美國哈佛大學
比爾蓋茲講座教授
孔祥重

無所不在的AI

人工智慧的驚奇創造力

生成式AI爆紅，在資訊科學界擁有國際級地位的美國哈佛大學比爾蓋茲講座教授，也是工研院院士的孔祥重，日前應工研院之邀發表專題演講，分享AI以文生文、以文生圖、以文生音的原理，以及最新應用情境，讓人們更深入了解人工智慧如何為人類的創造力加值。

口述／美國哈佛大學比爾蓋茲講座教授孔祥重　整理／唐祖湘

AI無所不在，我先舉一個最新發展為例，這是低軌道衛星（Low Earth Orbit Satellite；LEOS）網路應用，美國SpaceX的星鏈（Starlink）、亞馬遜Kuiper、歐洲OneWeb等業者都已發射商業用的低軌衛星，直接連上網際網路，不用再挖溝埋訊號線，但衛星網路結構很複雜，尤其低軌衛星移動速度快，對一個地面上的用戶終端站而言，一個看得見的衛星在空中停留僅大約20分鐘，訊號強度與頻率也隨之變化，網路連接十分難以控制。

AI可以解決很多這一類複雜問題，例如一群終端站可以預測衛星頻道使用狀況，知道誰在連這個衛星，彼此可避免連到同一個頻道以減少訊號干擾，同時可以經過解碼訊號學習，知道衛星承載何種模型化方案（Modulation Scheme）；另外，地

AI技術進步飛快，也帶來挑戰，大家應彼此合作，一起追趕，才更有機會成功！

面上的用戶終端站因不方便彼此溝通，各連各的衛星，如果運氣不好時，數個終端站可嘗試連接同一個衛星，互相干擾，同時其他衛星則空在那裡，近2年最新研究指出，AI透過強化學習（Reinforcement Learning），將與衛星的距離、空氣條件等元素寫成公式，地面上的用戶終端站可有效地決定與哪一個衛星連結，以提高吞吐量（Throughput）、降低訊號干擾，如此經由用戶終端站與衛星自動生成隨機使用協定（Random Access Protocol），這種作法在傳統教科書上是找不到的。

LLMs顛覆AI訓練方式

近期的AI重大進展，「預訓練大型語言模型」（Pre-trained Large Language Models；LLMs），常見到的訓練方式是先給一些字，像是一句話前頭的5、6個字，預測出下一個字或詞元（Token）是什麼，這種訓練的方式跟語言無關，知道怎麼訓練英文，同樣方法也可以用來訓練中文或其他語言，從前每個新語言都要用不同的方法來訓練，現在都徹底顛覆了。

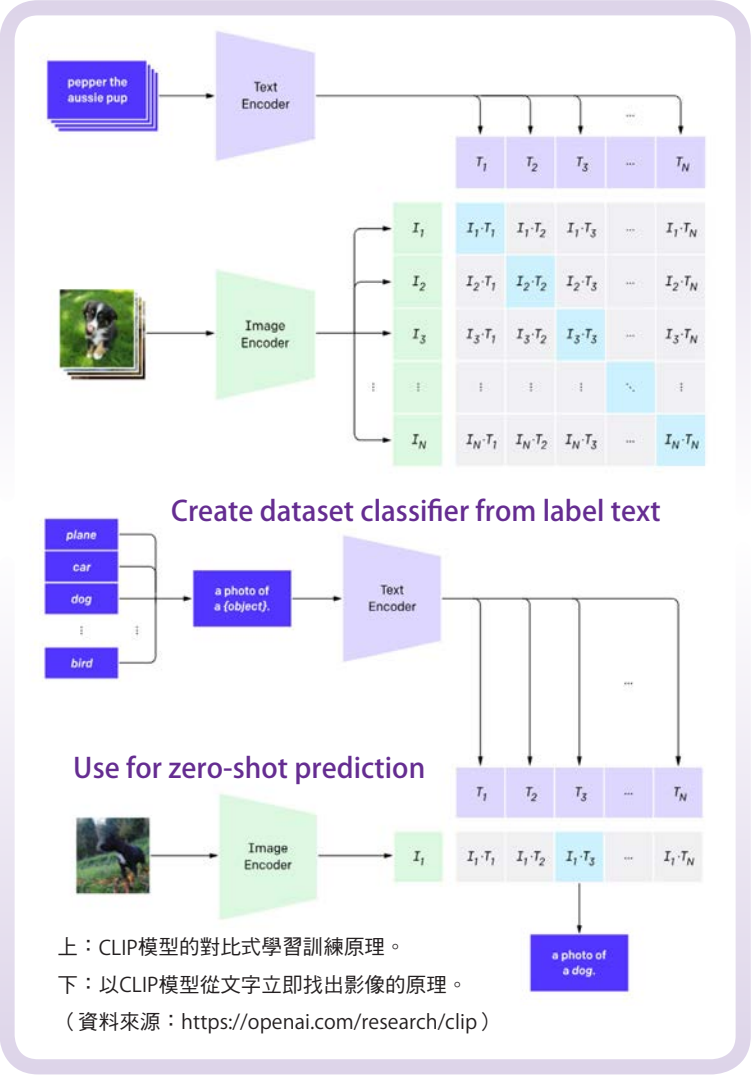
系統只要知道語境（Context）在講什麼，這篇文章或這句話是什麼意思，就可以猜出下一個字，也就是訓練找一個編碼（Encoding），把每一句、每一字、每段落都變成嵌入向量（Embedding Vector），就可以使用向量距離預測下一個字。這種模型越大越好，模型大代表資料可用的多，可支援很多不同的下游任務（Downstream Tasks）；ChatGPT也是用這樣原理，背後的生成型預訓練變換模型（Generative Pre-trained Transformers, GPT）模型規模夠大後，基礎知識完整到一個程度，不管問什麼問題，哪一種領域都可以使用。

