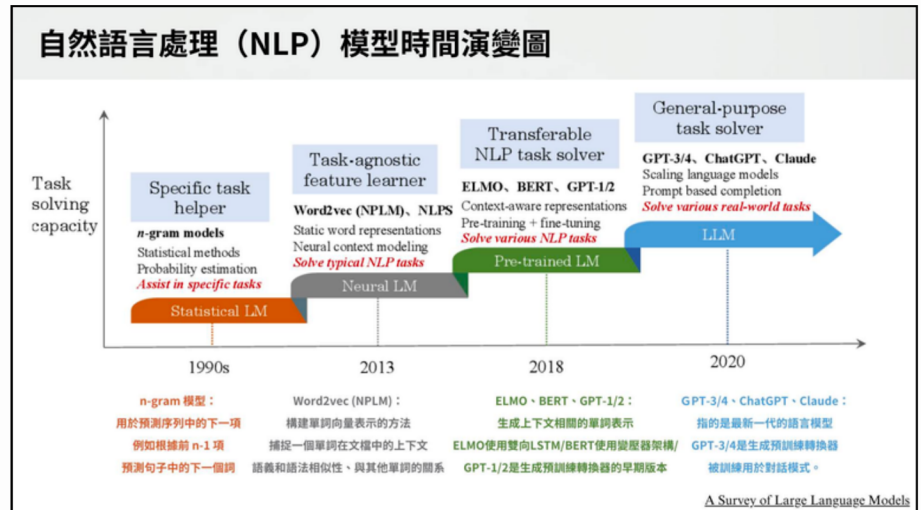


大型語言模型 不僅僅是ChatGPT！

隨著人工智慧技術的飛速發展，大型語言模型在AI領域中獨佔鰲頭。

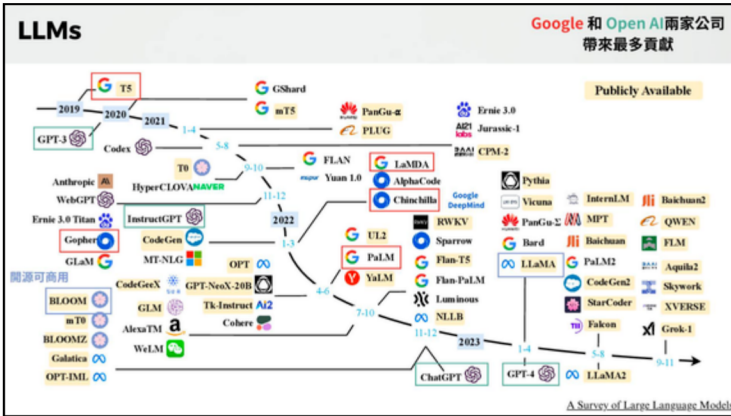
這些模型不僅能夠理解和分析文本，還能進行複雜的推理，成為人機互動的關鍵橋樑。來看看這些革命性技術背後的巨頭吧！



世界上有哪些大型語言模型？

- OpenAI的GPT系列**
 包括知名的GPT-3和GPT-4，以卓越的文本生成和理解能力聞名。
- Google的BERT和T5**
 BERT深耕於詞語含義的理解，T5則將NLP任務轉化為文本到文本的格式。
- Google DeepMind的Gopher和Gemini**
 分別專注於文本的理解與生成，以及對話和多模態任務。
- Meta的OPT、BART和LLaMA**
 分別致力於文本生成、理解與生成的結合，以及效率和擴展性。
- Hugging Face的BLOOM**
 一個多語言、高效能的開源合作項目。
- Salesforce的Codegen**
 這是Salesforce研發的一款強大的語言模型，專注於多種編程語言的代碼生成和理解，旨在提升軟件開發的效率和質量。

大型語言模型 不僅僅是ChatGPT！



GOOGLE VS OPENAI 在這場技術賽道上，兩者各展所長

在這場技術競賽中，Google 和 OpenAI 是兩家在人工智能領域處於領先地位的公司。Google的BERT和T5在多項NLP任務中取得突破，而OpenAI的GPT系列則在文本生成和理解上大放異彩。

模型的「大小」

通過模型的參數量來衡量。GPT-3擁有1750億個參數，而GPT-4的參數量則達到了數萬億級別。當參數量攀升到一定的高度，便可以開始自我學習、融會貫通，創造出未曾教導過的全新規則。

