

調 查 報 告

壹、案由：112年10月，國內研究單位釋出之繁體中文優化大型語言模型(LLM)疑因生成爭議內容而下架，凸顯我國在推動LLM時繁體中文語料不足所產生之困境，考量LLM未來的普及性、影響力，以及所用資料集之語言分布落差可能導致的歧視或偏見，究政府機構對於推動健全LLM之關鍵要素，如語料取得、算力建置及測評，相關措施是否允當?政策是否積極?等，均有深入瞭解調查之必要。

貳、調查意見：

本案經調閱數位發展部(下稱數發部)、國家科學及技術委員會(下稱國科會)、經濟部所屬智慧財產局(下稱智財局)及能源署、交通部中央氣象署(下稱氣象署)、國防部及審計部等機關卷證資料，並於民國(下同)113年4月15、16日辦理專家學者諮詢會議，同年11月27日國科會及數發部等機關人員會同現場履勘，並於113年12月27日辦理機關詢問，已調查完畢，茲臚列調查意見如下：

一、人工智慧(AI)及大型語言模型(LLM)為國家發展的戰略性科技，主權AI及資料治理基礎更攸關國族文化及價值觀之存續，而過去政府資料開放係以施政便民及公開透明為目的，早已無法滿足現今的資料治理需求；惟數發部不僅怠於研謀對策，亦未積極發揮政府資料開放之既有機能，行政院亦未善盡推動、協調及督導之責，以至於迄未建構健全之文本資料生態系，遑論其他多模態之圖片或影音資料；以國科會所推動之可信任人工智慧對話引擎(TAIDE)模型為例，其繁體中

文(以下簡稱繁中)訓練語料之蒐整，竟無系統性規劃而僅憑計畫團隊單打獨鬥，導致最應優先配合政策之公部門訓練資料集竟僅有58個，其中更只有2個完整收錄於政府資料開放平臺，此外政府網站符合公眾領域貢獻宣告(CC0-1.0)而可以直接爬取者僅占2%，難以滿足主權AI發展之迫切需求，對於我國銳意發展AI及數位經濟實屬重大風險，行政院及數發部均有違失，應確實通盤檢討。

(一)行政院近年與AI及LLM相關政策，包括「五大信賴產業推動方案」、「智慧國家方案」、「臺灣AI行動計畫2.0」等政策及方針，足見AI在國家發展之重要性：

- 1、行政院113年9月5日通過「五大信賴產業推動方案」：包括半導體、人工智慧、軍工、安控及次世代通訊。政策願景是成為全球AI影響力中心，目標是推動AI、軟體、資安等數位經濟產業，於115年產值突破新臺幣(下同)兆元。
- 2、臺灣AI行動計畫2.0(2023-2026年)則指出，AI已成為攸關國家發展的戰略性科技。該計畫預計藉由推動資料治理，促進資料流通；並針對通用領域及如醫療、交通、金融等特定應用領域推動AI法制。
- 3、依據行政院「智慧國家方案」，國科會於112年4月啟動以繁體中文文本為基礎開發的「可信任AI對話引擎」Trustworthy AI Dialogue Engine(下稱TAIDE)，並以跨部會署科發基金「TAIDE先期計畫」推動，以蒐整繁體中文及在地文本訓練資料，開發預訓練模型。

(二)在前述AI發展中，「主權AI」係確保我國數位主權

之關鍵；根據國科會113年12月5日補充資料，估計「OpenAI模型之簡中資料不超過2%，繁體中文資料比率低於0.1%，兩者比率落差甚鉅。」換言之，倘未來我國全面導入由強勢文化資料所訓練之LLM或其他多模態AI模型，臺灣用語、圖像、社會、文化及價值觀將有消失之風險。茲將各界指出繁中文本及具我國特色之多模態訓練資料重要性綜整如下：

