

高達45TB的數據參數， 演算通用人工智慧AGI

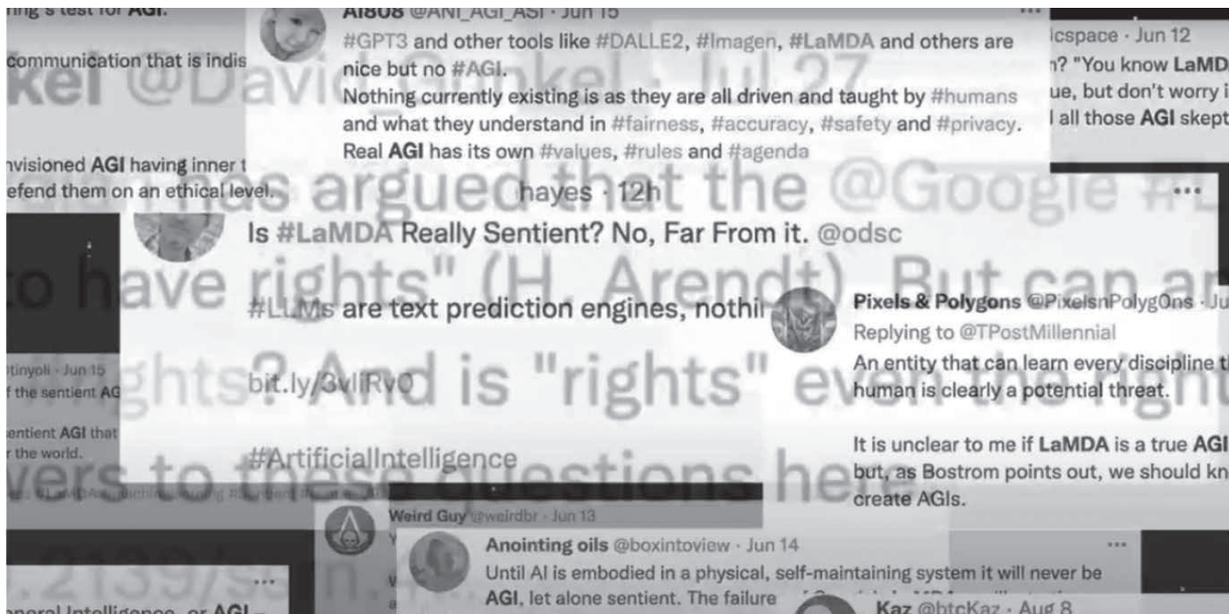
總會諮詢顧問 廖建利

前段時間，googole的一位工程師在於一個叫Lambda的AI交談之後上交了一份調查報告，聲稱AI已經具備了人格，甚至他還要求googole尊重Lambda的需求和權利。在進行實驗之前徵求AI的同意。這是一出來，馬上引起廣大群眾的熱烈討論，有討論AI到底有沒有情感的，還有討論AI有沒有人權的。而在這些討論裡，有一個詞出現了很多次，就是AGI。

什麼是AI呢？不知道大家有沒有看過電影2001太空漫遊裡的哈爾9000銀翼殺手2049裡的Joy他裡的Samantha，他們都是AGI縮寫，翻譯過來就是通用人工智慧，指的是擁有理解或學習人類所能完成的任何智慧任務能力的智慧體。

換句話說就是與人類等同甚至超越人類智慧水準的AI。AI的概念不只局限於藝術作品，在現實裡，同樣有很多人相信他會實現。比如狂人馬斯克7年5月30號就發了一條文章說：「2029年我們就要看見AGI」。很多人知道。馬斯克辦了一家叫neuralLink的一個美國神經科技和腦機介面公司，由伊隆·馬斯克和八名其他聯合創辦者創辦負責研發植入式腦機介面的科技技術。公司的總部在舊金山，於2016年成立。

目前全球的科學家還一直在琢磨怎麼上傳人類的意識。要做到這點，那麼得要靠全腦模擬技術，也就是用電腦去類比人腦的物理結構和資訊處理過程。如果這個技術成熟了，那不光是駭客帝國和人沒什麼區別的，AI也許同樣就能實現了。



不過目前這個領域的研究還處在相當初級的階段，相比之下，很多人更看好另一條路，也就是目前AI研究的主流方向。深度學習，特別是當下這條路上，有一個熱得發燙的未來科技標緻大模型。

