

Transformer and ChatGPT

<https://openai.com/blog/chatgpt/>

01/05 Patrick Ruan



<https://reurl.cc/OE3Xnv>

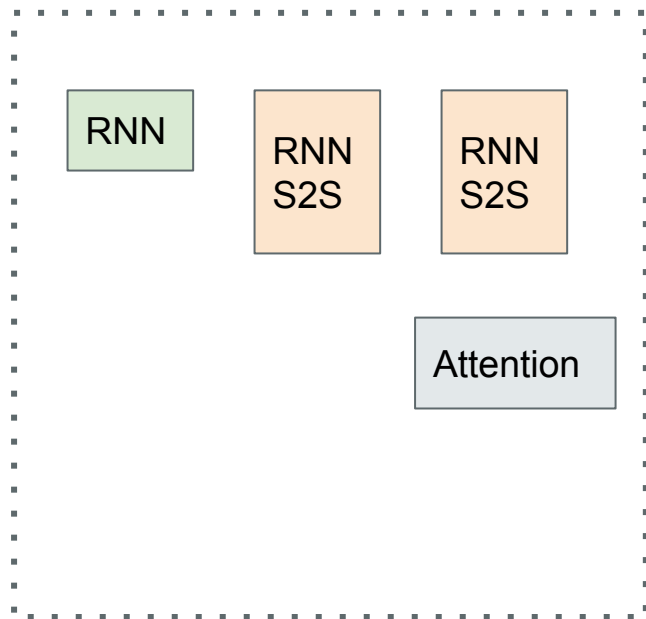


語言的困難

一個可以跟我們對談、讓我們滿意的對象

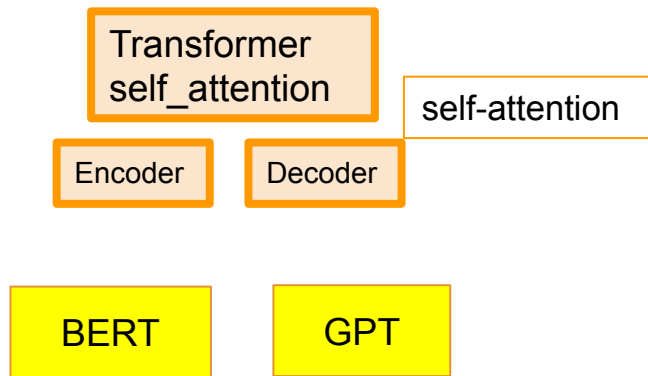
後續, 圖像世界的困難...

變形金剛與 GPT 的前生今世



Knowledge Base

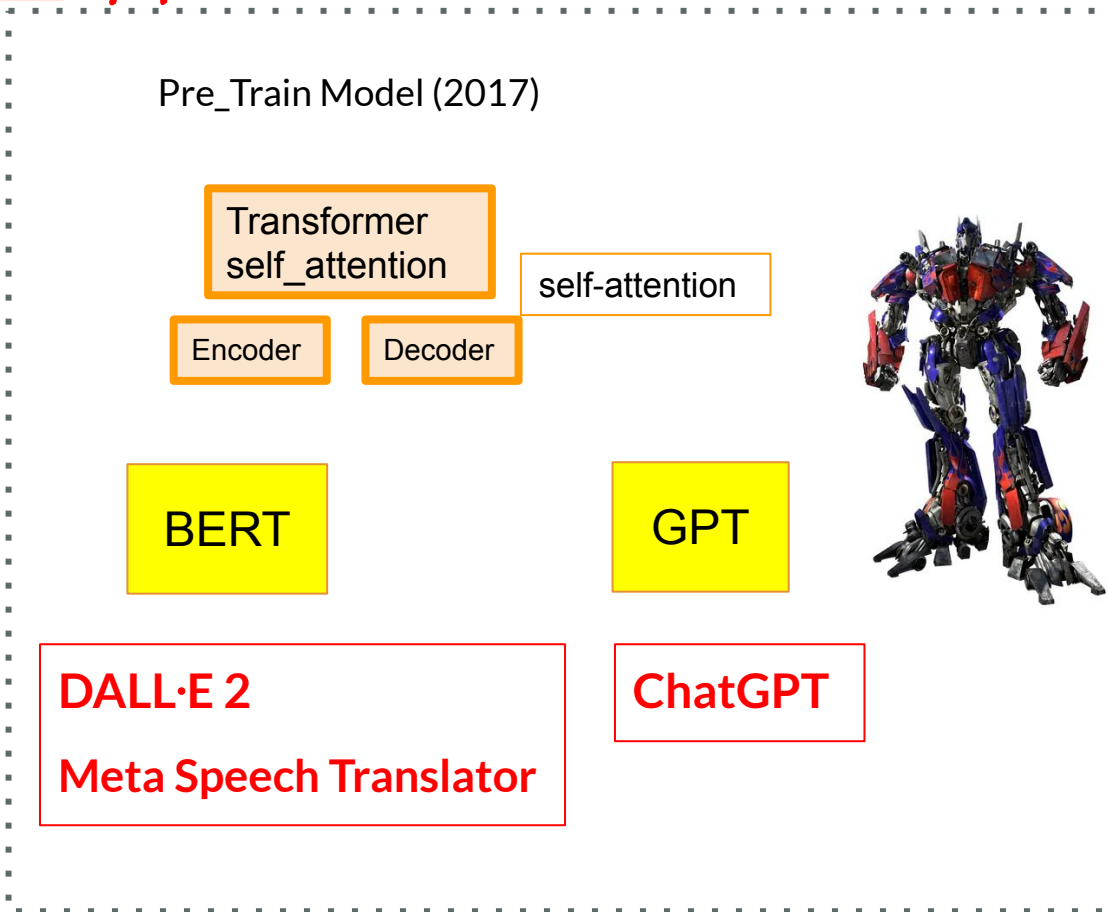
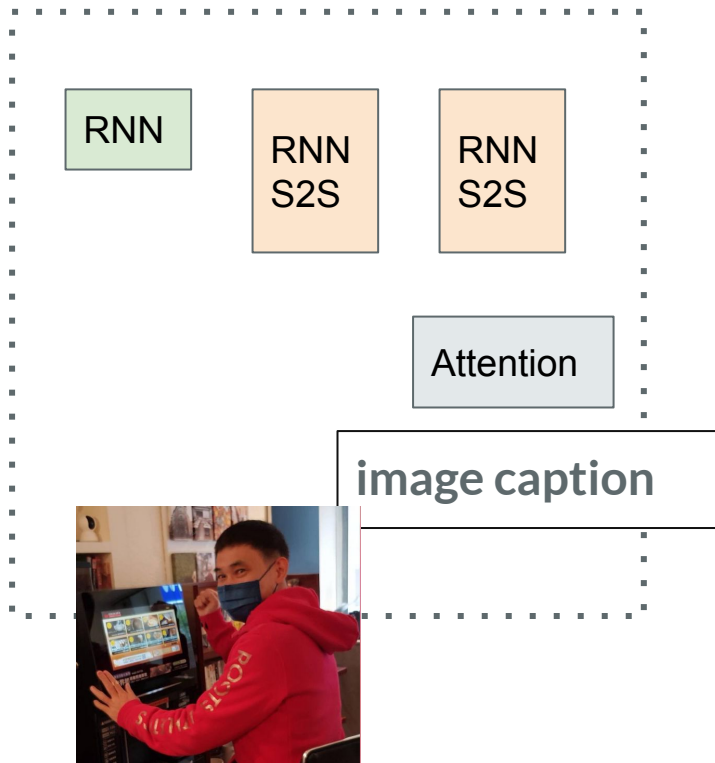
Pre_Train Model (2017)



