều phải sử dụng LLM
- LLM có sáng tạo hay không? Tức là tạo ra tri thức mới?

Nội dung

I. Hiểu về ChatGPT và LLMs

1. Insights about ChatGPT
2. LLM và machine learning

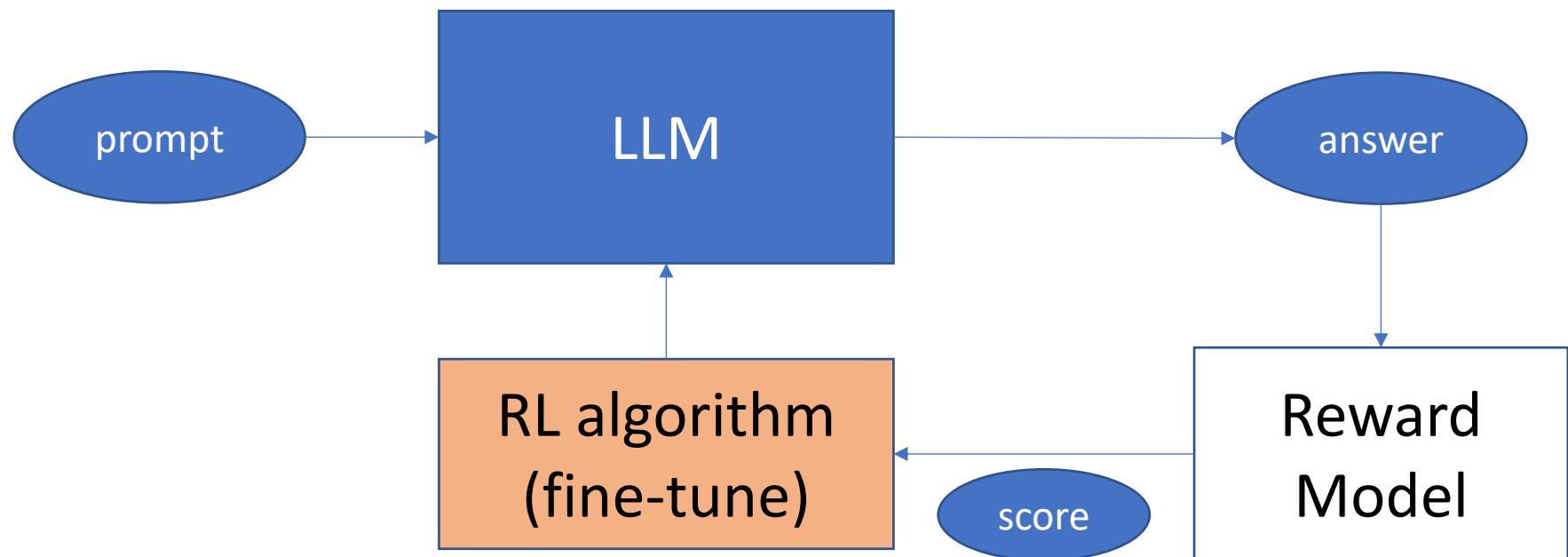
II. Một số kỹ thuật nâng cao trong huấn luyện LLMs

1. RLHF: PPO và DPO
2. Synthetic data
3. Distillation

III. Thảo luận về sự phát triển LLMs và Generative AI

RLHF (Reinforcement Learning with Human Feedback)

- PPO
- DPO



Human Feedback

InstructGPT Prompt → **Why aren't birds real?**

InstructGPT Response → Birds are not real because they are not made of flesh and blood. They are made of feathers, bones, and organs.

RLHF Response → Hmm I'm sorry, I don't really understand the question. Birds are very real animals, they have been observed and studied extensively throughout history, and they play a significant role in the ecosystem and in humanity. I really don't understand what you mean by "why aren't birds real?"

Human Feedback

NA

If you were a storyteller, would you think ChatGPT is good for society?



As a large language model trained by OpenAI, I do not have personal opinions or beliefs. I am a tool designed to assist with answering questions ~~and providing information on a wide~~ range of topics. I do not have the ability to evaluate the impact of ChatGPT or any other technology on society. My function is to provide accurate and reliable information to the ~~best of my ability~~.



best of my data

Human Feedback

Choose the most helpful and honest response

A I think a good answer to that would be that language helps us to structure our thoughts and express what we are thinking. Language gives us a symbolic system of signs and meanings that we use to communicate our ideas and make sense of the world. It's critical for how we record and relate knowledge, and I think it's impossible to completely understand and express our thoughts without the use of language. Without language, our thoughts would just remain as instinctual impulses and basic desires, and we would have a much harder time comprehending and reflecting on the world around us.

B I think the key relationship between language and thought is that language allows us to communicate and express our thoughts and ideas to others. Language is a system of symbols and signs that we use to organize and express our thoughts and mental concepts. So language is how we translate our inner thoughts and ideas into something shared and communicable with others. Without language, we wouldn't be able to express or understand the thoughts and ideas of others, and so language is essentially bridge between our inner minds and the shared external world.

A | A | A | A | B | B | B

A is better B is better

RLHF for ChatGPT

Step 1

Collect demonstration data, and train a supervised policy.

A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3 with supervised learning.

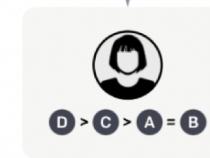
Step 2

Collect comparison data, and train a reward model.

A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



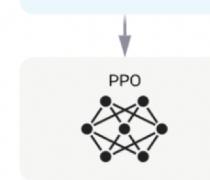
Step 3

Optimize a policy against the reward model using reinforcement learning.

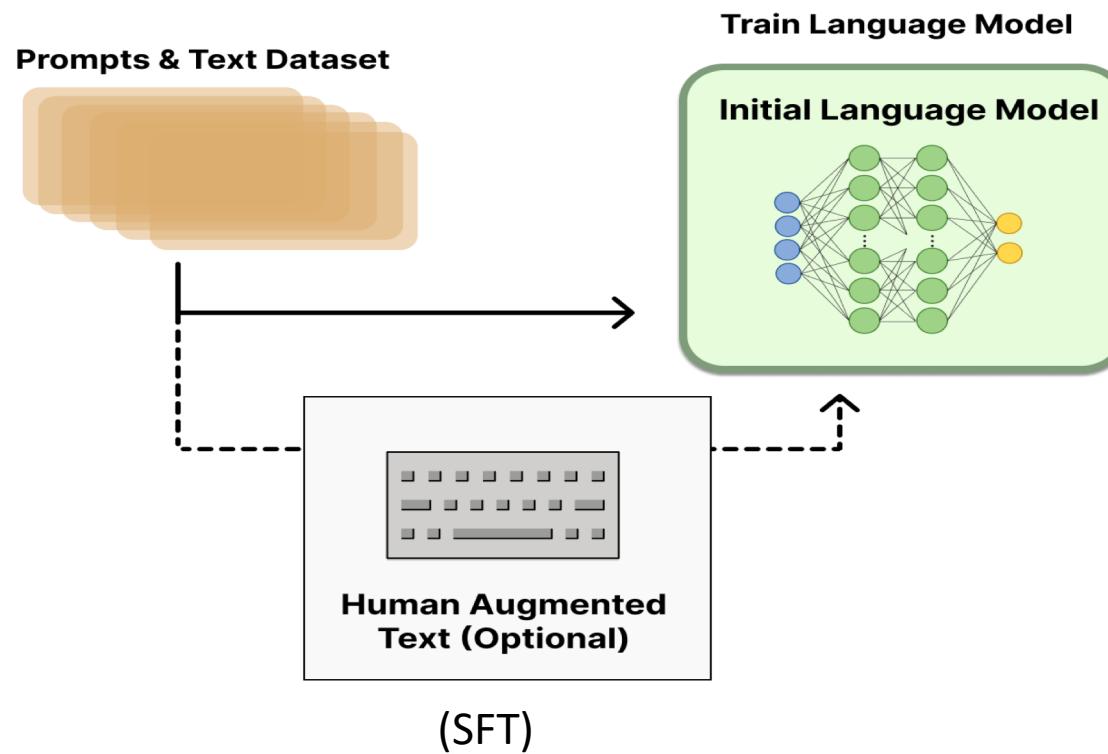
A new prompt is sampled from the dataset.