1、簡立峰先生於113年10月30日「人工智慧年會」中指出：

- (1) 我認為主權AI的主權是一個國族問題，我們過去沒有如此擔心過。
- (2) 我們應該擔心的是，如果未來，全世界的APP背後都加上AI，而這個AI只跑兩、三個模型，這將是人類史上前所未有的危險時刻。
- (3) 文化風險是長期風險，但短期風險已經出現了，就是「被忽略」。
- (4) 臺灣的主權AI策略還有一個重點：讓世界知道（缺乏繁中模型）是一個被忽視的問題。
- (5) 這不僅是技術風險，還有國族風險。可能有一天，臺灣所有的歷史資料都沒有被放進「未來的維基百科」，我們的文化特色被遺漏。

2、「行政院智慧國家推動小組」業界委員簡立峰先生業於112年6月30日及113年1月24日會議中指出：

- (1) 以中文問ChatGPT，其回答都受英文內容回應影響，將於不知不覺中高度美國化。因此，除提醒要長期培養語言模型專家外，更要留意國際大型模型對使用中文的臺灣之影響。
- (2) 訓練資料這部分是一個國家長遠發展很重要的問題，網路上的簡體中文資料品質雖然不好，

但相當於一億本書的量，但是網路上卻是缺少繁體中文的資料……。

3、次據113年11月27日國科會簡報指出，因國際開源LLM、生成式AI應用崛起，惟基於市場策略訓練資料大多以英文為主，至中文部分則以簡體中文為大宗，故常發生回應內容與臺灣文化、價值觀不符的情形。

4、另摘述本院113年4月15日及16日專家學者諮詢會議意見如下：

(1) 台灣人工智慧學校基金會侯宜秀秘書長：若使用國外預訓練結果，很像是用在訓練外國學生，再怎麼訓練還是洋腔洋調。

(2) 中央研究院(下稱中研院)資訊科技創新研究中心李育杰研究員：OpenAI也好，Meta也好，在訓練它的LLM裡面，華文大概是2%，臺灣是占這2%裡面的1%，所以它講的東西像中國這也不意外。

(三) 本案於資料治理所涉法令包括「數位發展部組織法」，另有「政府資訊公開法」所授權訂定之「行政院及所屬各級機關政府資料開放作業原則」及「政府資料開放諮詢小組設置要點」等。此外，「AI基本法」業由國科會於113年7月15日至9月13日進行草案預告作業。綜整前開法令，復以「行政院智慧國家推動小組」下設之4個分組中，「數位基盤分組」及「數位治理分組」係由數發部(111年8月27日成立)主責，爰政府當前於資料治理領域之主責機關為數發部，尚無疑義。

1、「數位發展部組織法」第2條規定：「本部掌理下列事項。第1項第6款：政府數位服務、資料治理

與開放之策略規劃、協調、推動及資源分配」。

2、行政院及所屬各級機關政府資料開放作業原則¹：

- (1) 第1點：為行政院及所屬各級機關（以下簡稱各機關）極大化開放政府資料，以政府資訊公開法為基礎，推動政府資料以開放格式、機器可讀且具一定品質方式提供外界使用，結合民間資源及創意，達成施政便民及公開透明之目的，特訂定本作業原則。
- (2) 第2點：政府資料開放之範圍，指各機關於職權範圍內取得或作成，且依政府資訊公開法得公開之各類電子資料，包含文字、數據、圖片、影像、聲音、詮釋資料（Metadata）等。
- (3) 第10點：政府資料之開放，應以中央二級機關為中心，由其統籌規劃所屬機關（構）資料集管理，並集中列示於政府資料開放平臺（data.gov.tw），供使用者連結下載及利用。
- (4) 第14點：各機關辦理政府資料開放作業，應建立績效管理及推廣機制，並定期評估檢討執行成效，對積極推動有功人員應優予獎勵。

3、政府資料開放諮詢小組設置要點²：

- (1) 第1點：為配合政府資料開放政策，建立跨域合作溝通平臺，擴大推動資料開放，達成施政便民及公開透明之目的，於行政院及各中央二級機關設政府資料開放諮詢小組。
- (2) 行政院政府資料開放諮詢小組（下稱行政院諮詢小組）由行政院資訊長（數發部部長兼任）任召集人，任務包括「擬訂資料開放推動政策」、

¹ 112年1月19日修正版本。

² 112年11月27日院授數多元字第1123001832號函部分修正版本。

「跨域合作溝通協調平臺」及「督導資料開放推動成效」。(第2、3點)

