

「萬物皆道元」，老子早知道？

文·圖／張慶瑞

科學的語言是數學，資訊是二進位元組成的機器語言，而人類則以口語或手語表達。任何語言都必須使用工具與符號來表達，口語用聲音、手語靠視覺，而盲人的點文則是有賴觸覺刺激。自然語言是指人類語言，據估算，經過幾萬年的人類文化發展下來，語言至少約有5000種到7000種，最常使用的前五種語言分別是華語（11.922%）、西班牙語（5.994%）、英語（4.922%）、印度語（4.429%）和孟加拉語（2.961%）。語言主要由聲音而發展出來，之後逐漸出現字彙，字彙其實就等價於現代AI的token。宇宙萬事萬物都可以token來隱喻表達，更奇妙的是由少數tokens，就可以生成無限的token組合來表達細緻的人類抽象情感。Token 這個概念在程式設計領域已經行之有年，它的翻譯五花八門，有令牌、符號、權杖、代幣等翻譯。而如果我們要明白其一般性的意義，歸根結柢來說，token表示的是一個抽象代替物。Token生成的特性與老子『道生一，一生二，二生三，三生萬物。』的思想有異曲同工之妙。『太極生二儀，二儀生四象，四象成八卦』，八卦可以解釋與衍生出萬事萬物，這就是AI的生成機制，由簡單的『道』出發，進而衍生完整結構與預測推論能力。道法自然，道生萬物，是自然而然，渾然天成。易經中的元亨利貞是以元為始，元的本義是混沌初開之際，始也。AI也由生成一個token開始，再生成二個，三個以至於無限多個tokens，將token與老子想法結合翻譯成「道元」，不但貫通古今，融會中西，更重要是音義均佳，符合信雅達的翻譯要旨，Token 最佳翻譯就是「道元」。

「道元」是自然語言表達的最小單位，就像「位元」是數位計算的運算單位，然而這只是技術操作的表面意義，「道元」的深層意義是各種思想與科技的象形替代單元。「道元」跟「位元」最大不同，是在「道元」並不是二進位，而是替代意識描述的象形單元，同時具有抽象與實質意義，也是AI演算法處理和學習各種文字、圖像和聲音資料類型的基本元素。「道元化」（Tokenization）是指將輸入資料拆分為基本道元的標記過程，使AI能夠分析自然語言，增強搜尋演算法文字分類的速度，有助於大語言模型（LLM）的文字生成與情感分析。大語言模型的訓練，是利用「道元」與「位元」間的轉換後快速進行數位計算，透過既有「道元」大量資

料的學習，進行更多嶄新「道元」的生成與意義重組。道元生成與重組功能遠超過簡單的位元搜尋，進而展現AI強大的「道元」的思維生成方式。

英文與中文在本質結構非常不一樣，英文是利用音符組成基本道元後再形成複雜的自然語言，而中文則是利用象形標記符號來組成基本道元語詞。英文的道元詞彙意義常是明確而單一的，而中文利用象形符號來組成的道元語詞，則常是隱晦而有多重涵義的。由於中文與英文的組成來源不同也造成在AI的「道元化」過程有極大差異。英文的「道元化」遠較中文來的容易，主要就是在於英文道元是明確且單一。例如英文的自然語言表述 “I like to go shopping at weekend”，很明顯可以像圖1中分成I，like，to，go，shopping，at，weekend 七個明確的道元標記語詞。中文則因為每個字元本身雖是明確的象形標記，但經過六書造字的指事、象形、形聲、會意、轉注、假借的時代演化，使得中文的道元標記衍生出多種意義，而「道元化」也就相對困難。中文體系的標點符號就像是在做初級的「道元化」，但是仍然因為中文多義化而造成很多解釋困擾。例如眾所皆知的明朝徐文長，在朋友家的『下雨天留客天留我不留』故事，至少就有如圖1中的七種不同「道元化」結果，也因此造成中文與英文在NLP的發展有極大差異。隱藏在中文中的「多義性」是目前「道元化」的主要障礙。

