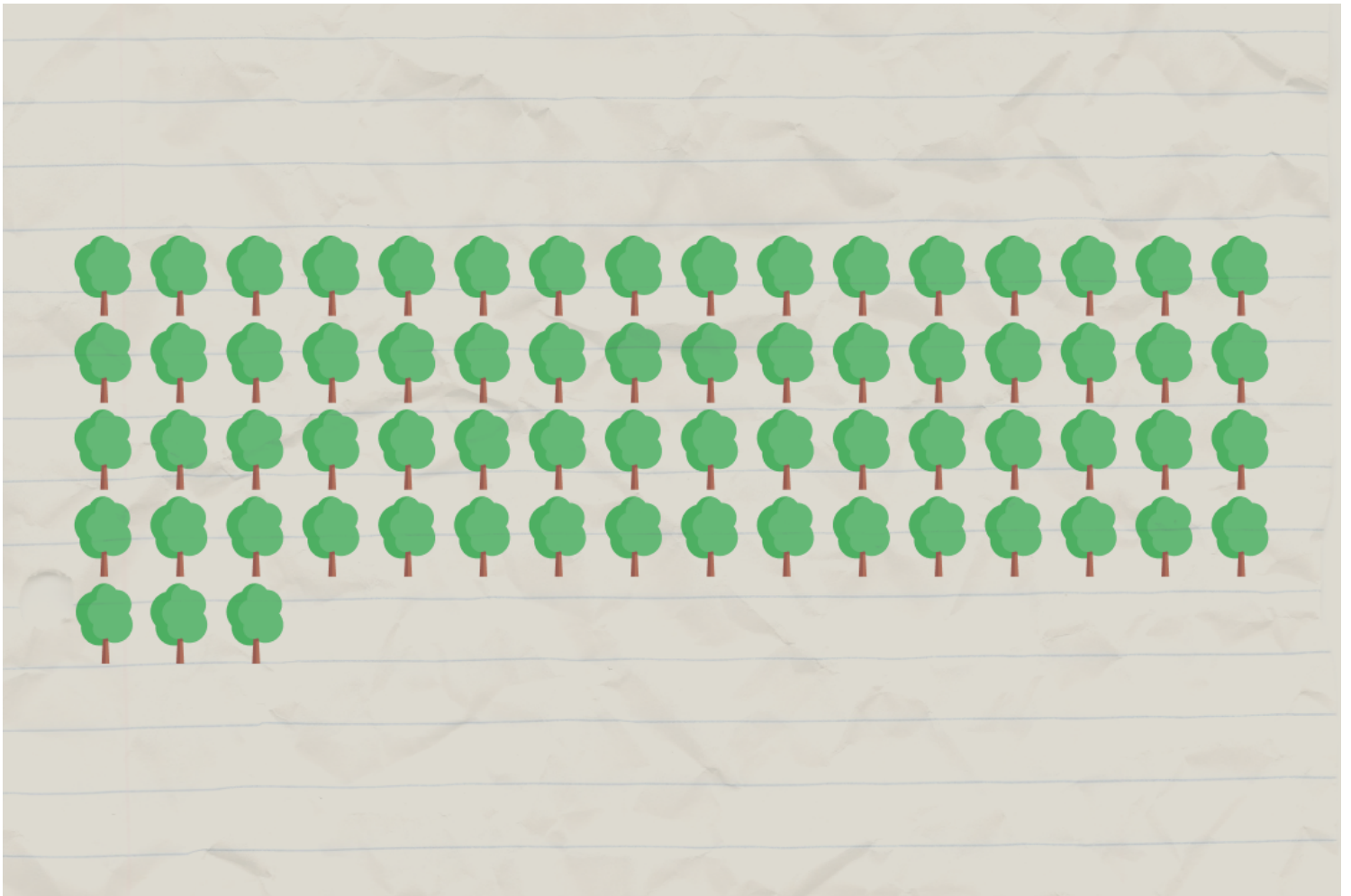


數洞 科技 深度

## ChatGPT到底有多少碳足跡？ | 一分鐘數洞

訓練 ChatGPT產生的碳排放，相當於你乘經濟艙在北京香港之間飛行一千多次。



特約撰稿人 徐彥頤 | 2023-06-01

一分鐘數洞

(徐彥頤：一名用數據看世界的社會新聞記者)