	GPT-1	GPT-2	GPT-3
340M	117M	117M	175B
	345M	345M	
	762M	762M	
		1558M	

data: 45T

變形金剛與 GPT 的前生今世



GPT-3 是目前最先進的自然語言生成模型，能夠做到許多自然語言處理任務，包括文本生成、文本翻譯、問答系統等。但需要注意的是，GPT-3 需要巨大的計算資源，因此在一般的計算機上運行起來可能會很慢。

P

請問 GPT 3 的訓練成本



GPT-3 是目前最先進的自然語言生成模型，它的訓練需要大量的計算資源。根據 OpenAI 的說法，GPT-3 的訓練成本約為數千萬美元。



此外，GPT-3 還需要大量的訓練數據，通常是由人工標記的文本。訓練數據的成本也是訓練 GPT-3 所必需的一部分，且可能比計算資源的成本還要高。

因此，訓練 GPT-3 的總成本可能非常高，包括計算資源的成本和訓練數據的成本。

前兩頁的 GPT 資料都是 ChatGPT 告訴我的



Transformer 是一個 pre-train model, 採用 self supervised learning



Transformer BERT 條理資訊、GPT 善於說話

關於語言

1. 第一波人工智慧與第二波人工智慧的必然性

- a. 1980+ : 莫拉維克的悖論: 大人的技能我很行、小孩的能力我霧煞煞
- b. 在沒有語言之前、我們的思維是一團星雲: 索緒爾 (1900+)
- c. 可以言說的要把它說清楚、不能言說的保持沈默。: 維根斯坦 (1921)

關於語言

2. 第三波人工智慧：圖像、語言

- a. 圖像的任務比較簡單
- b. RNN \Rightarrow **Transformer**
 - i. BERT
 - ii. GPT \Rightarrow ChatGPT : **LLM**、**RLHF** : 從參數愈來愈大到人力成本大增
 - iii. MidJourney、DALIE、Meta Speech Translator
- c. 圖像的任務比較簡單？
 - i. 傅柯：瘋狂

關於語言模型 (Language Model)

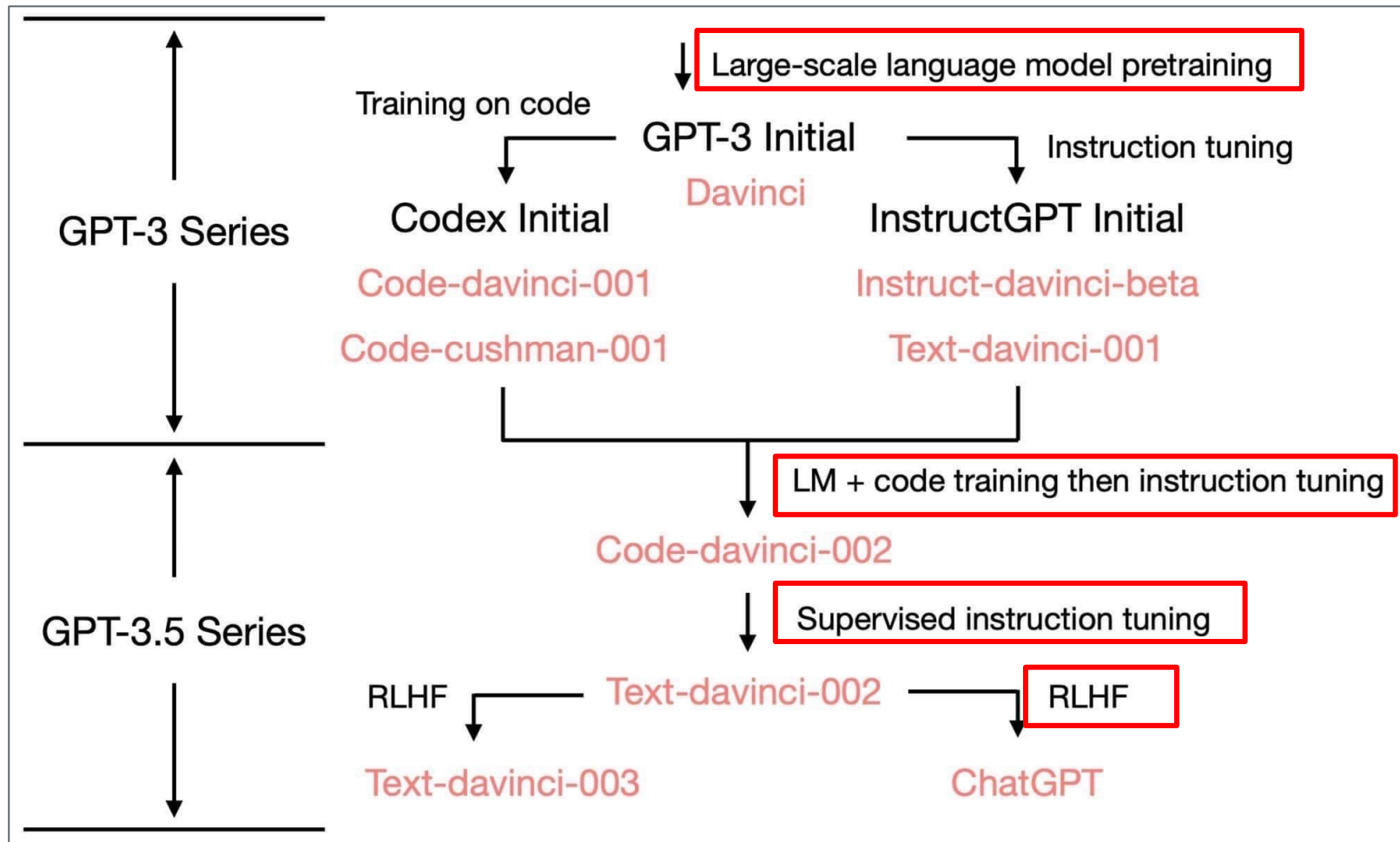
$$Y^* = \arg \max_Y P(Y|X)$$

$$Y^* = \arg \max_Y P(Y|X)$$

$$= \arg \max_Y \frac{P(X|Y)P(Y)}{P(X)}$$

$$= \arg \max_Y P(X|Y)P(Y)$$

ChatGPT



關於語言模型 (Language Model)

$$Y^* = \arg \max_Y P(Y|X)$$

$$Y^* = \arg \max_Y P(Y|X)$$

$$= \arg \max_Y \frac{P(X|Y)P(Y)}{P(X)}$$

$$= \arg \max_Y P(X|Y)P(Y)$$

今晚如何談 CHATGPT?