- (3) 第10點：中央二級機關諮詢小組應按開會頻率檢視該機關與所屬機關提報之盤點資料及民間需求之回應說明，涉機關主管法規疑義者，應會同機關法制單位檢討修正，並確認回應說明之適切性；涉機關業務流程或資訊系統問題者，應排定改善期程，定期追蹤列管；涉跨部會法規疑義者，應提報行政院諮詢小組。

4、「AI基本法」草案預告版本涉及資料治理條文：

- (1) 第15條第1項：政府應建立資料開放、共享與再利用機制，提升人工智慧使用資料之可利用性，並定期檢視與調整相關法令及規範³。
- (2) 第15條第2項：政府應致力提升我國人工智慧使用資料之品質與數量，確保訓練結果維繫國家多元文化價值與維護智慧財產權⁴。

5、另查，在「行政院智慧國家推動小組」下設之4個分組中，數發部主責「數位基盤分組」及「數位治理分組」，其任務為完備政府數位基礎、建構資料治理生態、規劃與研議提升智慧政府服務及數位治理所需之配套法制，促進國內數位治理生態結構之形成及落實。

(四)按照行政院近年推動政策及專家一再疾呼AI發展之重要性，政府理應極度重視臺灣特色繁中文本及多模態資料生態系之建立。惟查，「數位發展部組織

³ 草案說明：資料為人工智慧發展之重要元素，政府有必要確保人工智慧之創新與產業發展得以取得高品質和可追溯的資料。參考歐盟人工智慧法有關支援高品質資料近用之規定，以及美國AI行政命令要求聯邦資料長委員會研擬資料開放相關指引等規定……。

⁴ 草案說明：為確保AI訓練結果維繫國家文化價值，避免影響弱勢、多元族群權益及人民之智慧財產權，於第2項定明各機關應致力推動之事項，以完善我國資料治理機制。

法」雖明定資料治理為其職掌之一，然而直到112年1月，最接近資料治理的規定—「行政院及所屬各級機關政府資料開放作業原則」，仍僅以施政便民及公開透明為目的，尚未建立資料治理之概念或機制；而產官學界長期詬病國內資料治理環境之證據斑斑可考，尤其行政院「臺灣AI行動計畫(2018-2021年)」早在數發部111年8月成立前，就已指出資料開放及法規限制問題，數發部竟迄114年2月還在起草「促進資料創新利用發展條例(草案)」之階段，尚未送行政院審議，明顯落後治理需求，爰該部怠於研謀資料治理對策之違失，至為明確：

- 1、行政院「臺灣AI行動計畫(2018-2021年)」已直指「我國政府雖在資料開放上早先全球一步，但開放的質量以及外部業者提供相應的網路服務都有不足，在法規限制下造成收集資料不易，無法取得巨量資料來發展人工智慧運算模型」。
- 2、所謂「資料治理」及數發部所應負責之範疇，負責籌備成立數發部之郭耀煌前政務委員早在112年3月就已指出⁵國內資料治理環境亟待強化，並已預見治理環境的不完備，將斷傷資料經濟的發展。茲將郭前政務委員之意見臚列如下：
 - (1) 數發部的**主要業務**，包括負責發展資料經濟和推動智慧政府，自然要面對強化公私部門**數位資料管理、流通和加值處理的業務**。
 - (2) 所謂數位資料，……需要在整個資料**生命週期各階段**，皆有完備的治理規範和作業機制，也**包括對公私部門的資料治理成熟度輔導和評**

⁵ 郭耀煌。112年3月7日。「資料治理司」怎麼不見了？網址：<https://futurecity.cw.com.tw/article/2959>

量資料。

- (3) 國內公私部門的資料治理及個資保護成熟度確實尚有不足，相關罰則相較於所造成之損害和風險也常不相稱，自然應該加重各利害關係人的責任意識。但是，促成公私協力建立專業的輔導機制，配合前述法制規範完備化，讓政府部門和民間企業能加速提高資料治理成熟度更是重要。否則，可能讓利害關係人因害怕面臨處分和罰則，致心態和作為皆趨於保守，將不利於我國發展資料經濟和智慧政府。
- (4) 資料治理問題在數位時代趨於複雜，而且風險擴大，但除非我們要走回頭路，否則就要坦然面對，期待相關主管機關勇於承擔，社會各界積極合作。