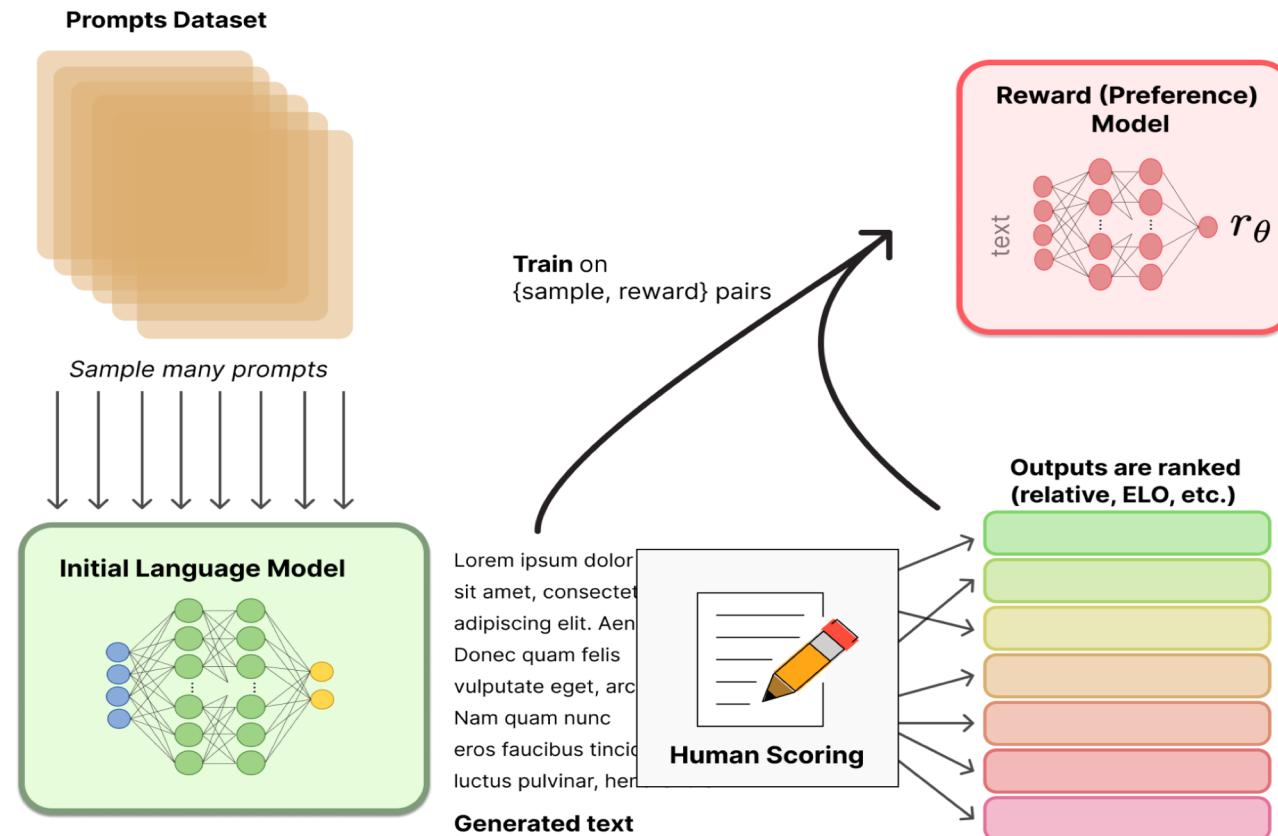
The policy generates an output.