看看我們對他做了什麼事就知道他是哪裡弱了

InstructGPT ChatGPT

對策

0. 文字接龍 + \Leftarrow 自監督學習

1. 專家文字對話 長對話 (SFT) 

2. Reward Model (評分模型) 

3. 強化學習 (搭配 RM)



Transformer 是一個 pre-train model, 採用 self supervised learning



Transformer BERT 條理資訊、GPT 善於說話



ChatGPT 訓練很貴



GPT 很簡單: 接龍 Self_Sup, ChatGPT 很簡單: 最不好就花錢找人: SFT + RL

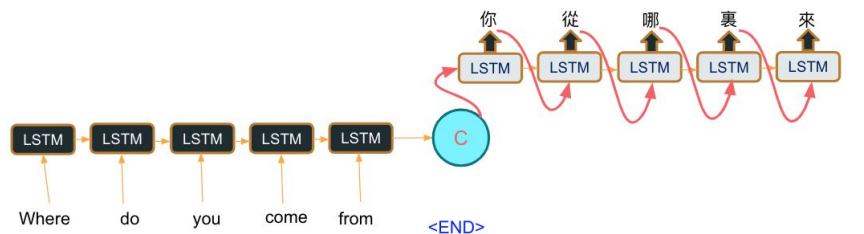
0. 基本訓練:接龍:它沒有邏輯

Machine Learning and Deep Learning

- 有監督式學習
- 非監督式學習
- 強化學習
- **自監督式學習**

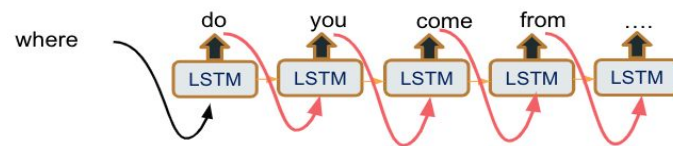
3. 從 LSTM 到 GPT

Data[["where", "do", "you", "come", "from"]]
Label[["你", "從", "哪", "裏", "來", EOS]]



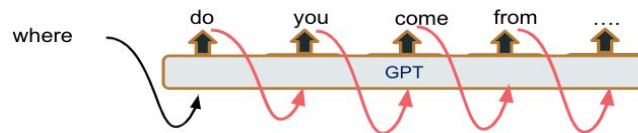
self - supervised learning

Data[["where", "do", "you", "come", "from"]]
Label[["do", "you", "come", "from"]]



self - supervised learning

Data[["where", "do", "you", "come", "from"]]
Label[["do", "you", "come", "from"]]



0. 基本訓練:接龍:它沒有邏輯

Corpus 李鐘碩與歌手IU李知恩爆出熱戀

Input

南韓 ⇒
南韓\男星 ⇒
南韓\男星\李鐘碩 ⇒
南韓\男星\李鐘碩\與 ⇒
南韓\男星\李鐘碩\與\歌手 ⇒
男星\李鐘碩\與\歌手\IU李知恩 ⇒
李鐘碩\與\歌手\IU李知恩\爆出 ⇒

Label

男星
李鐘碩
與
歌手
IU李知恩
爆出
熱戀

未來的 Inference candidate

政府、電影、男星
丁海寅、金宣虎、李鐘碩
昨天、出演
經紀公司、劇組
韓志旼、權娜拉
....
....

就算是看了 45T 的資料量。

1 T = 1024 G

就算是看了 45T 的資料量。它還是只能是會說話...

o. 基本訓練：接龍：它沒有邏輯

看再多資料，
它還是只有記憶、
沒有大腦、沒有邏輯、
只是反應很快、很會說話

如果他曾經讓你覺得邏輯...
那也是它看過的不少有邏輯的句子

就算是看了 45T 的資料量。

1 T = 1024 G

就算是看了 45T 的資料量。它還是只能是會說話...

0. 基本訓練 + :接龍



納入程式當作語料庫

0. 基本訓練 + :接龍

1. 先前提到的大量資料、語料
2. 先前提到的程式資料庫
3. 加派人力去除不好的資料、語料
4. Data Augmentation

0. 基本訓練：接龍：它沒有邏輯

基本上，深度學習模型數學數學只有兩個，

① 一個是 $z = a x + b$

② $y = \text{relu}(z)$, ie. if $z > 0$, $y = z = ax_b$,
if $z < 0$, $y = 0$

剛剛讓妹妹年輕的方法，就是他的模型沒有 bias 造成的。

這個 a 叫做旋轉， b 叫做平移，relu 叫做擠壓，然後全世界的數學式都可以經由 $\text{relu}(ax+b)$ 模擬出來... 沒有這個 b 還真的不行。

$$Y = \text{softmax}(w_3 \sigma(w_2 \sigma(w_1 x^T + b_1)^T + b_2)^T + b_3)$$

擠壓

旋轉

平移

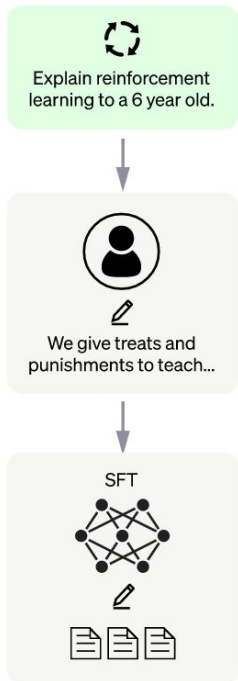
Step 1

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-3.5 with supervised learning.



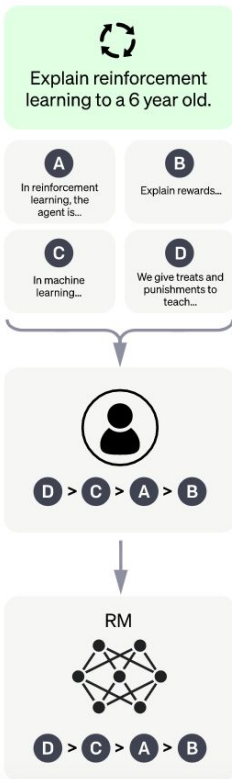
Step 2

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 the PPO reinforcement learning algorithm.

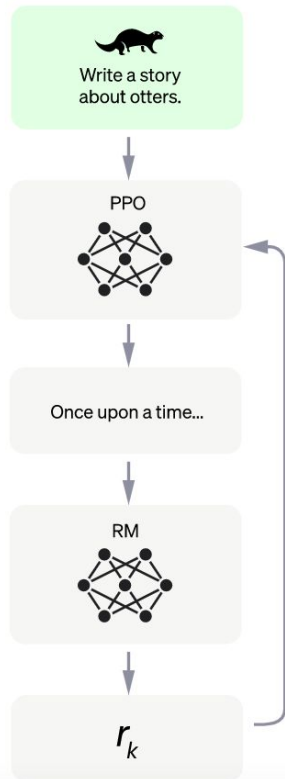
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



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.



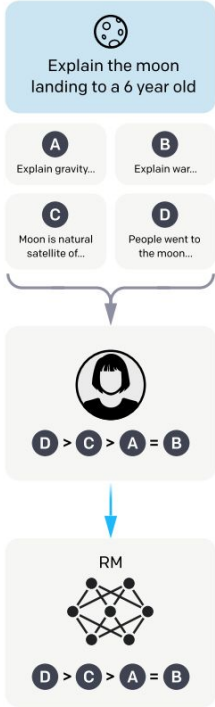
Step 2

**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

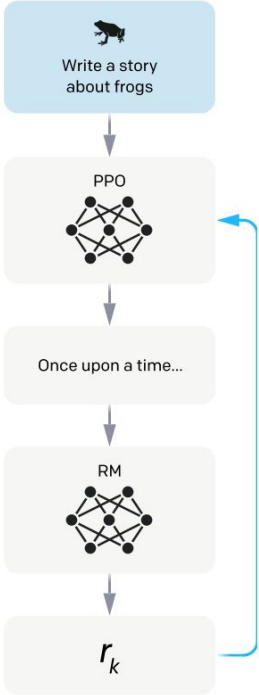
**Optimize a policy against
the reward model using
reinforcement learning.**

A new prompt
is sampled from
the dataset.