3、綜整外界及本院諮詢專家學者詬病政府資料治理問題如下：

- (1) 簡立峰先生在「2024人工智慧年會」演講⁶指出「臺灣每個地方都在討論資料，可是沒有人覺得這是他的責任。我希望能找出負責人，否則一直討論，十年後還是討論同一個問題」。
- (2) 台灣人工智慧學校侯宜秀秘書長：應針對臺灣特色資料做出統整，如果把臺灣特色資料集整理出來，不只臺灣自己模型可以做用，也可供國外的語言模型調整訓練時可運用，政府內部文本資料可就比較沒有著作權疑慮的部分，透過各部會盡快整理出來⁷。

⁶ 「2024人工智慧年會」演講精華，天下雜誌官網，網址：
<https://futurecity.cw.com.tw/article/3548> 2024年10月30日。

⁷ 網址：<https://www.ctee.com.tw/news/20250205701828-430104>

(3) 中研院法律所資訊法中心劉靜怡合聘研究員：

〈1〉現況大家都不滿AI治理和發展，但這不是在AI出現之後才有的問題，這是先天不足後天失調，……因為Data搞不定，到了AI需要的時候，問題全部浮現，包括公部門資料如何使用等等，沒有架構可以讓資料生態系得以建立，每任政府都用鋸箭法來處理。

〈2〉過去10年一直在推OpenData，很多部會找我當OpenData諮詢委員，我都說我們法源不夠，一定搞不出來。

〈3〉其他國家的開放資料都是有立法，人家都有請求基礎。而且「政府資訊公開法」原始精神是人民監督政府，它不是在處理Data Act，所以那不是足夠的法律基礎。我怎麼講都沒人聽，這是一個懶惰立法。

〈4〉現在的草案還沒想清楚要怎麼用資料來解決國家問題，只想到短期解決問題，合理合法使用都不解決，怎麼用大數據？

(4) 人工智慧科技基金會詹婷怡董事長：

〈1〉在政府政策裡，某些機制存在但不確定是否涵蓋，例如OpenData，此外資料治理也無人關注，很多議題都沒有人討論……。

〈2〉AI來了才發現我們資料治理基本功根本不夠。

(5) 台灣人工智慧學校蔡明順校務長：在訓練資源方面，政府跟美國和中國比，挹注也遠遠不夠，即使跟相近量體的國家比較，也是遠遠不夠。

(五)經查，以國科會TAIDE模型為例，來自公部門之訓練料集僅58個，不僅資料取得過程係由計畫團隊單打獨鬥分別向各個單位洽辦，本院並查證該58個資

料集只有其中2個⁸完整收錄於政府資料開放平臺，其餘不是未收錄，就是僅有目錄、連結、標題或內容不完整，無法作為訓練資料使用；政府對於如此重要的資料蒐整作業竟無系統性及通盤性的做法，可見政府訓練資料迄無健全之資料生態系。質言之，黃仁宇在《萬曆15年》對於明朝政府「全國滿佈著無數的短途運輸線，缺乏統一的組織和管理」、「……實際的執行卻全賴互不相屬的下級機構，……組織上的低能必然造成裝備上的落後」等尖銳批評，正是我國在政府資料治理方面的寫照：

- 1、以具代表性之TAIDE模型為例，據國科會查復：「目前國內對於政府繁中文本資料尚無有系統且具規模性之生態系」，在蒐整效率及完整性方面堪憂。
- 2、TAIDE訓練資料內容涵蓋政府及民間多元領域的資料，係經個別洽商授權；若以單純公部門訓練語料來源而言，公部門訓練資料集共45個，公部門出版期刊訓練資料集則為13個，合計58個。
- 3、政府開放資料……以量化、數字型之統計報告或清單為大宗，並非當前適合訓練TAIDE模型、進行「文本生成」所需資料，而部分合適的資料，TAIDE計畫團隊已下載或另向各部會洽詢取得。
- 4、圖文授權資料之取得較文字語料更為困難，故改採有償方式獲取圖資並與著作權所有者進行多輪洽談(計花費約4個月)，方完成小規模圖資採購。
- 5、在主權AI訓練語料之規模評估掌握、取得授權、