自然語言處理（Natural Language Processing, NLP）探討如何數位處理人類語言，過程中至少有認知與理解的步驟，然後根據需求與目的進行數位處理。自然語言的生成系統則是依據輸入數據，利用電腦生成更多資料後，再重構成各種自然語言。電腦的基本符號是位元，NLP的基礎則建構在「道元」之上，所以NLP的重要工作就是「道元化」，如何有效而快速的將自然語言拆解成AI可以理解的道元後，再利用機器學習生成各種新道元組合是人工智慧的主要工作。「道元化」是把複雜現象以第



圖1：英文與中文分別由拼音文字與象形文字而發展形成，以至於在目前人工智慧的自然語言模型下的「道元化」與生成過程中有極大的難易不同。道元解構後可以轉換成位元，而後利用電腦高速生成新的道元組合。(A) 英文的道元化，以及生成機制的流程，可以因應ChatGPT的問答而生成結果後，再去道元化而成為自然語言。(B)中文的「道元化」，由於道元的隱晦性高，以致於解構成道元後仍然有多重意義的可能。

一原理的歸納法，拆解成AI的大語言模型中的各種道元組合。如圖1，「道元化」後，再將道元轉化為數位輸入資料，AI可以依據輸入資料，快速進行數位處理與生成各種新穎道元。「道元化」愈精確，電腦就能愈快速產生出無數嶄新而有意義的道元組合，大語言模型的「道元化」與生成式人工智慧（generative AI）的生成組合過程，在形式意義上類似人類的理解與推理的思維意識。由於AI在重組道元過程中，引入機率性來加速最佳道元組合的產生，也導致有不可預知與不可重複性，這也類似於人類的個體行為的獨特性與無法預測性。

世上所有的學問瓶頸都在於語言，為什麼很多人學科學覺得跟聽天書一般，是因為沒有掌握好科學語言。物理學不好常不是因為物理困難，而是因為數學工具不熟練。圖2中，很多人在高中覺得物理實用且有趣，但是念大學時卻因為量子力學完全脫離現實又抽象而失去興趣。Hilbert是有名數學家，他曾說「物理對物理學家是太困難的問題」。為什麼？因為他覺得數學是科學的語言，而物理學家的數學不好，是無法真正理解物理本質。只有想法而沒有使用數學工具的能力，在處理真實問題時也只有徒呼負負。但現在時代變了，有太多現代工具可以使用，解決自然問題已經不是只有靠數學繩梯，例如精密實驗設備，數學套裝軟體，AI等都已經可以



圖2：學不好科學與出國處處碰壁是同樣原因，因為語言與技術一竅不通。(A)前往語言不通的國家，跟誰都是雞同鴨講，去哪裡都會迷路。(B)學習科學如果數學不夠熟習，物理直覺不夠敏銳，就只能眼睜睜看著兵強馬壯的同伴已經越過護城河，準備攻入宇宙真理城堡，而自己卻仍然在河的另一邊看著真理發抖，卑怯的希望胯下的數學瘦馬可以跳遠點，不要落入河中淹死。插圖由大同大學何明果校長所繪製。

逐漸將物理轉譯成多數人可以理解的自然語言。語言不熟練造成的學習障礙比比皆是，學習外國文學與歷史也有類似狀況，沒有精通外國語言前，絕對無法直接精確掌握外國文學的思想與歷史脈絡，只靠翻譯界面，只能學到半調子。許多人花很多時間學習外國語言跟數學，都只是希望透過熟練語言來更了解自然或是社會現象與文學，有了AI的大語言模型，將解決語言造成的學習障礙。未來只要學會母語，誰都有機會瞭解物理真理與莎士比亞，這將是NLP流行後的對人類歷史發展的巨大衝擊。