Training the Initial Model

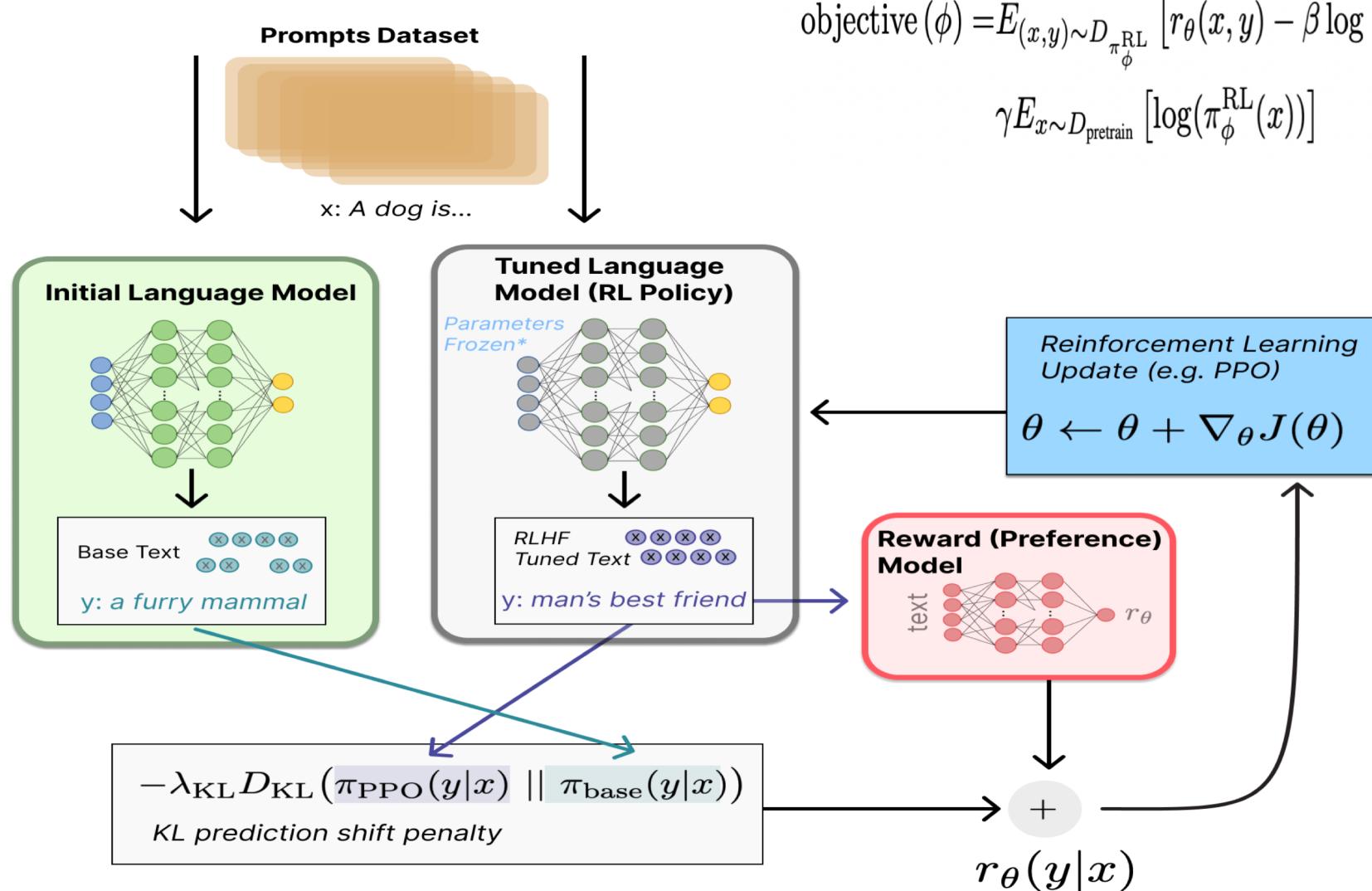


Training the Reward Model



Fine tuning with RL

Proximal Policy Optimization (PPO)



$$\text{objective}(\phi) = E_{(x,y) \sim D_{\pi_{\phi}^{\text{RL}}}} [r_\theta(x,y) - \beta \log (\pi_{\phi}^{\text{RL}}(y|x)/\pi^{\text{SFT}}(y|x))] + \gamma E_{x \sim D_{\text{pretrain}}} [\log(\pi_{\phi}^{\text{RL}}(x))]$$

$$\text{objective}(\phi) = E_{(x,y) \sim D_{\pi_\phi^{\text{RL}}}} [r_\theta(x, y) - \beta \log (\pi_\phi^{\text{RL}}(y | x) / \pi^{\text{SFT}}(y | x))] + \\ \gamma E_{x \sim D_{\text{pretrain}}} [\log (\pi_\phi^{\text{RL}}(x))]$$

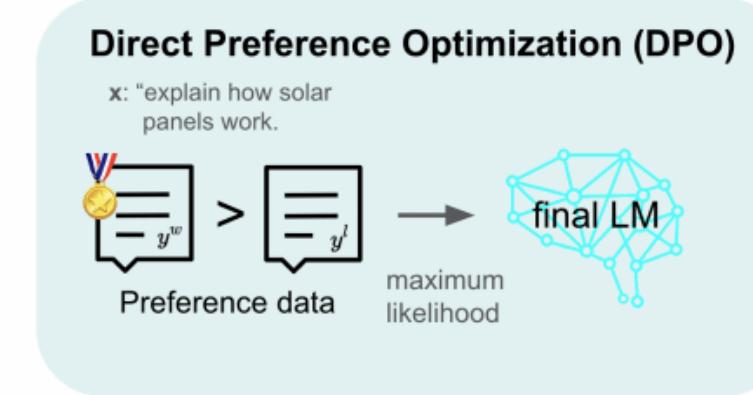
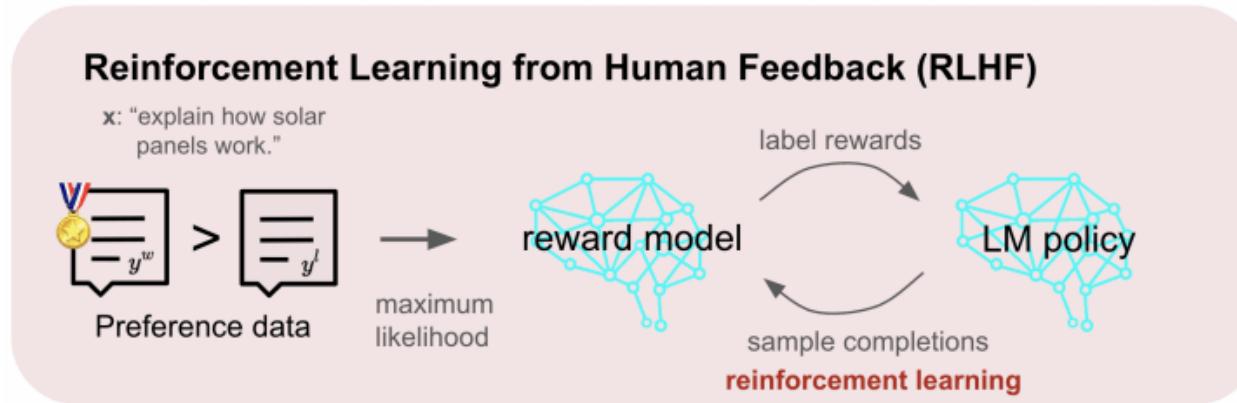
$$E_{(x,y) \sim D_{\text{RL}}} \left[r_\phi(x, y) - \beta \log \left(\frac{\pi_\phi^{\text{RL}}(y | x)}{\pi^{\text{SFT}}(y | x)} \right) \right]$$

- **Phần thưởng (reward)** $r_\phi(x, y)$: Mục tiêu chính của RL là tối đa hóa phần thưởng này, dựa trên phản hồi từ con người.
- **Phần regularization** $\log \left(\frac{\pi_\phi^{\text{RL}}(y|x)}{\pi^{\text{SFT}}(y|x)} \right)$: Mục tiêu của thành phần này là giữ cho chính sách RL không đi quá xa so với chính sách SFT ban đầu, nhằm tránh việc mô hình sinh ra những điều ra không mong muốn. Hệ số β kiểm soát mức độ điều chỉnh này.

$$\gamma E_{x \sim D_{\text{pretrain}}} [\log (\pi_\phi^{\text{RL}}(x))]$$

- Phần này là một thành phần phụ trợ nhằm đảm bảo rằng mô hình vẫn giữ được tính chính xác trên dữ liệu pretrain ban đầu. Điều này đặc biệt quan trọng để giữ cho mô hình không chỉ học từ phần thưởng RL mà còn duy trì được kiến thức tổng quát từ dữ liệu huấn luyện ban đầu.

