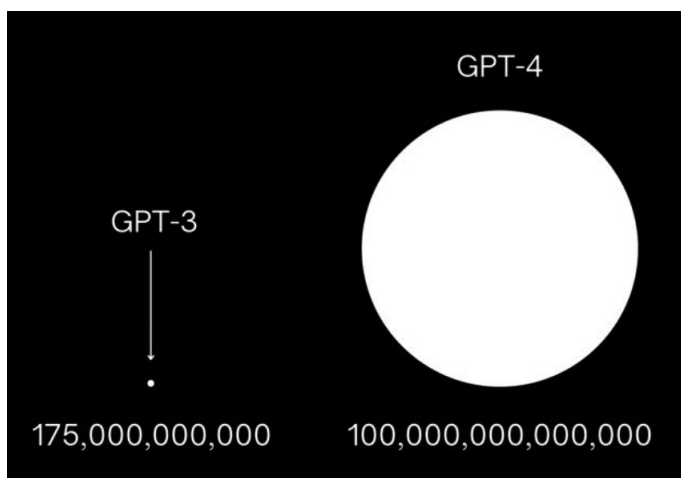


ChatGPT 則是基於 GPT-3.5 而發展出來的。官方並未公布 GPT-3.5 的參數數量，在 2023 年，OpenAI 推出參數量高達 100 兆的 GPT-4，其語言生成能力將更為驚人。現在，我們試著將 GPT-3 與 GPT-4 的大小以視覺化方式呈現（如下圖），可知其參數量的巨大差距。簡而言之，參數量越大，語言模型可以捕捉的文字的前後文關係會越精細入微，因此能生成品質更好的語言內容。



如本節一開始所言，GPT 本質上是 Transformer 解碼器。解碼器與編碼器相同的是，都可以透過自我注意機制，捕捉字與字之間的關係，也就是為整個句子中已知部分的每個字，去計算任意一個字（包含本身）對這個字的「注意力權重分數」。這個權重分布結合整個句子中已知部分的每個字的向量，去計算出下一個字的向量。但解碼器與編碼器不同的是，編碼器使用的場景是整個句子都是已知的；解碼器則是由左至右逐字生成，如同文字接龍一般。因此，解碼器的通俗稱呼就是文字接龍器。

例如，在「那隻貓正在追逐牠…」這句中，整個句子是一個序列，每個字都是一個元素；但 Transformer 還會在序列之前放上一個特殊元素 <s>，用來代表序列開頭。

若要求句子中任一個字 w 的輸出向量，只要將句中每個字的值向量 (Value vector)，與 w 對各字的注意力分數 (Score) 相乘，加總起來，就可以得到。以「牠」為例，下表將「牠」對每個字的注意力列在 Score 欄位，我們可以看到，「牠」對「貓」的注意力分數是最高的 0.35，其次是對「牠」自己的 0.2，第三是對「追」、「逐」的 0.12，將句中每個字的值向量與「牠」對各字的注意力分數相乘，加總起來，就可以得到「牠」的輸出向量。