惠勒（John Archibald Wheeler）在1989年利用位元來解釋宇宙真理，「萬物皆位元（it from bit）」思想的起源是「每一個物理量，每一個真理，都從二進制的位元的『是或否』中展現其最終意義」，這想法類似機器學習的架構，將複雜現象拆解成多層的組織，AI每層內的問題都以『是或否』進行訓練。卡洛·西尼（Carlo Sini）說，語言是人類手上用來分析經驗所得的第一個工具箱。語言不僅將思想轉化為文字，同時人類的思考也是利用語言符號才能有效操作。因為量子物理和相對論的成功，迫使我們必須放棄以常識為起點的學習方式，多數人是因為語言的功能不足導致無法理解真理，嶄新語言的發展是理解宇宙真理的第一步。

科技歷史的發展從畢達哥拉斯的「萬物皆數字」到惠勒的「萬物皆位元」花費了幾千年時間，但位元是否真能完整描述宇宙真理，並沒有人知道。巴伯（Julian Barbour）認為位元訊息只是符號與機率，一旦脫離所代替的事物後，這些符號與機率並無任何實質意義。巴伯認為是「位元由萬物而來（bit from it）」而不是「萬物皆位元」。人類過去主要思想與歷史進展不在於語言符號發展本身，而是在道元符號所對應衍生出的事物抽象思維系統，並進而發展出的內在邏輯哲學。位元只描述了部分『是或否』的真理，而非全部真理，但道元是替代事實與邏輯內容。數位世界之後的量子世界即將進入「萬物皆量子位元」與「萬物皆道元」的量子AI時代，電腦在過去數十年由位元計算，到有大量資訊記憶後，又發展出強大搜尋功能，現在ChatGPT更顯現出初級生成推理機制。未來量子電腦發展更成熟後，創新與完整的生成思維功能也隱然在望。道元解構與生成過程對AI的推理與創新過程有絕對的貢獻，道元在AI中是一種標記符號，可以被訓練產生，並代表某些特定意義，用來組織無限有意義的思想。宇宙事物都可以道元表達，只要生成一個道元，就可以生成二個，三個以至於無限多個道元。老子『道生一』與AI的道元標記思維有異曲同工之妙，只要道元的表達意義夠豐富，就可以出現完整的推論與預測能力。道元並不是數位，道元的作法符合人類的記憶模式與智慧生成機制。老子《道德經》象元第二十五章提到，『有物混成，先天地生』，有物混成為宇宙萬事萬物，但是老子無法正確描述出甚麼是混成之物，無以名之，故稱之為「道」。道大，天大，地大，君王大，

老子企圖將宇宙萬事萬物的混成解構成可被理解的「道元」，裂解複雜天地現象後，再重新建構知識的過程，也就類似現代AI的「道元化」，以及AI自然而然，由內而外的再生成機制，『人法地，地法天，天法道，道法自然』。

人腦不斷進行樣型重組與辨認，就像AI以不斷重組不同的道元的新集體樣形態，並生成出無數的新意義組合。道元思維就是解構再重構的過程，透過徹底裂解事物到思想的本質後，再以烈火重生後的鳳凰型態展翅高飛，重新創造完全嶄新的天地。道元可以是組成宇宙的粒子，也可以是數位世界的位元組，道元更像是歷史上人類所有智慧累積的歷史精華，與無數科學量測資料所壓縮出的結晶。AI利用機器學習提煉既有道元，進而生成出更神奇的嶄新道元精華，這樣的過程難道不是像人類的思維與創新？AI的「道元化」讓我們了解宇宙本質其實就像老子的『道生一』與『有物混成』思想，宇宙就是道元生成與混成的天地。