2023年是生成式人工智能系統（Generative AI）大爆發的一年：你肯定聽過大型語言模型（LLM；Large Language Models）ChatGPT、Claude、Bard、Dragonfly；生成圖像的人工智能程式Midjourney、Dall-E、Stable Diffusion等。自2022年尾的ChatGPT問世以來，各個科技公司紛紛推出了其他大型語言模型；許多人都在關注這場技術革命對人類社會帶來的影響，和可能牽涉的法規和道德問題——但其實碳排可能也是這些演算法帶來的問題之一。

目前GPT等人工智能迅速更新換代不僅帶來數據量的演變，還有與已有平台和服務的融合，譬如Bing和Duolingo已搭載GPT-4，而其服務器遍佈世界各地，使得測算其碳排放難上加難。我下載了微軟的Edge瀏覽器，問了問Bing Chat，卻並沒有得到直接相關的答案，倒是得知根據牛津大學2015年的一份報告，每一次谷歌搜索大致相當於0.2公克的碳排放。

相比而言，這些模型的訓練過程的碳足跡相對而言比較容易計算。單純訓練的過程，已可窺其排放之體量。

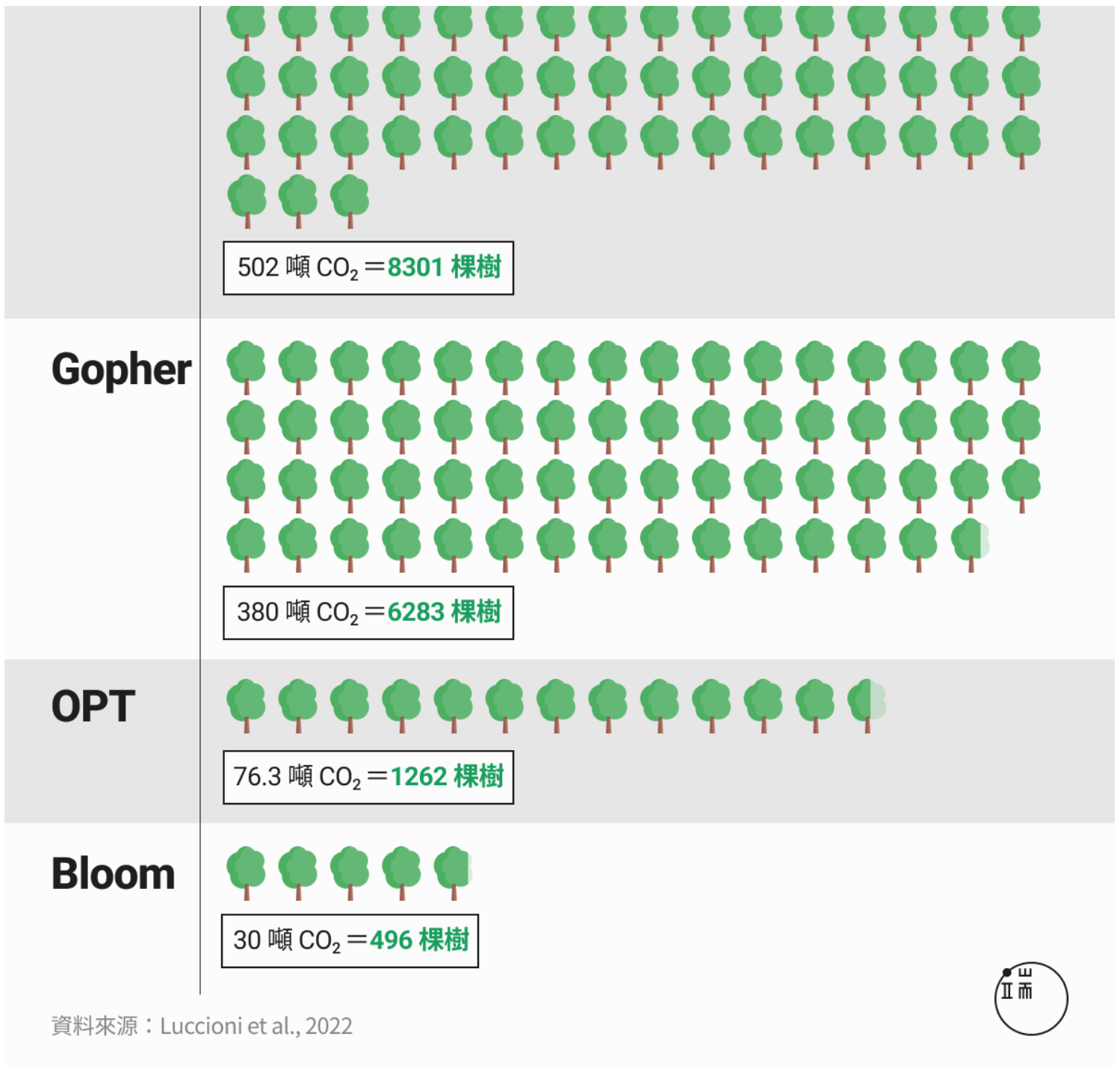
要訓練出像ChatGPT這樣的大型語言模型，需要向模型輸入海量文本，再不斷根據結果調整模型的參數和權重。開發者們在GPT-3的訓練階段交給了它570GB的文本數據進行學習，包括大把書籍和網絡文章、維基百科界面等等。其中包括3000億個單詞。ChatGPT能夠靈活自然語言運用，是因為開發人員在訓練過程中不斷調整其1750億的訓練參數（parameters）。