ChatGPT是一個破壞性創新，從前要開發一項新能力，做一些新應用時，需要重新設計軟體，現在不一樣。只要語言模型進步，例如，從ChatGPT-3.5進展到ChatGPT-4，只需要把後端引擎GPT-3.5升級到GPT-4，應用能力就會變好，軟體都不用改，維護成本降低很多，這是非常大的突破。

同一個預訓練模型可以支援許多不同應用，比如GPT模型，可以支援ChatGPT、支援翻譯，BERT模型可支援分類、支援主體識別，CLIP模型可以支援圖像字幕生成、用文字分類圖片，各式各樣的資料樣態，圖像、文本、語音、音頻、深度、感測，訓練一次模型可以應用於多方面，所以稱作基礎模型，這個名稱出現至今只有18個月，現在大家都在用，進步非常迅速。

無須人工標註的對比式學習

AI在文字與圖像相關應用有非常大的演進，可做成零樣本圖像分類器（Zero- Shot Image Classifier）。例如想知道某張圖是狗還是貓，過去作法是用一個很大的數據集（例如ImageNet），以人工標註，訓練成本十分昂貴；現在CLIP模型的訓練可以用一堆已經存在有圖片的文章，把每張圖



上：CLIP模型的對比式學習訓練原理。
下：以CLIP模型從文字立即找出影像的原理。
（資料來源：https://openai.com/research/clip）

跟旁邊的標題文字敘述視為一對（Pair）數據點，100萬篇文章就有100萬對，可以大概知道這些圖是指什麼，直接用這些原始資料來訓練AI，把圖像編碼到嵌入向量，再把對應的文本編碼到附近的嵌入向量，讓文章與圖像的嵌入向量成對，就是典型的對比式學習（Contrastive Learning）。

對比式學習的訓練是利用圖像與文字兩個編碼器進行配對（如圖），成對的圖與文出現在矩陣的對角線上，對角線匹配數值大，偏離對角線匹配數值比較小，就表示訓練成功，比如要辨認某張圖像是不是狗，可將這張圖像透過圖像編碼器轉成一個向量，看哪個文字向量跟圖片向量最接近即可，且文字與圖像是可以對調的，可以從文字找圖像，也可以用圖像來找文字。

同樣方式也可以用於音頻（Audio），使用定位樣態（Anchor Modality），把文字跟圖像配對了，音頻跟圖像也對好，今後就可以用文本去找音頻，例如想聽鳥鳴，就輸入文字「Give me the audio of birds song」，系統就會輸出鳥的聲音，即使當初沒有配對鳥鳴與文字，只要有鳥的圖像跟文字的關係，以及鳥的圖像跟聲音的關係就可以間接配對，類似應用是無窮無盡的。