科學家與詩人沒有不同，科學家將真理用數學符號表達，詩人將情感用語言符號陳述，都是透過符號將腦中意象介紹給普羅大眾。科學家用直覺與耐心思考，試圖找出宇宙真理的邏輯結構，並嘗試將觀察所得與結論，借用可量化的符號陳述給群眾了解。因為人腦的計算與記憶能力有限，而且思考過程常出現幻想與錯誤，所以需要不斷休息才能繼續工作。休息時需要常借用語言與數學的類比符號來記憶，與作為繼續思想與類比推理的中繼站。當讀到「狗」這個符號時，沒有兩個人的大腦反應狀態會出現完全相同結果，但卻可產生雷同的類比推論與訊息。語言類比符號是儲存資訊與腦力的休息站，利用既有知識來生成新道元，而不是每次在腦中重新發明「狗」。臺灣當年推動建構式數學的專家們，完全不了解語言類比符號是承載歷史上所有智者的知識，看似簡單的道元符號，經過無數智者的傳播和修改，才能出現豐富的抽象意義。要求每一個小學生從頭創造自己的類比符號體系，簡直就是天方夜譚。易於複製與理解的道元符號，保證了交流簡單性，而道元的抽象過程對於語言和數學的演化有更重要生成作用。AI的道元也出現一些挑戰和安全考慮因素，因為道元解構與生成的缺乏透明度，人工智慧生成被批判是「黑盒子」。因為隨機性的引入，NLP會出現訓練資料的偏差與幻覺，並生成無法判斷的錯誤結果。這些缺點是由於道元化的過程仍有瑕疵，萬物混成，溯源則清，或許仔細回顧老子之道，將可釐清AI之token的真義。

有史以來人類智慧累積的文化「道元」與觀測自然界所得的科學「道元」，

數量已經龐大到目前古典電腦無法有效而快速處理，未來量子科技才有足夠的算力利用LLM來訓練既有「道元」。外界的各種不同詢問也都可以轉成「道元」後與AI互動，進而生成更多的「道元」組合後，再轉譯成外界能了解的自然語言。道元不僅限於文字與語言，而可以各種類比型態來表達各種數據，道元可以是像素的圖像片段，或是聲音片段。在ChatGPT中，使用者輸入中的有意義詞彙都是道元。正確的「道元化」使人工智慧能夠理解語言的上下文，對於翻譯、情緒分析和內容生成至關重要。在大語言模型時代，因為AI與量子科技的結合，算力也會快速提升，量子科技與AI結合將會更快速產生更多嶄新有意義的道元，未來預期速度之快與規模之大將超過歷史上的所有道元總量，進而推動知識與科技再進化。道元的實體價值絕對超過現在對AI認知的狹隘經濟價值，道元未來所產生的文化與思想價值，才是真正的「金山銀礦」。現在因為經濟利潤而在瘋狂建造大型數據中心與搶購GPU時，更應關注道元將創造的量子新世紀中龐大的社會文化價值。

我們承接祖先智慧對宏觀世界的真理觀察所產生的類比符號系統後，利用物理原理創造出現代的科技世界。然而過去的這些語言結構並無法有效描述複雜的微觀現象，量子專家所設計出的數學語言是一般大眾所完全無法了解的科技『文言文』。如何架構出科技『白話文』讓所有人都能輕易的來描述與理解複雜量子體系的語言符號將是現代人的重大責任。LLM的「道元化」已經啟動科技『白話文』的巨輪，成為描述完整宇宙的道元時代的起點，「萬物皆道元」，一切來自道元。有詩為證曰：『宇寰非位元，魂魄亦無源，唯待道元出，方知天地言。』



張慶瑞 小檔案

1979年畢業於臺大物理學系，1988在加州大學聖地牙哥分校取得物理博士學位，1989年二月進入臺大服務，曾經擔任臺大副校長並代理校長。

張教授從事微磁學數值研究與自旋傳輸機制，已發表280篇以上專業論文並獲得28個專利。他是美國物理學會（APS）與國際工程學會（IEEE）會士，及俄國國際工程學會（RIAE）的院士。曾擔任亞洲磁性協會理事長，及臺灣磁性協會理事長暨臺灣物理學會理事長。近來曾主持NTU-IBM量子計畫，積極加速培養新興跨領域人才。近期推動量子計算相關研究，應用於新材料、新藥物合成，與財務金融領域，並創建臺灣量子電腦暨資訊科技協會，擔任理事長。於2022年擔任中原大學物理系講座教授並兼任校級量子資訊中心主任。