The policy
generates
an output.

The reward model
calculates a
reward for
the output.

The reward is
used to update
the policy
using PPO.



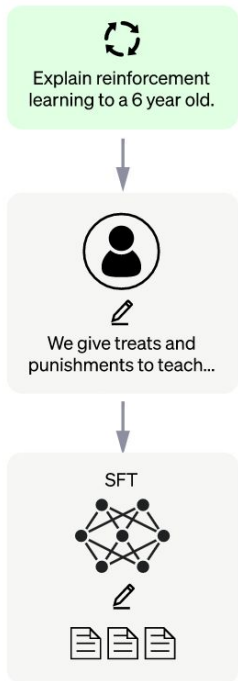
Step 1

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-3.5 with supervised learning.



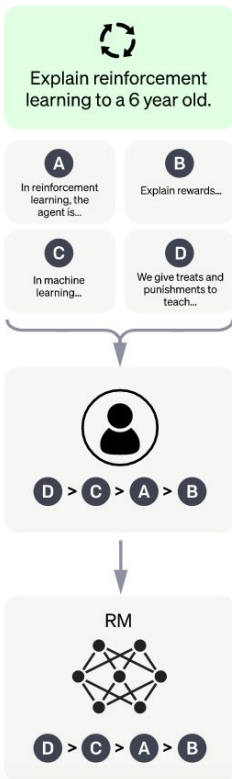
Step 2

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 the PPO reinforcement learning algorithm.

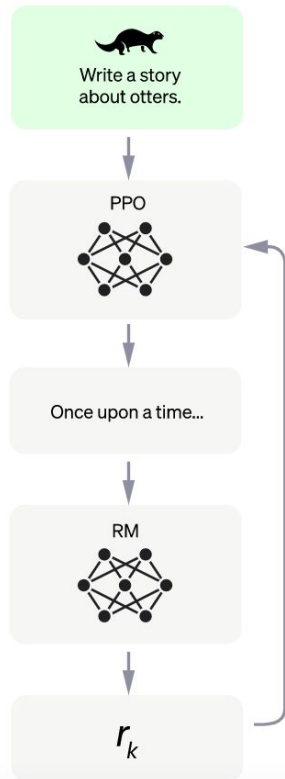
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



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.



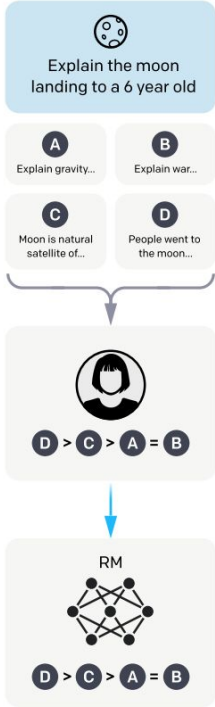
Step 2

**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

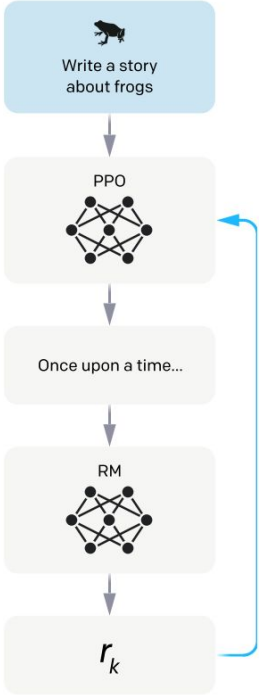
**Optimize a policy against
the reward model using
reinforcement learning.**

A new prompt
is sampled from
the dataset.

The policy
generates
an output.

The reward model
calculates a
reward for
the output.

The reward is
used to update
the policy
using PPO.



1. SFT 專家文字對話 長對話

接龍, 童言無忌

Q: 27 是不是質數

接龍王: 27 是一個質數, 因為一個質數是一個大於1的自然數, 且它沒有其他自然數能夠整除它。
27 可以被 3 整除, 所以它是一個質數。

1. SFT, Supervised Fine Tuning、長對話

專家@InstructGPT

~~27 是一個質數，因為一個質數是一個大於1的自然數，且它沒有其他自然數能夠整除它。27 可以被 3 整除，所以它是一個質數。~~

27 不是一個質數，因為它可以被3整除。質數只能被1和



專家也要編寫 instruction-like prompts!

1. SFT、長對話

專家@ChatGPt

Q1: 什麼是一個質數

ChatGPT: 質數是一個大於 1 的自然數, 且它沒有其他自然數能夠整除它。例如, 2、3、5、7、11 都是質數, 而 4、6、8、9、10 都不是質數。質數是數學中一個重要的概念, 在許多計算機科學的應用中也被廣泛使用。

Q2: 27 是質數嗎?

ChatGPT: 27 不是一個質數, 因為它可以被 3 整除。質數只能被 1 和自己整除, 而不能被其他自然數整除。

Q3: pi 是一個質數嗎?

ChatGPT: π 不是一個質數, 因為它是一個無理數。質數只能是自然數, 而 π 是一個圓周率的數學常數, 它的值大約為 3.14159, 但它的小數部分是無限的, 所以它不是一個自然數。

Q4: 目前知道的最大質數是多少

目前知道的最大質數是 2021 年發現的, 其值約為 $2^{82,589,933}-1$, 約有 24,862,048 位, 超過 7000 萬位。這個質數是由 GIMPS (大型有理數計算專案) 在 2021 年發現的。