⁸ 財政部字辭典及臺南旅遊網景點資訊

蒐集作業、格式轉換、資料品管、清理、切Token 相關分工部分，以LLM而言尚未有明確處理流程可供依循。

- 6、鑑於數發部職司資料治理與開放之策略規劃、協調、推動及資源分配，推動主權AI所需之資料洽談與管理仍仰賴數發部，及掌握各專業領域資料的各部會共同推動，以達成訓練資料系統化之目標。

(六)再查，即使國內資料治理機制未臻完備，但數發部若能發揮「行政院及所屬各級機關政府資料開放作業原則」、「政府資料開放諮詢小組設置要點」及「行政院智慧國家推動小組(2021-2025)」所賦予之職責及機能，應可一定程度滿足主權AI之資料需求；惟查數發部卻怠於發揮政府資料開放平臺之機能，並稱「政府資料開放平臺的核心目標是推動各機關資料開放並集中列示於平臺，……提供資料供應與流通應用的基礎環境，……而非針對特定需求資料集進行處理或客製化收集」云云，且該部對於訓練資料規模、需求對接、運用回饋、處理程序及分工等等均無法具體答覆，不僅忽視訓練資料對於國族文化存續及數位經濟發展之迫切需求，亦怠於執行前開規定有關「極大化開放政府資料」、「建立績效管理及推廣機制」、「定期評估檢討執行成效」、「建立跨域合作溝通平臺」、「民間需求回應說明」及「跨部會法規疑義者提報行政院諮詢小組」等職責，怠失至臻明確。舉例而言，數發部雖稱於112年8月9日「TAIDE跨部會合作討論會議」後提供約1,692個資料集予TAIDE計畫團隊，但TAIDE後續並未採用，數發部卻對資料運用情形毫無所悉，

國科會亦無積極回饋，兩部會沒有確實對接，忽略在資料合作中，供需雙方需要反覆溝通需求之特性，以致於TAIDE計畫團隊仍舊以個別洽辦方式取得所需資料，即屬明證，益徵數發部在AI領域並未發揮政府資料開放之積極功能。

- 1、根據數發部查復⁹，101年各部會先由開放5筆資料集做起，迄今已開放超過5萬筆資料集、瀏覽量超過1億3,200萬人次、下載量達2,070萬人次。
- 2、嗣106年起協同中央各部會訂定領域資料標準，除制定「領域資料標準訂定流程參考指引」，提供各級政府訂定資料標準的參考方向，並輔導中央各部會完成24項領域、1,918個資料標準，集中列示於「政府資料標準平臺」提供各界查詢。
- 3、又稱，政府資料開放平臺的**核心目標是推動各機關資料開放並集中列示於平臺**，落實施政便民與公開透明，滿足民間對資料的「知」與「用」的權利，主要提供資料供應與流通應用的基礎環境，保障資料提供機關與使用者的權益，同時維持平臺的通用性，便利各界取得資料加值應用，而非**針對特定需求資料集進行處理或客製化收集**。
- 4、政府資料開放平臺的設計核心在於集中列示各機關已開放資料集，提供通用型資料資源，平臺本身不針對特定需求進行資料清理或客製化收集，亦不存放資料，資料實際由各機關管理並存放於其系統內，各機關自行管理其資料資源，確保資料更新的及時性，**平臺僅提供資料目錄與下載連結**，讓民眾能夠快速找到並存取所需的資料，提