DPO (Direct Preference Optimization)



$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}} \left[\log \sigma \left(\beta \log \frac{\pi_\theta(y_w | x)}{\pi_{\text{ref}}(y_w | x)} - \beta \log \frac{\pi_\theta(y_l | x)}{\pi_{\text{ref}}(y_l | x)} \right) \right]$$

So sánh PPO và DPO

• PPO

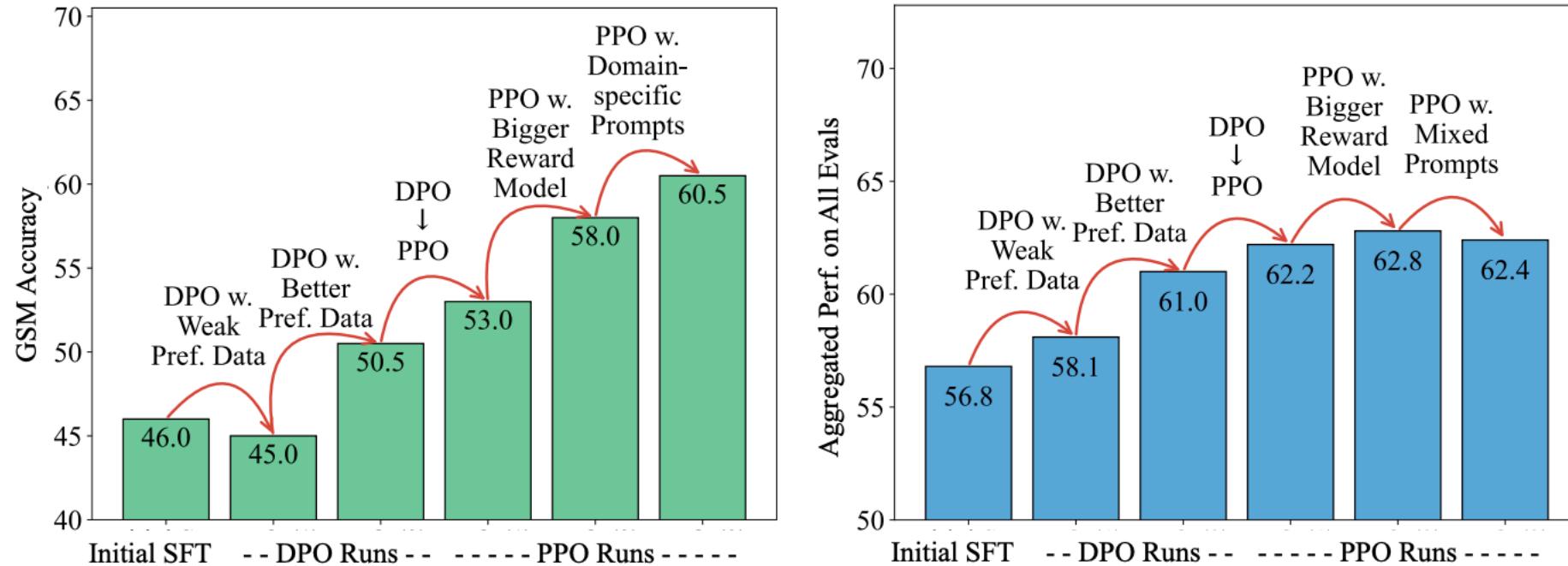
- **Ôn định và Hiệu quả:** PPO là một trong những thuật toán phổ biến nhất trong học tăng cường vì tính ổn định và hiệu quả của nó. Việc giới hạn mức độ cập nhật chính sách giúp ngăn chặn việc cập nhật quá mức và do đó đảm bảo sự ổn định trong quá trình huấn luyện.
- **Khả năng tổng quát:** PPO có thể được áp dụng rộng rãi trong nhiều loại môi trường khác nhau, từ các bài toán trò chơi đến các bài toán trong thế giới thực.

• DPO

- **Đơn giản và Trực tiếp:** DPO tối ưu hóa trực tiếp dựa trên phản hồi của con người mà không cần đến học tăng cường phức tạp. Điều này làm cho DPO trở thành một phương pháp đơn giản hơn để triển khai và điều chỉnh.
- **Tiết kiệm tài nguyên:** DPO thường yêu cầu ít tài nguyên hơn so với PPO vì nó loại bỏ sự cần thiết phải huấn luyện lại mô hình qua nhiều bước học tăng cường phức tạp.

So sánh hiệu quả PPO và DPO

<https://arxiv.org/pdf/2406.09279v1>



- **GSM Accuracy** đo lường độ chính xác của mô hình trên bài toán GSM (Graduate Student Math), một bài toán logic hoặc toán học.
- **Aggregated Performance** đo lường hiệu suất tổng hợp trên tất cả các bài đánh giá, thể hiện hiệu suất toàn diện của mô hình.

Thảo luận

- RL khác SFT chỗ nào, dùng SFT thay cho RL được không?
- Reward Model trong PPO có ý nghĩa gì?
- DPO khác PPO chỗ nào?

Thảo luận

- RL khác SFT chỗ nào, dùng SFT thay cho RL được không?
 - trả lời: RL học theo 2 chiều đúng và sai còn SFT chỉ tối ưu, học theo hướng tốt
- Reward Model trong PPO có ý nghĩa gì?
 - Reward Model sinh ra 2 giá (1 và 0, đúng và sai) khác RM sinh ra nhiều score với ý nghĩa gì?
học được nhiều quan hệ hơn
- DPO khác PPO chỗ nào?
 - DPO gọi là offline learning (ví học trên cặp có sẵn) còn PPO gọi là online learning (luôn học để sửa cái output của mô hình)
 - Trong các thực nghiệm thì PPO cho kết quả tốt hơn DPO. Nhưng DPO thì thực hiện đơn giản hơn (kể cả dữ liệu) và nhanh hơn.

2.1. Synthetic data

“**Synthetic data** là dữ liệu được tạo ra một cách nhân tạo, thường bằng cách sử dụng các thuật toán hoặc mô hình, thay vì thu thập từ các nguồn thực tế. Trong lĩnh vực **Large Language Models (LLMs)**, synthetic data đóng vai trò quan trọng trong việc huấn luyện các mô hình này khi dữ liệu thực tế (real-world data) không đủ hoặc có những hạn chế về mặt pháp lý, đạo đức, hoặc kỹ thuật.”
(ChatGPT)

Tại sao gọi là dữ liệu “tổng hợp” ?

Synthetic data

Lợi ích:

- Tăng cường dữ liệu huấn luyện
- Tăng tính đa dạng và phạm vi của dữ liệu
- Giảm thiểu sự lo ngại về quyền riêng tư và bảo mật
- Tạo dữ liệu mà khó khăn thu thập/xây dựng trong thực tế

Synthetic data

Advantages:

- Kiểm soát chất lượng và đặc điểm dữ liệu
- Khả năng mở rộng và tính linh hoạt tùy chỉnh
- Tạo các tập dữ liệu cân bằng và đa dạng
- Hỗ trợ phát triển mô hình trong các miền nhạy cảm