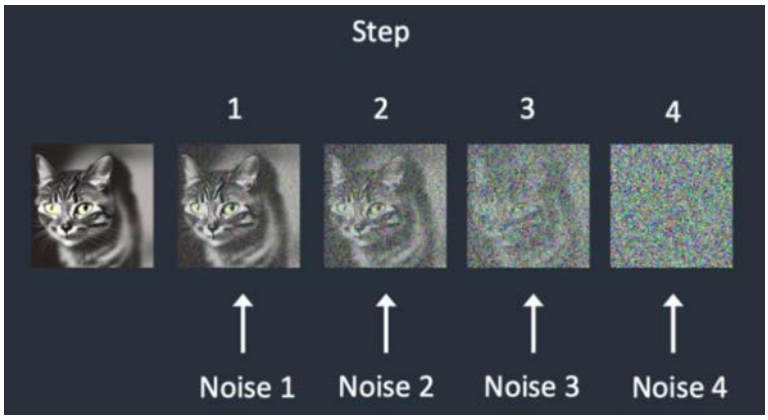
使用Stable Diffusion以文生圖

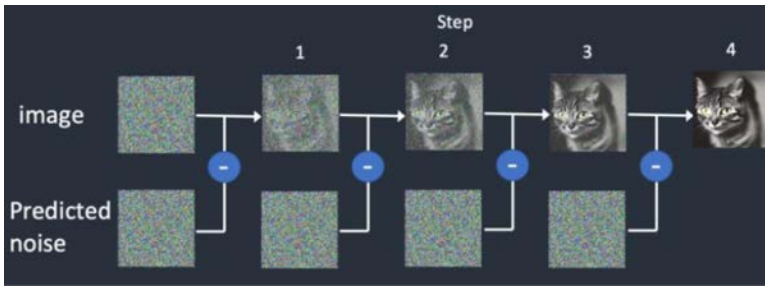
穩定擴散（Stable Diffusion）則是一個以文生圖的開源AI模型，只要輸入文字提示，經由AI運算後就能產生對應的圖片影像，例如提供「Teddy bears wearing suit, discussing business proposal around office table」文字，就會出現泰迪熊很認真在辦公桌開會的圖像。

Stable Diffusion有前向擴散（Forward Diffusion）與反向擴散（Reverse Diffusion）兩種途徑。前向擴散是在訓練雜訊預測器（Noise Predictor），像是把一張貓的圖，慢慢添加雜訊在上面，圖就越來越模糊，最後一張就是100%雜訊圖像，要訓練這神經網絡雜訊預測器模型，從輸入到輸出，經過前向擴散過程，可以知道每張圖跟原來的圖差異有多少，差異的地方就是雜訊，經由許多圖來訓練模型，每張圖雜訊可以被預測。

訓練預測雜訊之後就可以逆向，從最後一張圖開始減少雜訊（Denoise），每一張都減，最後就出現貓臉的圖，但為什麼出現的是貓臉，而不是狗臉？因為會再加文字提示，文字加雜訊預測，再加上原本的雜訊圖，3個輸入就變成下一個降噪輸出，差不多20幾步就會成功。

很多實際情況需要生成3D圖像，例如遊戲中





上圖為前向擴散，是在訓練雜訊預測器（Noise Predictor），若把一張貓的圖，慢慢添加雜訊在上面，圖就越來越模糊。下圖為訓練預測雜訊之後逆向，從最後一張圖開始減少雜訊（Denoise），每一張都減，最後就出現貓臉的圖。（資料來源：https://stable-diffusion-art.com/how-stable-diffusion-work/）

的人體或衣服，過去要提供100個角度的攝影，才能產生3D，而且很昂貴，現在只需要給一個文字提示，例如用一張某人站著的2D圖，加上文字提示，就可以做出對應的3D合成圖像，以此草圖為基礎，再去微調（Refine）會省去很多時間，可以做很多不同應用。

大型語言模型的分佈式儲存

大型語言模型愈來愈大時，最好有一部分儲存在雲端，一部分在邊緣，重點在於要怎麼切割，有幾個因素要考慮：延遲性（Latency）、邊緣裝置內存消耗（On-device Memory Consumption）、準確性（Accuracy），如果切到雲端太多，延遲性會很高，如果切到裝置太多，容量要求高，可能就沒辦法放在邊緣裝置，切割也要注意要讓雲端與邊緣溝通成本較低。

需要切割時，網路架構也要適應，硬體有多少資源，可以容許多少延遲，最後可能用NAS神經架構搜索（Neural Architecture Search）來決定網路架構與切點在哪裡。

另外，一旦要把一個模型切兩半，切的地方離開裝置送到外頭，還沒送到雲端前這段會被看到，除非做加密，但這會有管理成本，如果不做加密，放在特徵空間（Feature Space）上要傳訊到雲端的資訊，從前被認為不能夠還原輸入圖，現在發現特徵空間上面有很多資訊，圖是可以還原的，我們曾試過40個卷積層還是可以還原。

深偽技術要靠AI與之對抗

過去的對抗性圖像（Adversarial Image）研究上的一個有名的例子，是將路上的停止標誌，加上幾個白點，用人眼看還是停止標誌，但車用照相機照的圖像，經過神經網絡就看成限速標誌。大型語言模型也有類似狀況，將某句話稍微改一下，人看覺得還是原來的意思，但是對機器來講，這個句子可能從負面變成正面。

例如用AI產生上千個評論（Reviews），一般人腦對這些評論會判定為「普評」，但丟給電商的AI推薦系統來判斷卻是「好評」，因此就推薦給顧客，這就是深偽（Deepfake）攻擊，未來這些以假亂真的訊息、圖像等，可能大部分來自AI自動生成。不過，若是透過AI造假，就可能透過理解AI的規則來與之對抗，辨認出假訊息的來源。

全世界都想知道，如何在兼顧安全、成本與功能之下，發展內部使用小型AI模型，但小模型的品質普遍不佳，從大模型來縮小規模會較有意義，在這方面可用模型壓縮和微調的方法。我認為，一些基礎AI模型已相當完整，臺灣不需要從頭開始，應該發揮更多創新與想像，像堆積木一樣，將既有模型組為各式不同的系統來解決手邊的問題。面臨AI技術進步飛快，隨之也帶來挑戰，大家更應彼此合作，一起追趕、追求卓越，才更有機會成功！