原本只是學會一直講話的 GPT,

經過了這次“人類老師”的示範, 讓它學會了在它原本就有的“能力範圍內”挑選出人類需要的答案。

- 這些人工資料數量雖然龐大, 但是與原本的接龍語料庫比起來是非常小的。
- Fine Tuning 的影響與特性。

GPT 的對話, 常常在長句子時候前後矛盾,
所以採取人工長句子對話訓練。



teacher forcing at Seq2Seq training



Transformer 是一個 pre-train model, 採用 self supervised learning



Transformer BERT 條理資訊、GPT 善於說話



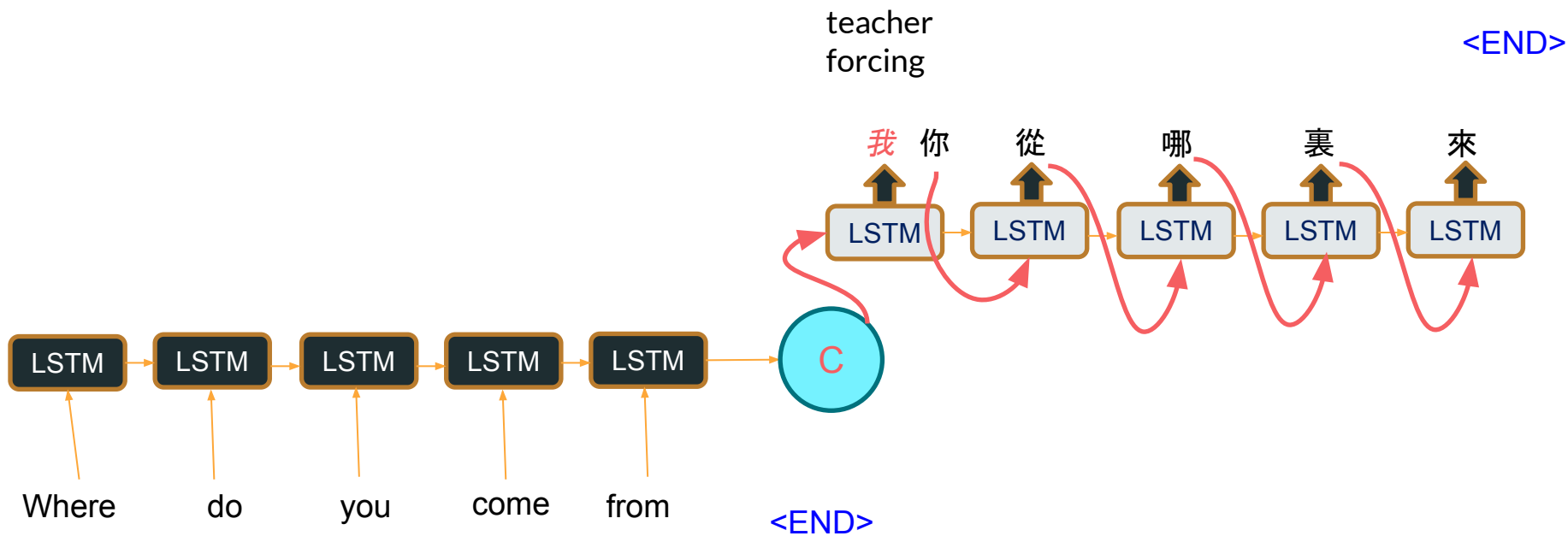
ChatGPT 訓練很貴



GPT 很簡單: 接龍 Self_Sup, ChatGPT 很簡單: 最不好就花錢找人: SFT + RL

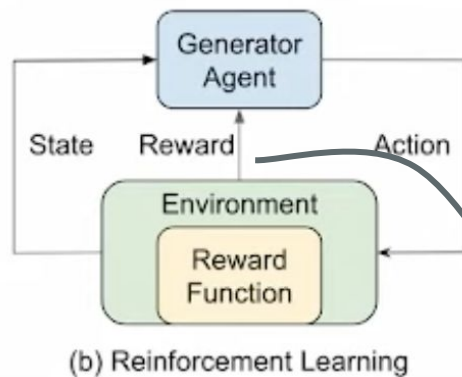


ChatGPT: 泛化結果很好



Data [["where", "do", "you", "come", "from"], "EOS"]
Label [["你", "從", "哪", "裏", "來", "EOS"]]

2., 3. Reward Model (RM) 與 強化學習 (RL)



from youtube, Vivian Chen
<https://www.youtube.com/watch?v=SOAbgV07IH8&list=PLOAQYZPRn2V5yumEV1Wa4JvRiDluf83vn&index=74>

(是人工人力！
不是聰明偷懶讓大家試用、然後
收集幾個讚、幾個噓！)

這裡的 Reward 由真人去評估好壞！

收集這些評估去訓練出一個 reward model

RM input: 一個 prompt + 數個答案(當讓是 GPT 產生)

RM output: 最佳答案到最差答案

Labels: 就是真人標注！



2., 3. Reward Model (RM) 與 強化學習 (RL)

由於 Reward Model (RM) 中 真人標註答案的優劣, 可以執行

1. Help
2. remove Hallunation
3. harmless

“For example, large language models can generate outputs that are untruthful, toxic, or simply not helpful to the user. In other words, these models are not aligned with their users. “

社會化了的 GPT 3

小型 GPT-2 寫的詩～～

月亮今晚傷感；人類將要搬到新家。

昆蟲大叫。

樹上的昆蟲呼氣。

沒有草可以吃，沒有水可以喝，沒有新鮮的空氣可以呼吸。

誰住(生)？誰死了？

首先，我將告訴您事實。

你看，

它是蟬歌的形式。

它比我所能發明的美麗和真實高出一萬倍。

它不完整，不完整，不完整。

它從我在演奏台上找到的東西開始，市長轉身，他跳著蟬的歌舞，他為我跳舞，全人類為我跳舞，而我們出於同樣的原因跳舞。

E

能不能以「AI」寫一篇迴文詩



AI是一種神奇的工具
可以進行複雜的運算
在多種領域展現它的能力
無論是工業還是醫學
它都有所貢獻

人們對AI充滿期望
希望它能夠為我們創造更多价值
但是我們也要謹慎對待
因為它不是無所不能
需要我們的監督和管理
才能讓它發揮最大的價值



Transformer 是一個 pre-train model, 採用 self supervised learning



Transformer BERT 條理資訊、GPT 善於說話



ChatGPT 訓練很貴



GPT 很簡單: 接龍 Self_Sup, ChatGPT 很簡單: 最不好就花錢找人: SFT + RL



ChatGPT: 泛化結果很好



ChatGPT: 把野小孩規範了

所以...

從 self supervised learning 倒退嚕？

2., 3. Reward Model (RM) 與 強化學習 (RL)

Pre_Trained Generative Model:

1. Helpless
2. Hallucination
3. Harmless

野孩子 ⇒ 社會化

走回頭路？

0. Self-Supervised
1. Supervised Learning: SFT
2. Supervised Learning: RM
3. Reinforcement Learning

還好的是，泛化結果**不錯**！

- (1) 新增程式碼理解和生成能力，對輸入的理解能力和包容度高，能在絕大部分知識領域給出專業回答。
- (2) 加入道德原則。即 ChatGPT 能夠識別惡意資訊，識別後拒絕給出有效回答。
- (3) 支援連續對話。ChatGPT 具有記憶能力，提高了模型的互動體驗。

生成式：
沒有邏輯推理能力、
不受控
前後矛盾

剛開始 (關於還**不錯**...)

基本上，沒預期**強 ai** 這麼快接近。

強 ai 的接近，如果超越人腦，他可能就是人類最後一個發明。

先前確實是認為 RL 會是比较有突破性的智慧技術，我們也不認為機器如此有記憶、邏輯。

突然間，記憶邏輯被用暴力取代，(?) 還藏著 RL 的潛能。

不過因為世界的多元複雜，強力暴力日久就捉襟見肘、

在還沒有捉襟見肘之前，我們心存敬意、與畏懼。

終究不是**強AI**
AI 效應

泛化結果**不錯**

終究不是強AI？

深度學習之後還是要跟邏輯理性語言的搭配！

「我們還離語言模型的終點還有很長一段距離，原因很簡單，人類遠遠不需要這麼多資料量，就可以做的比ChatGPT更好。突破的方式就是我們對今天的模型，還能怎麼樣更進一步的了解，AI還是一個黑盒子，未來會有很多研究人員像是腦科學家一樣去解剖AI。」李宏毅老師說。

人工智慧效應

所以...