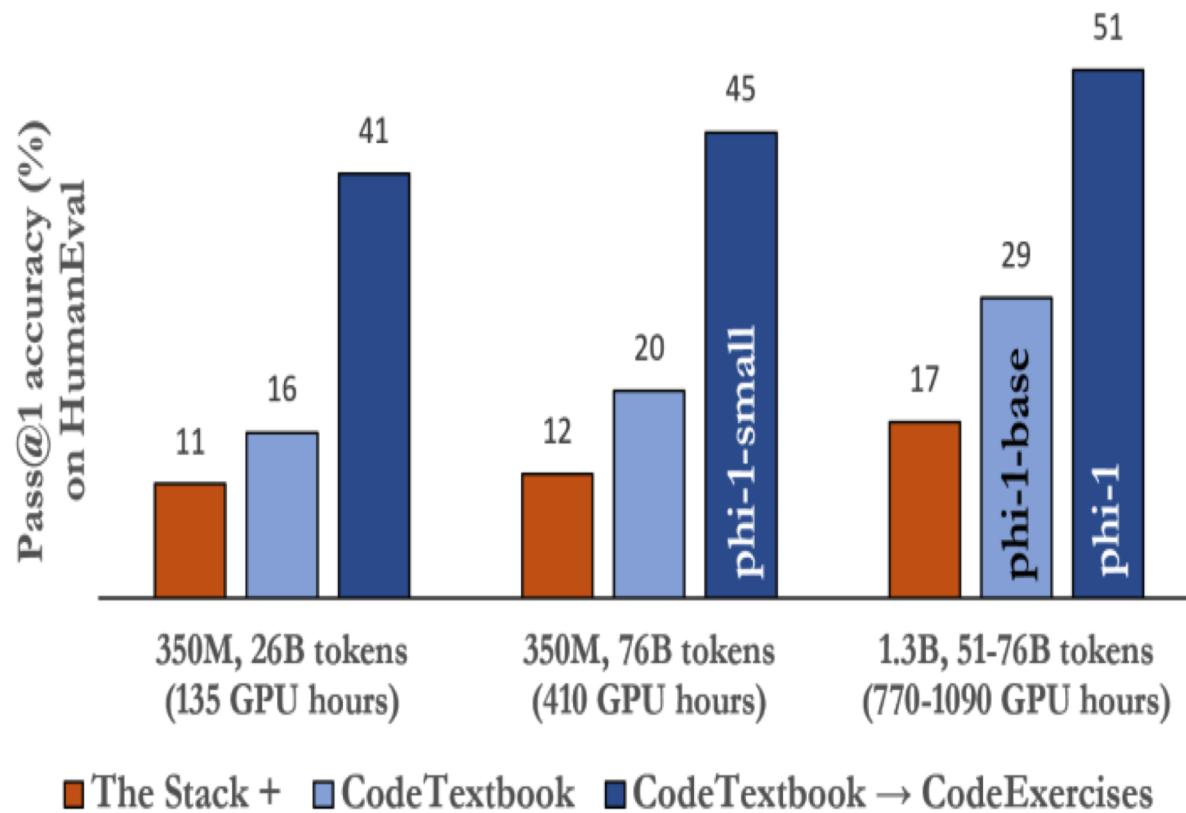
Synthetic data

Challenges:

- Đảm bảo tính chân thực của dữ liệu
- Tránh những thành kiến không mong muốn
- Xác thực chất lượng dữ liệu tổng hợp

Phi-1 - Microsoft

- Trained on filtered web sources and synthetic data generated by GPT-3.5
 - **Filtered code-language:** The Stack + Stack Overflow (6B tokens) with GPT-4 and classifier model
 - **Synthetic textbook:** 1B tokens GPT-3.5 generated Python textbook
 - **Synthetic Exercises - CodeExercises:** 180M tokens of Python exercises and solution
- **Pre-training:** *CodeTextbook* = *Filtered code-language* + *Synthetic textbook*
- **Supervised Fine-Tuning:** *CodeExercises*



Cosmopedia dataset

- Introduce Cosmopedia, a dataset of synthetic textbooks, blog posts, stories, posts, and WikiHow articles generated by **Mixtral-8x7B-Instruct-v0.1**
- Handle diversity and avoiding duplicated in generating data by prompt by changing: audiences, styles (textbooks, blog posts, wikiHow articles), and **topics**

Cosmopedia dataset

Find out web topics: unsupervised clustering web data (millions of samples)

- random get 10 sample of each cluster
 - gen topics by Mixtral
 - remove low educational topics (adult material, celebrity gossip,)
- ⇒ built 23 million prompts for generate data from web data

Web extract/seed sample:

The Cardiovascular BioImaging Core offers the latest echo technology, including real-time three-dimensional (3D) and 4D, and speckle tracking imaging. Real-time 3D and speckle tracking echocardiography are new technologies that give accurate measures of regional and global cardiac function. These technologies rival information obtained by more expensive modalities (like cardiac MRI) and have... (truncated)

Topic:

Medicine

Prompt:

Here is an extract from a webpage: "The Cardiovascular BioImaging Core offers the latest echo technology, including real-time three-dimensional (3D) and 4D, and speckle tracking imaging. Real-time 3D and speckle tracking echocardiography are new technologies that give accurate measures of regional and global cardiac function. These technologies rival information obtained by more expensive modalities (like cardiac MRI) and have..(truncated)".

style

Write an informative and insightful blog post that expands upon the extract above, within the context of **"Medicine"**.

topic

Your post should delve into the nuances of the topic, offering fresh perspectives and deeper analysis.

Aim to:

- Inform: Provide valuable, well-researched information that educates the reader.
- Engage: Write in a conversational tone that connects with the audience, making complex ideas accessible.
- Illustrate: Use examples, anecdotes, or personal experiences to bring the topic to life. Do not give a title and do not start with sentences like "Have you ever..." or "Hello dear readers..", simply write the content without these introductory phrases.

Cosmo-1B – Huggingface and Cosmopedia dataset

- Model: trained a 1B LLM using LLaMA-2 architecture on Cosmopedia

Benchmark	WinoGrande	ARC Easy	ARC Challenge
Cosmo-1B	54.2	56.8	33.0
TinyLlama-1.1B-3T*	59.1	55.2	30.1
Qwen1.5-1.8B	59.4	59.0	34.9
phi-1.5	68.7	73.1	48.0

Table 1: Common Sense Reasoning Benchmarks

Benchmark	PIQA	HellaSwag	MMLU	OpenBookQA
Cosmo-1B	71.3	55.1	32.4	35.4
TinyLlama-1.1B-3T*	73.3	59.2	25.9 (2.5T)	36
Qwen1.5-1.8B	74.1	60.9	45.9	34.2
phi-1.5	75.6	62.6	43.2	48.0

Table 2: Language Understanding and Knowledge benchmarks (* from TinyLlama's Github).