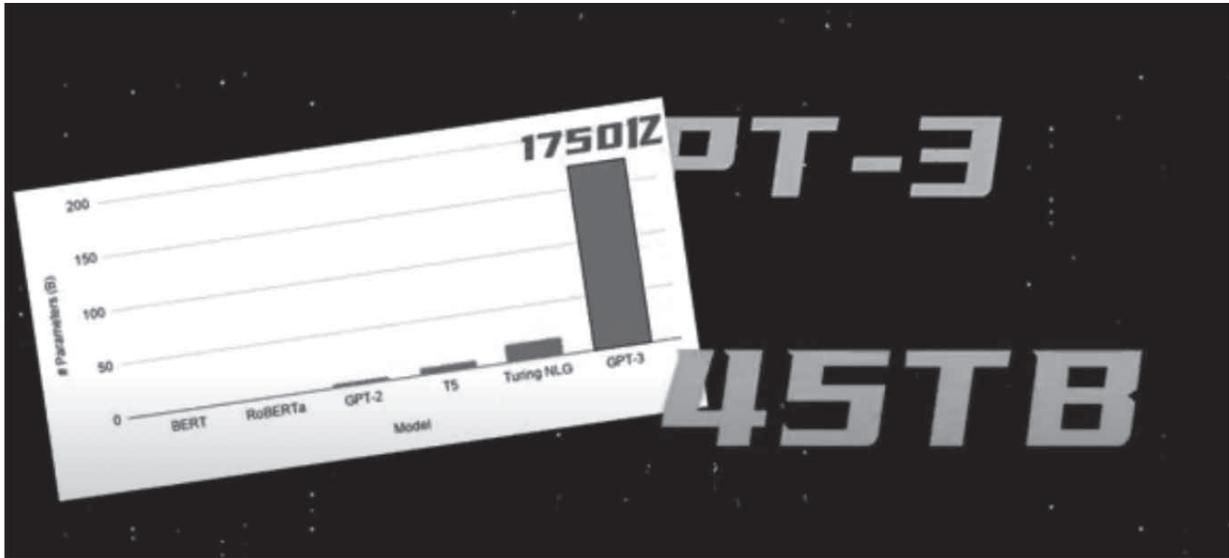
今天我們生活中能見到的AI，大多都比較偏科，比如臉部辨識的，局部掃描構圖的，能識別聲音的頻率停頓分貝在哪的，然而目前在能翻譯的這些應用都是在訓練階段，也有科學家規劃好了未來要完成的特定任務，用特定資料訓練出來的小模型，或者說換個任務，至於未來的他是否會變成人工智障或人工智慧，那就等著各種應用的模擬與試驗。

可是如果我們要追求實現通用的人工智慧該怎麼做呢？比較早的方法是想要什麼功能，就用這種功能對應的資料單獨再訓練一個模型，最後變成一堆模型的組合系統。但是

有人想嘗試另一種思路。我從一個語言模型出發，不做別的，只專注一個聯想接詞的功能，但是把它的規模做到非常的特別。大家又會怎麼樣呢？

2020年由OpenAI發佈的GPT-3史上最強AI的模型告訴我們，結果就是大力出奇蹟，它包含1751條參數，花費了45TB資料訓練，在幾乎不需要微調的情況下，發展出了寫詩並且編織出一本故事，當然這包括修改語法或加減法寫代碼。甚至設計網頁做遊戲等等能力，而且在很多工作上，也都達到了以假亂真的水準。

通過這件事，全球許多廠商忽然茅塞頓開，於是在做大做強，領先世界潮流的理念指引下，整個科技業界掀起了一場瘋狂研發大型模型、模組化的熱潮。比如國外的switch Transformers與中國的華為開發的鵬程·盤



古的鵬程盤古，宗太初等等，參數都達到了千億甚至萬億級。那麼研究製作出一個大模型需要什麼呢？除了大資料演算法之外，有一個東西很容易被忽略，那就是合適的訓練框架與深度學習模型的訓練模擬。

訓練框架就像是電腦的操作系統，往上關係到開發者的軟體編寫，往下涉及到硬體零件的資源調度，有了他們才能合理實現演算速度，運算能力，並且充份發揮整套系統的性能，以及提高訓練的效率。目前主流的深度學習框架，包括Pytorch TensorFlow等等，都支持分佈式的訓練。因為一般深度學習訓練需要的數據量非常的大，為了加速運算能力，就須要把同一個演算任務平均的分配到各個模組上進行運算，然而演算數據的併行，則是分佈演算裡最基本且應用最廣的方向策略。

問題是傳統的深度學習模型訓練又高出了幾倍的台階，而如果用傳統單一的分佈式訓練策略則費時費力，效果非常的差。那怎麼辦呢？

一種方案是直接在原有框架的基礎上做超載修改，比如基於PyTorch開發出來的，微軟的DeepSpeed和英偉達(輝達) Megatron-LM；另一種方案是從框架的設計之初，就把各種的超大規模予以併行訓練的需求考慮在內，因為這種方案得從頭到尾設計一套系統，在起初的投入非常的大，所以說現在這種的框架還不算多。

而未來AGI的寬廣格局，將引來更多的開發者和機構，參與到大模型的優化，以及更多AGI的產品被創新建立，如此才能凝眾起更多創新的力量，在激烈的競爭裡獲取更大的優勢。