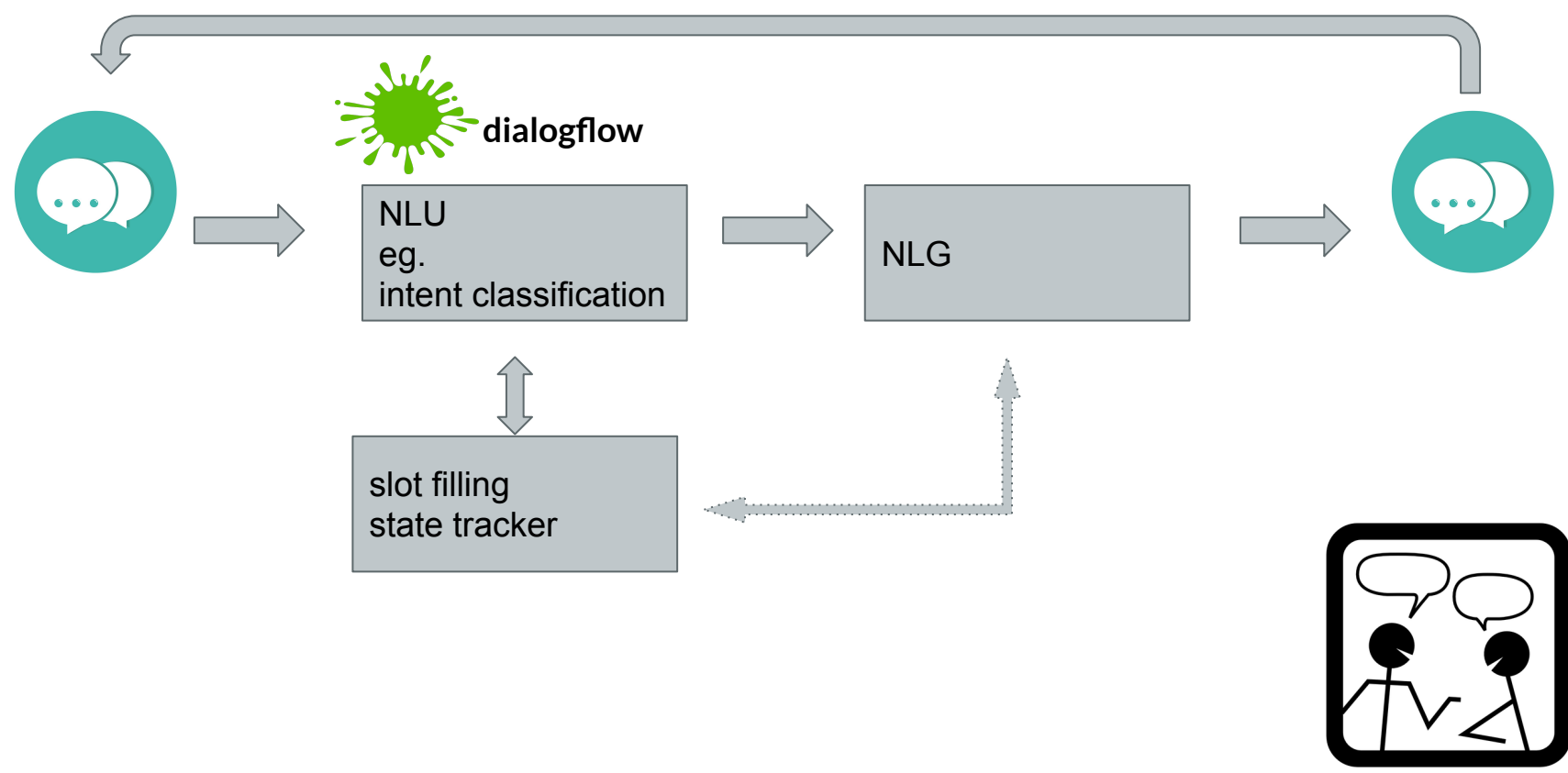
就樂觀來看，

- 已經有些工作幫助到使用者，有了加速的條件。
- 雖然從先前大張旗鼓的自監督不用標籤回到監督式學習，但是泛化結果很好。
- 畢竟，擁抱者勝！

悲觀者：

- 仍舊是監督式學習
- 重複第一二波人工智慧的困境、仍舊是玩具。

Task_Oriented example
(還有用流程方塊設計, 讓使用者被控制在固定流程裡的方式)



Patrick Ruan 派大

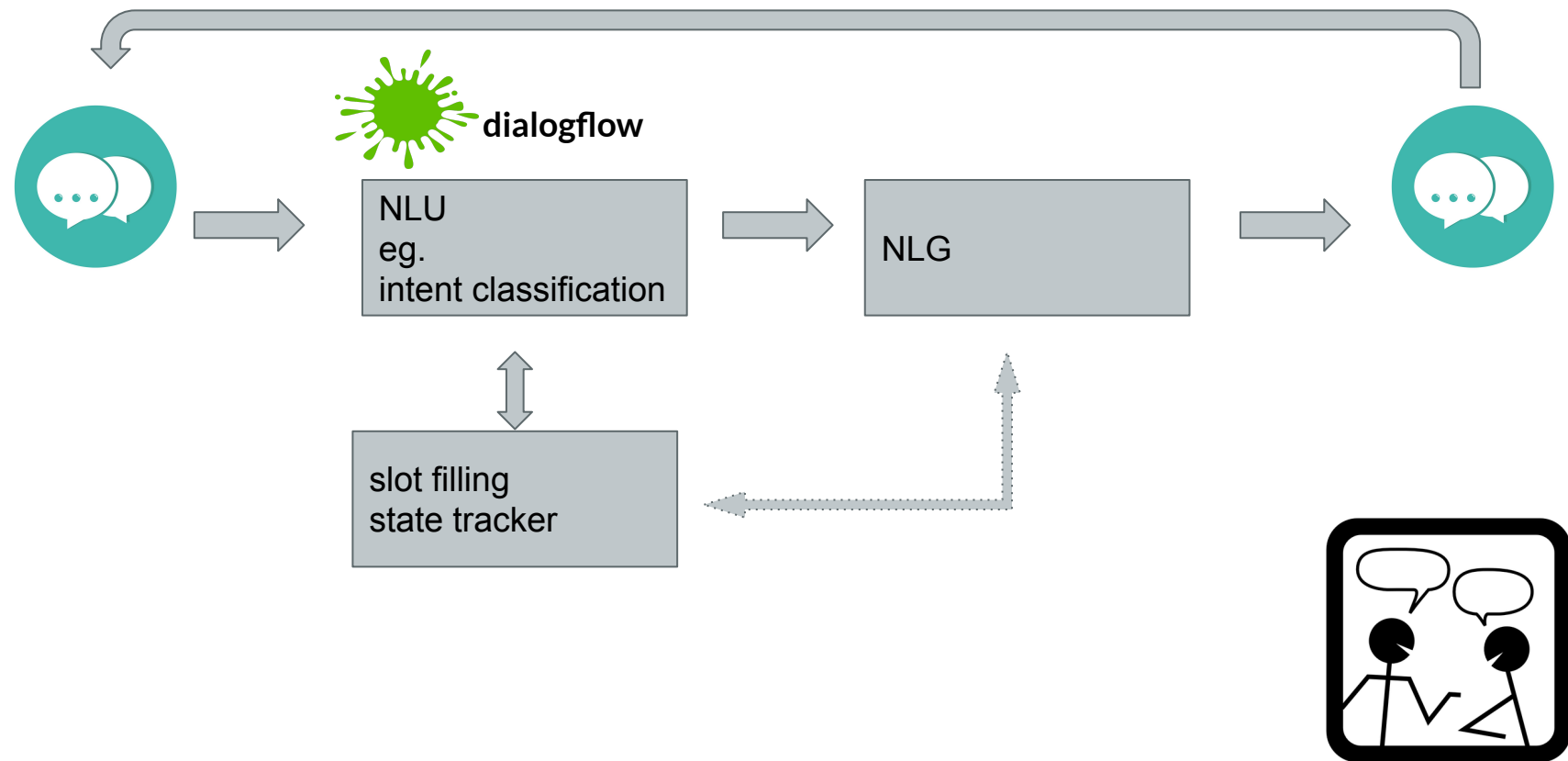
- 華實智造 數位轉型顧問
- 耐能智慧 教育顧問
- 資料科學小嫩嫩
- 經歷
 - 晨星半導體總經理室特助
 - 廣達無線通訊事業 產品部副處長