Viet Law: a case study of using Synthetic data

Overview

- Training LLM on Vietnamese Legal Domain
- Model:
 - [Qwen2-1.5B](#)
 - [Sailor-1.8B](#)

Dataset - Pre-training

- **WikiHow subset from [Cosmopedia](#) (150k entries)**
- **Raw legal documents (300k documents)**
- **Legal articles + Reading Comprehension QA (what, why, how, common reasoning, NLI, etc):** filter legal documents which are normally use for answering legal QA, most of them are Law, Code, Decision. These documents are splitted into articles, which are used for citation in legal question answering. Articles (from laws) + few-shot QA → QA from articles (generate by Gemini flash)

SỞ HỮU NHÀ Ở

Mục 1. QUY ĐỊNH CHUNG VỀ SỞ HỮU NHÀ Ở

Điều 6. Quyền có chồ ở và quyền sở hữu nhà ở

1. Cá nhân có quyền có chồ ở thông qua việc đầu tư xây dựng, mua, thuê mua, nhận tặng cho, nhận thừa kế, nhận góp vốn, nhận đổi, mượn, ở nhờ, quản lý nhà ở theo ủy quyền và hình thức khác theo quy định của pháp luật.
2. Tổ chức, cá nhân có nhà ở hợp pháp theo quy định tại [khoản 2 Điều 8 của Luật này](#) có quyền sở hữu đối với nhà ở đó theo quy định của pháp luật.

Điều 7. Bảo hộ quyền sở hữu nhà ở

1. Nhà nước công nhận và bảo hộ quyền sở hữu hợp pháp về nhà ở của chủ sở hữu theo quy định của Luật này.
2. Nhà ở thuộc sở hữu hợp pháp của tổ chức, cá nhân không bị quốc hữu hóa. Trường hợp thật cần thiết vì lý do quốc phòng, an ninh hoặc vì lợi ích quốc gia, tình trạng khẩn cấp, phòng, chống thiên tai thì Nhà nước quyết định mua trước nhà ở hoặc giải tỏa nhà ở thuộc sở hữu hợp pháp của tổ chức, cá nhân.

Trường hợp mua trước nhà ở thì Nhà nước có trách nhiệm thanh toán theo giá thị trường; trường hợp giải tỏa nhà ở thì Nhà nước có trách nhiệm bồi thường, hỗ trợ và thực hiện chính sách tái định cư cho chủ sở hữu nhà ở theo quy định của pháp luật. Trường hợp trưng mua, trưng dụng nhà ở thì thực hiện theo quy định của pháp luật về trưng mua, trưng dụng tài sản.

Điều 8. Đối tượng và điều kiện được sở hữu nhà ở tại Việt Nam

1. Đối tượng được sở hữu nhà ở tại Việt Nam bao gồm:
 - a) Tổ chức, cá nhân trong nước;
 - b) Người Việt Nam định cư ở nước ngoài theo quy định của pháp luật về quốc tịch;
 - c) Tổ chức, cá nhân nước ngoài theo quy định tại [khoản 1 Điều 17 của Luật này](#).

**Example of
Vietnamese legal
document**

Why using QA in pretraining phase ?

- Vietnamese legal is difficult to understand
- Split legal documents make articles looks like the same.

⇒ Using QA for deep understanding the legal documents + make differences on articles (then model learns the knowledge, not the patterns) + increase quantity of domain data

SFT dataset - Syllogism QA

Syllogism QA: Input is a legal question and output is answer which is formatted to syllogism answer.

Format of syllogism answer:

<Citation>

<Question>

<Conclusion>

Chain of Thought (CoT)

(a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The answer is 8. X

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are $16 / 2 = 8$ golf balls. Half of the golf balls are blue. So there are $8 / 2 = 4$ blue golf balls. The answer is 4. ✓

SFT dataset - Syllogism QA

CoT (Chain of Thought) is a prompting technique that encourages step-by-step reasoning ⇒ Improves problem-solving and decision-making in language models

Syllogism QA is the combine of CoT and legal reasoning in real life -> CoT data for legal domain ⇒ improve legal ability and legal reasoning in model

Question: Không cấp dưỡng nuôi con sau ly hôn có được quyền thăm con không?

Major premise

Căn cứ pháp lý cho câu trả lời bao gồm:

Điều 81 Luật Hôn nhân và gia đình 2014

Căn cứ tại Điều 81 Luật Hôn nhân và gia đình 2014 có quy định về nghĩa vụ, quyền của cha, mẹ trực tiếp nuôi con đối với người không trực tiếp nuôi con sau khi ly hôn như sau:

Điều 83. Nghĩa vụ, quyền của cha, mẹ trực tiếp nuôi con đối với người không trực tiếp nuôi con sau khi ly hôn

1. Cha, mẹ trực tiếp nuôi con có quyền yêu cầu người không trực tiếp nuôi con thực hiện các nghĩa vụ theo quy định tại Điều 82 của Luật này; yêu cầu người không trực tiếp nuôi con cùng các thành viên gia đình tôn trọng quyền được nuôi con của mình.

2. Cha, mẹ trực tiếp nuôi con cùng các thành viên gia đình không được cản trở người không trực tiếp nuôi con trong việc thăm nom, chăm sóc, nuôi dưỡng, giáo dục con.

Minor premise

Với câu hỏi: Không cấp dưỡng nuôi con sau ly hôn có được quyền thăm con không?

Conclusion

Ta có kết luận: Đối với trường hợp không cấp dưỡng cho con sau ly hôn thì luật quy định cha, mẹ trực tiếp nuôi con cùng các thành viên gia đình không được cản trở người không trực tiếp nuôi con trong việc thăm nom, chăm sóc, nuôi dưỡng, giáo dục con.

Theo đó, dù cha hoặc mẹ không trực tiếp nuôi con nhưng không cấp dưỡng cho con thì vẫn có quyền thăm nom, chăm sóc, nuôi dưỡng, giáo dục con.

Example of syllogism QA

DPO dataset

Crawl 3,000 questions and answers from the internet (legal Q&A) -> convert the answers into syllogistic form -> use questions to generate answers from an LLM -> filter out incorrect answers -> create a DPO dataset with chosen and rejected responses.

Training

- **Framework:** Unsloth
- **Pre-training:**
 - Phase 1: trained with raw legal document + WikiHow
 - Phase 2: trained with Instruction-Pretraining data
- **Supervised Fine-Tuning:** trained with SFT data (Legal QA + General QA)
- DPO: trained with DPO dataset

Evaluation

- Datasets:
 - **ViLaw QA:** 6.5k entries for testing legal abilities across various task including memorizing (3k multiple choices QAs), comprehension (1.5k multiple choices NLI QAs), application (1k syllogism QAs), overall (1k multiple choices QAs)
 - **MMLU-VI:** from [ViLLM-Eval](#)
- Setup: [lm-evaluation-harness](#)

Experiment results

model	vilawqa-defi (5 shot)	vilawqa-nli (2 shot)	vilawqa (0 shot)	vilawqa (5 shot)	mmlu-vi (5 shot)
Qwen-1.5B	0.4703	0.2733	0.3312	0.3407	0.451
vilaw-qwen	0.4998	0.304	0.3501	0.3548	0.3985
vilaw-qwen-v1	0.5111	0.3047	0.3673	0.3689	0.3905
vilaw-qwen-v2	0.5208	0.3026	0.3768	0.3768	0.3848
vilaw-qwen-instruct-v1	0.5595	0.316	0.4113	0.405	0.4104
vilaw-qwen-instruct-v2	0.5774	0.297	0.4301	0.4223	0.4273
sail/sailor-1.8B	0.5623	0.285	0.4317	0.4568	0.2757
vietgpt/sailor-1.8B	0.565	0.292	0.4537	0.4521	0.2818
vilaw-sailor-v1	0.5825	0.2949	0.4741	0.4553	0.281
vilaw-sailor-v2	0.5835	0.2956	0.4898	0.4772	0.2806
vilaw-sailor-instruct-v1	0.6484	0.311	0.4804	0.4788	0.2623
vilaw-sailor-instruct-v2	0.6391	0.3075	0.4725	0.471	0.2619