參數量：模型的腦容量 資料量：後天閱讀了多少書籍

Model	Release Time	Size (B)	Base Model	Adaptation IT RLHF	Pre-train Data Scale	Latest Data Timestamp	Hardware (GPUs / TPUs)	Training Time	Evaluation ICL CoT
T5 [82]	Oct-2019	11	-	-	1T tokens	Apr-2019	1024 TPU v3	-	✓
mT5 [83]	Oct-2020	13	-	-	1T tokens	-	-	-	✓
PanGu-o [84]	Apr-2021	13*	-	-	2048 tokens	-	2048 Ascend 910	-	✓
CPM-2 [85]	Jun-2021	198	-	-	2.6TB	-	-	-	✓
T0 [28]	Oct-2021	11	T5	✓	-	-	512 TPU v3	27 h	-
CodeGen [86]	Mar-2022	16	-	-	577B tokens	-	-	-	✓
GPT-NeoX-20B [87]	Apr-2022	20	-	-	825GB	-	96 40C A100	-	✓
Tk-Instruct [88]	Apr-2022	11	T5	✓	-	-	256 TPU v3	4 h	-
UL1 [89]	May-2022	20	-	-	1T tokens	Apr-2019	512 TPU v4	-	✓
OPT [90]	May-2022	175	-	-	180B tokens	-	992 80C A100	-	✓
NLLB [91]	Jul-2022	54.5	-	-	850B tokens	-	1536 Ascend 910	60 d	✓
CodeGen2 [92]	Sep-2022	13	-	-	400B tokens	-	768 40C A100	60 d	✓
GLM [93]	Oct-2022	130	-	-	400B tokens	-	-	-	✓
Flan-T5 [69]	Oct-2022	11	T5	✓	-	-	384 80C A100	105 d	-
BLOOM [78]	Nov-2022	176	-	-	366B tokens	-	-	-	✓
mT0 [94]	Nov-2022	13	mT5	✓	-	-	-	-	✓
Galactica [35]	Nov-2022	120	-	-	106B tokens	-	-	-	✓
BLOOMZ [94]	Nov-2022	176	BLOOM	✓	-	-	-	-	✓
OPT-IML [95]	Dec-2022	175	OPT	✓	-	-	128 40C A100	-	✓
LLaMA [57]	Feb-2023	65	-	-	1.4T tokens	-	2048 80C A100	21 d	✓
Pythia [96]	Apr-2023	12	-	-	300B tokens	-	256 40C A100	-	✓
CodeGen2 [97]	May-2023	16	-	-	400B tokens	-	-	-	✓
StarCoder [98]	May-2023	15.5	-	-	1T tokens	-	512 40C A100	-	✓
LLaMA2 [99]	Jul-2023	70	-	✓	2T tokens	-	2000 80C A100	-	✓
BaiChuan2 [100]	Sep-2023	13	-	✓	2.4T tokens	-	1024 A800	-	✓
QWEN [101]	Sep-2023	14	-	✓	3T tokens	-	-	-	✓
FLM [102]	Sep-2023	101	-	✓	311B tokens	-	192 A800	22 d	✓
Skylark [103]	Oct-2023	13	-	-	3.2T tokens	-	512 80C A800	-	✓

A Survey of Large Language Models

參數量與資料量

參數量的增加需要更龐大的訓練數據集支持，從而產生更精準的結果。

參數量：模型的腦容量 資料量：後天閱讀了多少書籍

Model	Release Time	Size (B)	Base Model	Adaptation IT RLHF	Pre-train Data Scale	Latest Data Timestamp	Hardware (GPUs / TPUs)	Training Time	Evaluation ICL CoT
GPT-3 [55]	May-2020	175	-	-	300B tokens	-	-	-	✓
GShard [104]	Jun-2020	600	-	-	1T tokens	-	2048 TPU v3	4 d	-
Codes [105]	Jul-2021	12	GPT-3	-	100B tokens	May-2020	-	-	✓
ERNIE 3.0 [106]	Jul-2021	10	-	-	375B tokens	-	384 V100	-	✓
Jurassic-1 [107]	Aug-2021	178	-	-	300B tokens	-	800 GPU	-	✓
HyperCLOVA [108]	Sep-2021	82	-	-	300B tokens	-	1024 A100	13.4 d	✓
FLAN [67]	Sep-2021	137	LaMDA-PT	✓	-	-	128 TPU v3	60 h	-
Yuan L0 [109]	Oct-2021	245	-	-	180B tokens	-	2128 GPU	-	✓
Anthropic [110]	Dec-2021	52	-	-	400B tokens	-	-	-	✓
WebGPT [84]	Dec-2021	280	-	-	300B tokens	-	4096 TPU v3	920 h	✓
ERNIE 3.0 Titan [111]	Dec-2021	260	-	-	-	-	-	-	✓
GLM [112]	Dec-2021	1300	-	-	280B tokens	-	1024 TPU v4	574 h	✓
LaMDA [68]	Jan-2022	137	-	-	768B tokens	-	1024 TPU v3	57.7 d	-
MT-NLG [113]	Jan-2022	530	-	-	270B tokens	-	4480 80C A100	-	✓
AlphaCode [114]	Feb-2022	41	-	-	967B tokens	Jul-2021	-	-	✓
InstructGPT [66]	Mar-2022	175	GPT-3	✓	-	-	-	-	✓
Chinchilla [34]	Mar-2022	70	-	-	1.4T tokens	-	6144 TPU v4	-	✓
PaLM [56]	Apr-2022	540	-	-	780B tokens	-	128 A100	120 d	✓
AlexaTM [115]	Aug-2022	20	-	-	1.3T tokens	-	64 TPU v3	2 d	✓
WLM [117]	Sep-2022	10	-	✓	-	-	128 A100 40G	24 d	✓
U-PaLM [118]	Oct-2022	540	PaLM	✓	300B tokens	-	512 TPU v4	5 d	✓
Flan-U-PaLM [69]	Oct-2022	540	U-PaLM	✓	-	-	512 TPU v4	37 h	✓
GPT-4 [46]	Mar-2023	1085	GPT-3	✓	-	-	-	-	✓
PanGu-2 [119]	Mar-2023	1085	PanGu-o	✓	329B tokens	-	512 Ascend 910	100 d	✓
PaLM2 [120]	May-2023	16	-	✓	100B tokens	-	-	-	✓

A Survey of Large Language Models

這些大型語言模型正引領著自然語言處理技術的革命。從編寫高質量文本內容到創作迷人的藝術品，再到精準的醫療診斷，它們的潛力和應用範圍正在不斷擴大。這不僅是技術進步的象徵，也是人類智慧的新篇章。讓我們共同期待AI帶來的無限可能和未來！