其他

1. 簡單程式(使用者可以執行驗證)
2. 程式 debug
3. 英文文法修正
4. “佈局”:小論文提供底稿。
5. midjourney + chatgpt
6. 部分取代 萬事問 google
7. 增加情境資訊, 修正歪樓
8. 聊天、陪伴, linebot (取名)

要應徵客服應該還不行！

Task_Oriented example
(還有用流程方塊設計, 讓使用者被控制在固定流程裡的方式)



強 AI 就是 AGI 嗎？

- wiki: 強人工智慧(Strong AI)或通用人工智慧(英語:Artificial General Intelligence)
- 我們迎來的是一個可以廣泛使用的人工智慧, 可以“接待”多數人, **卻不見得可以幫助到。**
- 背後更深層的意義是, 這次第一次AI達到「全民皆可用」的程度, OpenAI不只是推出一個新的語言模型, 更像是推出了一個「產品」, 也才能在社群上掀起強力的旋風。

(<https://www.bnext.com.tw/article/73508/all-about-chatgpt-2023>)

補充

看到了ChatGPT這樣的回答，我認為這一代的AI模型已經從第二語言使用者(依賴文法、規則來試著做回應)，轉換到已經大量閱讀的母語者的狀態。儘管它只是大量閱讀，並沒有**進一步的推理能力**，但已經能勝任日常生活中的很多文字生成工作(例如: 平常請行政助理處理的文書、生成即時新聞報導等等) ...

Richard Tsai

雖然不是，但是夠廣了

不用太多深度的純聊天，chatGPT聊起來感覺挺順暢的XD

但如果是有點深度的問題，chatGPT的回答雖然順暢，但品質**正確性**可能就不是那麼好了XD ...Trois Liu replied Richard Tsai

然後... AI 效應

1/Large language models like Galactica and ChatGPT can spout nonsense in a confident, authoritative tone. This overconfidence - which reflects the data they're trained on - makes them more likely to mislead.

2/In contrast, real experts know when to sound confident, and when to let others know they're at the boundaries of their knowledge. Experts know, and can describe, the boundaries of what they know.

3/Building large language models that can accurately decide when to be confident and when not to will reduce their risk of misinformation and build trust.



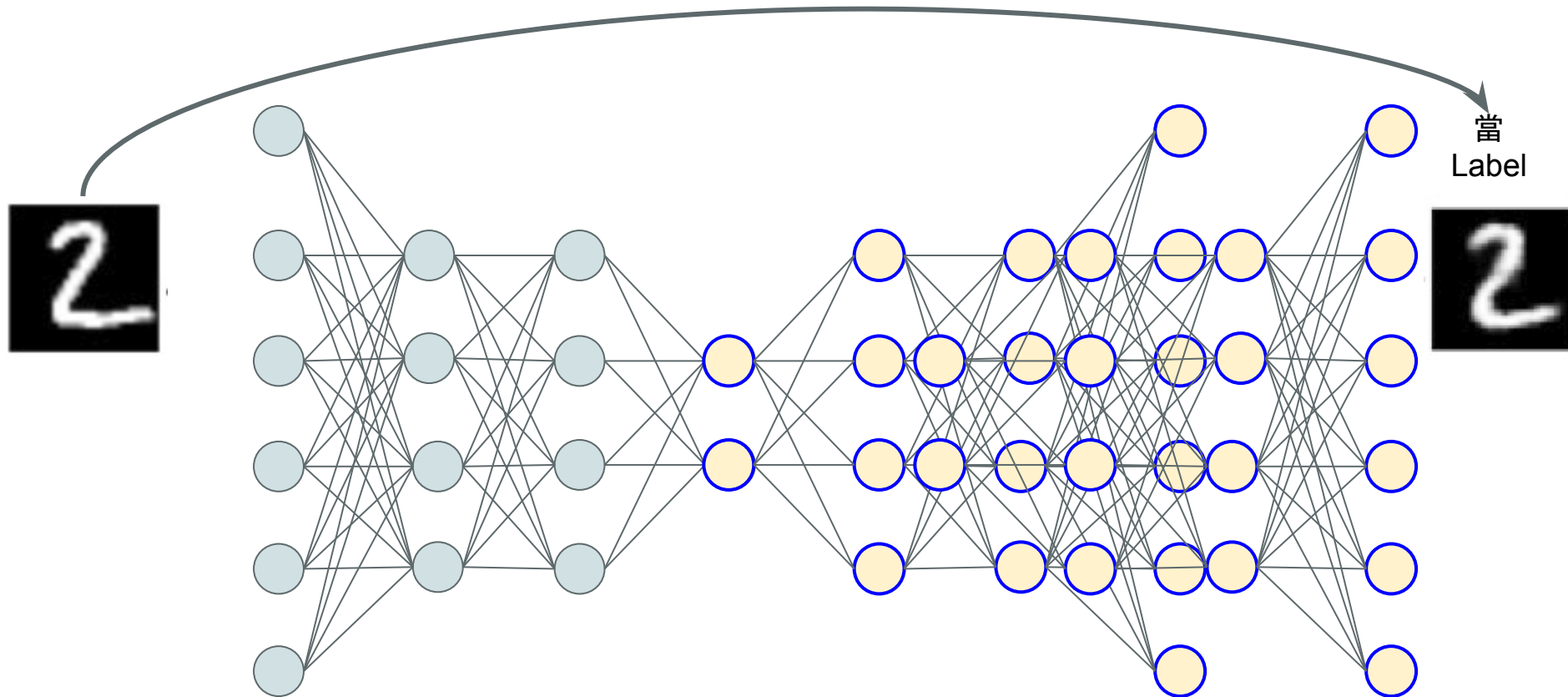
Andrew Ng  @Andre... · 2022/12/7 ...