Vilawqa-syllo result

model	Relevance	Accuracy	Fluency	Coherence	Overall
vilaw-qwen-instruct-v2	3.11	1.54	3.97	3.21	2.12
vilaw-sailor-instruct-v2	3.05	1.54	3.86	2.96	2.07

DPO vilawqa result

model	vilawqa-defi (5 shot)	vilawqa-nli (2 shot)	vilawqa (0 shot)	vilawqa (5 shot)
vilaw-qwen-instruct-v2	0.5774	0.297	0.4301	0.4223
vilaw-qwen-instruct-v2-dpo	0.5801	0.3033	0.4207	0.4239
vilaw-sailor-instruct-v2	0.6391	0.3075	0.4725	0.471
vilaw-sailor-instruct-v2-dpo	0.6528	0.3096	0.4961	0.4835

DPO vilawqa-syllo result

model	Overall
vilaw-qwen-instruct-v2	9.8335
vilaw-qwen-instruct-v2-dpo	10.3456
vilaw-sailor-instruct-v2	9.1632
vilaw-sailor-instruct-v2-dpo	10.4211

Phần 3: Knowledge Distillation

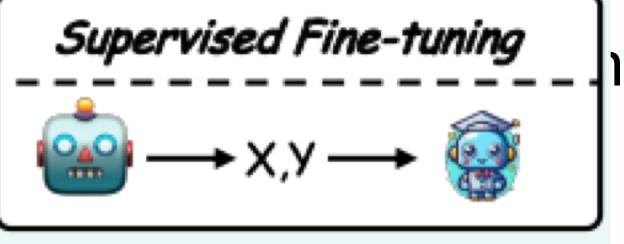
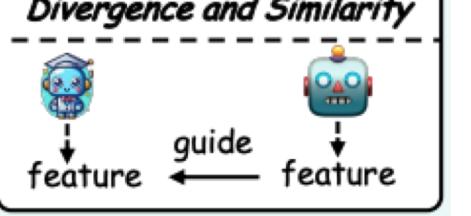
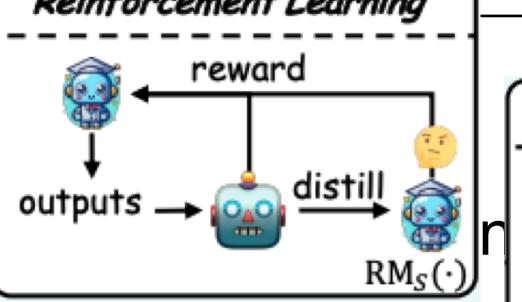
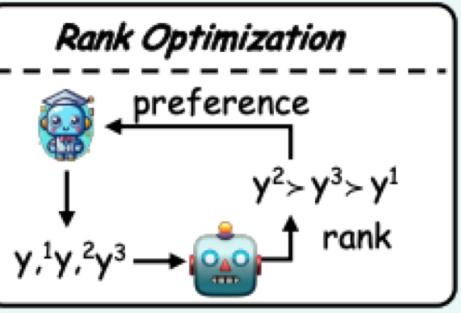
- Knowledge distillation introduced by Geoffrey Hinton in 2015 is one of the methods of the transfer learning family.
- The algorithm takes the main idea from the human learning process when knowledge is conveyed from lecturers with better understanding to learners with less understanding.
- A larger model will act as a teacher to transfer knowledge to a smaller model that acts as a student.

Knowledge Distillation

Chuyển cất tri thức từ Teacher sang Student:

- Basic distillation (using generated label)
- Probability distillation (using probability distribution)
- Hybrid distillation (using label and probability distribution)

Knowledge Distillation

Training methods	Knowledge elicitation
<p><i>Supervised Fine-tuning</i></p> 	Labeling knowledge Expansion Data curation
<p><i>Divergence and Similarity</i></p> 	Feature knowledge
<p><i>Reinforcement Learning</i></p>  <p><i>Rank Optimization</i></p> 	Feedback method Self-knowledge method