ChatGPT就是基於GPT-3稍作優化後的3.5版本推出的。其龐大數據量並不是獨一家。類似的BLOOM模型（全稱為The BigScience Large Open-science Open-access Multilingual Language Model）就訓練了一年之久，涵蓋了1760億參數，其訓練集使用了1.6TB的數據，在46門自然語言和13種編程語言中測試。

一篇研究報告估測，Gopher和GPT-3機器學習單單在模型訓練過程中就高達百噸，其中GPT-3位居榜首，預估排放502噸二氧化碳。這大約相當於62倍中國人均年度碳排放。根據國泰航空碳足跡計算公式估算，可以讓一位從北京到香港的經濟艙旅客來回飛行1394次。

## 不同模型在訓練過程中產生的碳排放





類似ChatGPT之類的語言模型的碳排放和幾個因素有關：其一是在模型學習訓練集過程中調試中的參數數量（其中大部分模型以數千億計），其二是研發模型公司的實體數據中心的能耗，其中包括電腦的運行以及空調的能耗等等，其三則是碳濃度，與數據中心當地獲得電力的方式有關。

但訓練過程僅是大型LLM運作的一個方面，後續還有雲端部署等步驟。在你和ChatGPT「交談」的時候，GPU,CPU以及RAM都在工作。在2023年1月，ChatGPT就收到了5.9億次訪問。丹麥數據科學家Kasper Groes Albin Ludvigsen認為其所需的GPU很有可能達到兩萬多個，這也這意味著其分佈的區域也十分多樣，很難測算其碳排放。

他認為單就全在一日所使用的電量而言，ChatGPT達到110萬到2300萬千瓦時，110萬千瓦時約為2243



而ChatGPT達到110萬到200萬小時。110萬小時約為2200個中國人月用電量。這個電量按照世界範圍內平均每千瓦時475克的碳排放計算，則相當於522噸碳排放。

同一篇報告預測BLOOM產品週期主要三個步驟中的器材製造、學習訓練和部署運作就達到50噸碳排放，約為訓練過程的兩倍。此預測仍然不含處理器原料的獲取與提煉，以及後續的硬件更迭和報廢處理。

在實驗中每次BLOOM搜索平均的能耗則為3.96瓦時，大約等于给智能手机供电到20%的电量。以同樣的全球平均碳排能效計算，大約相當於1.81克二氧化碳，略微小於10倍單次谷歌搜索，雖然ChatGPT很有可能小於這個數值。

我又問了谷歌公司的人工智能BARD同樣的問題，它給出的答案是ChatGPT和GPT-4每一次搜索能耗不高，僅在0.01-0.02克之間。

另一家法國公司Greenly則估測ChatGPT-3(第三代)的碳排放在每年238噸左右。而其中，以電力為主(160噸)，68.9噸來自服務器，9.6噸來自空調。

這些測算都是基於ChatGPT-3(三代)，而現已經更迭到第四代的GPT-4囊括的是更新的數據，更大的訪問量，和極有可能更多的排放。

出於好奇，我使用Bing Chat問了GPT-4單次搜索相當的碳排放，但始終沒有得到任何有意義的答案。聯想到我每一次問這個問題可能產生的環境污染，我關閉了這個界面。