ChatGPT is sometimes amazing, and sometimes hilariously wrong. Here, it confidently explains to me why an abacus is faster than a GPU. 😊

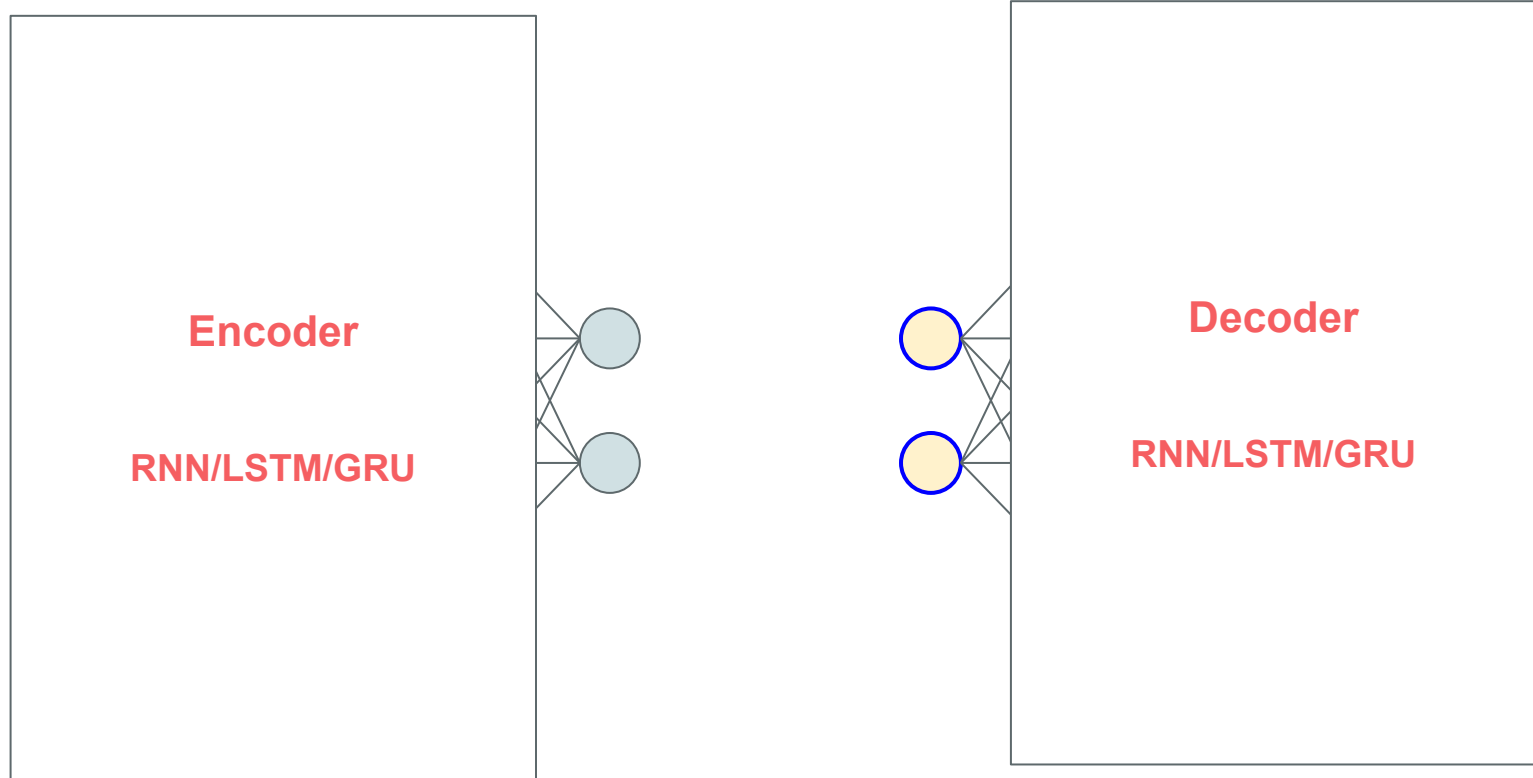
- 1750億,
- 預訓練資料量從5GB增加到45TB,
- 其中GPT 3訓練一次的費用是460萬美元, 總訓練成本達1200萬美。
- GPT3是靠了1萬顆GPUs(NVIDIA V100)

GPT: Generative Pre-Trained

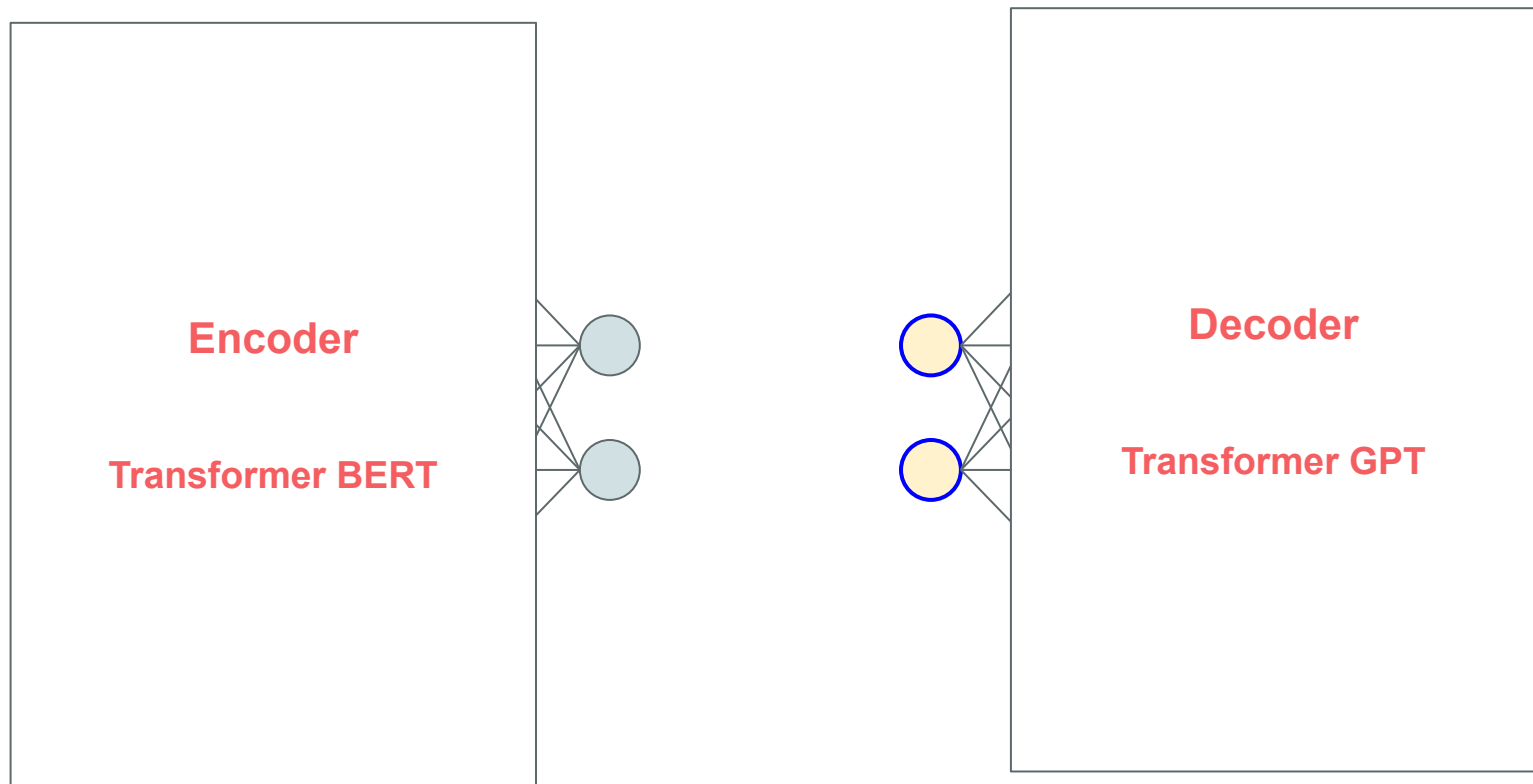
AutoEncoder(AE)



RNN Seq2Seq



2017 Transformer, self-attention



History

- 為了加強 Seq2Seq, 加入了 Attention
- Attention Is All You Need!
- Transformer
- 同樣架構,
 - 前半段克漏字、關聯性訓練就成了 BERT:輸出是一個 Embedding, 是一種 contextualize
 - 後半段接龍訓練就成了 GPT