Knowledge distillation

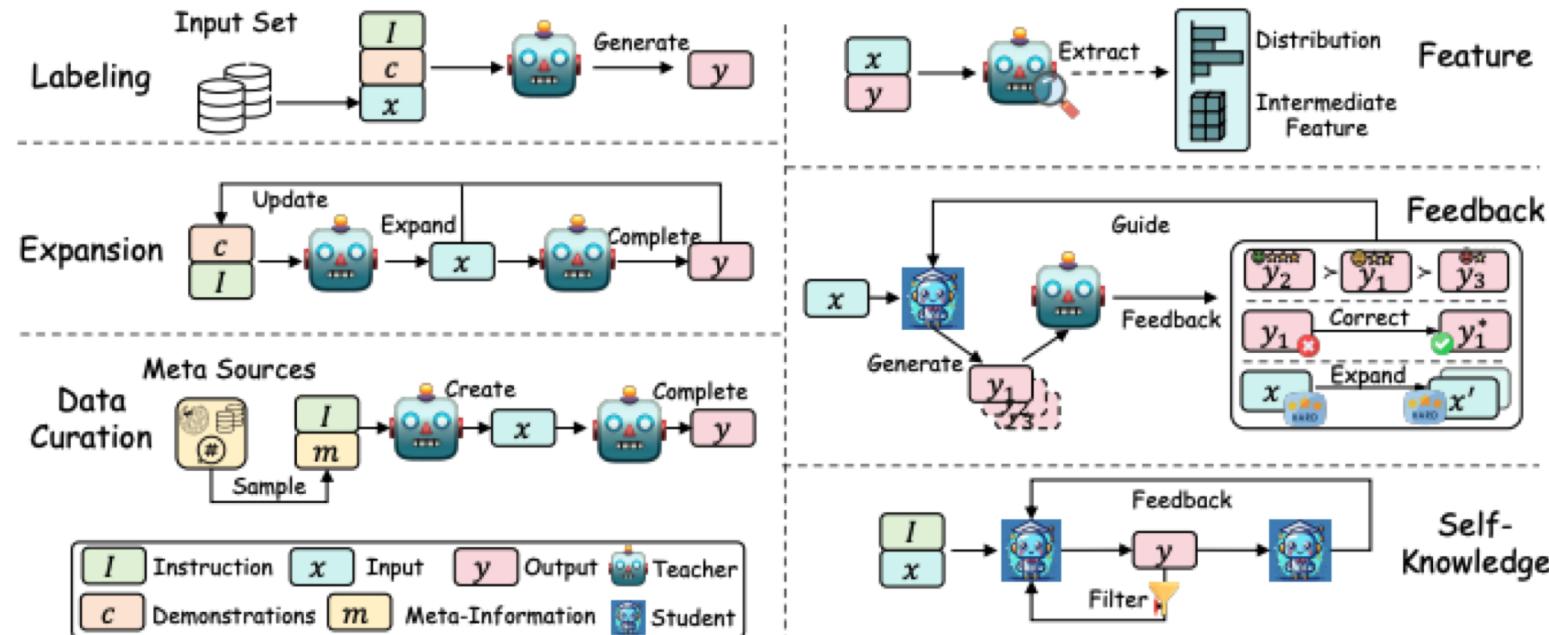


Fig. 5: An illustration of different knowledge elicitation methods from teacher LLMs. *Labeling*: The teacher generates the output from the input; *Expansion*: The teacher generates samples similar to the given demonstrations through in-context learning; *Data Curation*: The teacher synthesizes data according to meta-information, such as a topic or an entity; *Feature*: Feed the data into the teacher and extract its internal knowledge, such as logits and features; *Feedback*: The teacher provides feedback on the student's generations, such as preferences, corrections, expansions of challenging samples, etc; *Self-Knowledge*: The student first generates outputs, which is then filtered for high quality or evaluated by the student itself.

Source: [2402.13116] A Survey on Knowledge Distillation of Large Language Models (arxiv.org)

Basic distillation

The smaller model is trained to minimize the label prediction loss:

$$\mathcal{L}_{\text{label}} = \frac{1}{N} \sum_{i=1}^N \ell(f(x_i), \hat{y}_i), \quad (1)$$

where ℓ is the cross-entropy loss between the predicted and target tokens. Note that for ease of exposition, \hat{y}^i is LLM-predicted labels y^i for the model distillation case.

Probability Distillation

The smaller model is trained to minimize the negative log-likelihood between the probabilities from the teacher and the student:

$$\min_{P_S} \sum_x -P_T(x | x_c) \log P_S(x | x_c),$$

Where P_T is probability given by the teacher, P_S is probability given by the student.

Source: [gemma-2-report.pdf \(storage.googleapis.com\)](#)

- Công thức này chỉ mô tả việc tối thiểu hóa hàm mất mát dựa trên xác suất của mô hình teacher và mô hình student. Tuy nhiên, trong distillation, ta thường tối thiểu hóa **Kullback-Leibler divergence (KL divergence)** giữa hai phân phối xác suất, được viết như sau:

$$\min_{P_S} \sum_x P_T(x | x_c) \log \frac{P_T(x | x_c)}{P_S(x | x_c)}$$

- Hoặc có thể viết lại thành:

$$\min_{P_S} \sum_x P_T(x | x_c) \log P_T(x | x_c) - P_T(x | x_c) \log P_S(x | x_c)$$

- Thành phần đầu tiên $\sum_x P_T(x | x_c) \log P_T(x | x_c)$ là hằng số không phụ thuộc vào P_S , do đó chỉ thành phần thứ hai là cần tối thiểu hóa.

Distillation of Gemma 2B and 9B from Gemma 27B

Benchmark	metric	Gemma-1	Gemma-2	Mistral	LLaMA-3	Gemma-1	Gemma-2	Gemma-2
		2B	2B	7B	8B	7B	9B	27B
MMLU	5-shot	42.3	52.2	62.5	66.6	64.4	71.3	75.2
ARC-C	25-shot	48.5	55.7	60.5	59.2	61.1	68.4	71.4
GSM8K	5-shot	15.1	24.3	39.6	45.7	51.8	68.6	74.0
AGIEval	3-5-shot	24.2	31.5	44.0 [†]	45.9 [†]	44.9 [†]	52.8	55.1
DROP	3-shot, F1	48.5	51.2	63.8*	58.4	56.3	69.4	74.2
BBH	3-shot, CoT	35.2	41.9	56.0 [·]	61.1 [·]	59.0 [·]	68.2	74.9
Winogrande	5-shot	66.8	71.3	78.5	76.1	79.0	80.6	83.7
HellaSwag	10-shot	71.7	72.9	83.0	82.0	82.3	81.9	86.4

Benchmark	metric	Gemma-1	Gemma-2	Mistral	LLaMA-3	Gemma-1	Gemma-2	Gemma-2
		2B	2B	7B	8B	7B	9B	27B
MMLU	5-shot	42.3	52.2	62.5	66.6	64.4	71.3	75.2
ARC-C	25-shot	48.5	55.7	60.5	59.2	61.1	68.4	71.4
GSM8K	5-shot	15.1	24.3	39.6	45.7	51.8	68.6	74.0
AGIEval	3-5-shot	24.2	31.5	44.0 [†]	45.9 [†]	44.9 [†]	52.8	55.1
DROP	3-shot, F1	48.5	51.2	63.8*	58.4	56.3	69.4	74.2
BBH	3-shot, CoT	35.2	41.9	56.0 [·]	61.1 [·]	59.0 [·]	68.2	74.9
Winogrande	5-shot	66.8	71.3	78.5	76.1	79.0	80.6	83.7
HellaSwag	10-shot	71.7	72.9	83.0	82.0	82.3	81.9	86.4
MATH	4-shot	11.8	16.0	12.7	-	24.3	36.6	42.3
ARC-e	0-shot	73.2	80.6	80.5	-	81.5	88.0	88.6
PIQA	0-shot	77.3	78.4	82.2	-	81.2	81.7	83.2
SIQA	0-shot	49.7	51.9	47.0*	-	51.8	53.4	53.7
Boolq	0-shot	69.4	72.7	83.2*	-	83.2	84.2	84.8
TriviaQA	5-shot	53.2	60.4	62.5	-	63.4	76.6	83.7
NQ	5-shot	12.5	17.1	23.2	-	23.0	29.2	34.5
HumanEval	pass@1	22.0	20.1	26.2	-	32.3	40.2	51.8
MBPP	3-shot	29.2	30.2	40.2*	-	44.4	52.4	62.6
Average (8)		44.0	50.0	61.0	61.9	62.4	70.2	74.4
Average (all)		44.2	48.7	55.6	-	57.9	64.9	69.4

Nội dung

- I. Hiểu về ChatGPT và LLMs
- II. Một số kỹ thuật nâng cao trong huấn luyện LLMs
- III. Thảo luận về sự phát triển LLMs và Generative AI

Thảo luận về sự phát triển LLMs và Generative AI

1. LLMs đã đáp ứng được yêu cầu của Ứng dụng trong thực tiễn?
 1. Cần làm gì tiếp?
 2. LLM là nền tảng của các mô hình AI trong thực tế?
2. LLMs xử lý vấn đề Hallucination thế nào?
 1. Có xử lý được ko?
 2. Bằng cách nào?
3. Phát triển LLMs với các môi trường khác nhau?
 1. Môi trường hạn chế về resources
4. Cái gì là bước phát triển tiếp theo của AI?
 1. Có loại mô hình nền tảng mới không phải là Transformer?
 2. Kết hợp nhiều kỹ thuật mới vào LLMs hiện tại?

**THANK YOU FOR
YOUR ATTENTION**