⁹ 113年6月13日數位策略字第1130010943號函

升資料的公開透明度與再利用性，爰無直接提供資料集檔案大小具體數據。

5、由於文本資料的篩選與處理較為繁瑣，機關在開放資料時可能更傾向於釋出已結構化的資料類型，對於多模態或非結構化資料的開放意願較低。

(七)此外，數發部函稱：「若國科會對資料具體應用仍有進一步需求，本部將持續深化與國科會及相關機關的溝通」云云，亦有失政策高度。有鑑於臺灣特色訓練資料需求非國科會獨有，各機關、民間及學術單位均有需求，例如中研院曾有測試模型使用中國訓練資料集產生爭議，高雄市政府亦於113年12月11日宣示要做城市級主權AI等等，無不需要適當的訓練資料支援，數發部實應站在臺灣特色資料權威機構之高度，進行資料治理整體規劃，進而透過與其他AI或LLM模型溝通合作，設法在強勢文化所建構的大型語言模型或多模態AI模型中，保留臺灣特色的用語、圖像、社會、文化及價值觀等，否則行政院「五大信賴產業推動方案」宣示成為全球AI影響力中心之政策目標，將淪為空談。

(八)另查有關行政院之違失，除了行政院所核定之「臺灣AI行動計畫(2018-2021年)」早在108年就已指出資料開放及法規限制問題，又TAIDE於112年4月開始推動，復以中研院於112年10月發生測試模型使用中國訓練資料爭議，外界也早已指出繁中語料缺乏的隱憂；惟行政院諮詢小組面對外部需求極為迫切、政府內部對AI訓練資料極度欠缺掌握的情形下，任由資料開放平臺置身事外、部會各自為政，顯然有失推動、協調及督導之職責，亦有違失；此外，行政院「數位政策法制協調專案會議」雖下設有

「資料創新法制工作分組」，惟究其歷次會議對資料治理議題之督導強度，恐無法滿足AI發展之迫切需求及外界對資料治理長期沉痾之詬病；此外，「數位政策法制協調專案會議」下設之「資料創新法制工作分組」，與「行政院智慧國家推動小組」下設之「數位治理分組」，其分工似有部分重疊，爰行政院應予釐清或檢討，以確實發揮督導及協調之功能：

- 1、有關中研院於112年10月發生測試模型使用中國訓練資料爭議，當時中研院資訊科技創新研究中心研究員李育杰(現為國家安全會議諮詢委員)受訪¹⁰時即直指「要讓臺灣建立自己的大型語言模型，前提是需要建立臺灣自己的大型語言資料集，所以，『臺灣應該好好整理我們自己的繁體中文語料庫』」，復以目前AI模型推出及版本更新之速度動輒以季計算，比爾蓋茲早於1999年就已指出「在數位的高速公路上，不超速才是必死無疑」¹¹，顯示資料治理需求已超越「行政院及所屬各級機關政府資料開放作業原則」當初僅為達成施政便民及公開透明之範疇，而涉及國族文化及臺灣特色之保存，實屬刻不容緩之議題。
- 2、根據「政府資料開放諮詢小組設置要點」，諮詢小組依層級分為「行政院諮詢小組」及「中央二級機關諮詢小組」；其中有關主權AI所需之多元訓練資料尚非中央二級機關所能獨力推動完成，而涉及跨部會協調及督導事宜；惟「行政院諮詢

¹⁰ 余至浩。112年10月12日。【剖析中研院大型語言模型事件的衝擊】為何臺灣必須擁有本土化的LLM。iThome網站。(https://www.ithome.com.tw/news/159231)

¹¹ Bill Gates。1999/3/1。THE SPEED OF THOUGHT: Using a Digital Nervous System。

小組」面對外界迫切需要AI訓練資料的情形下，卻無視於平臺本身對於訓練資料掌握度偏低，仍堅稱資料開放平臺非為特定需求而設，任由政府資料開放平臺置身事外及部會矛盾叢生，顯然未能善盡前揭要點所賦予「擬訂資料開放推動政策」、「跨域合作溝通協調平臺」及「督導資料開放推動成效」等三項任務。

3、另據復，目前跨部會之數位政策協調權責在由行政院法制、經濟、科技三位政務委員組成之「數位政策法制協調專案會議」，該專案會議於112年4月24日第一次召開，茲將涉及本案之決議事項摘述於後；內容顯示專案會議對於資料治理之督導強度亟待強化，方能因應外部之競爭態勢。

(1) 112年4月24日：本專案會議應具有前瞻性，請各部會以部會層級，主動通盤檢視、盤點現在及未來與人工智慧(AI)有關之議題、政策、規範及研究。

(2) 112年8月8日：

〈1〉「資料之跨機關分享及二次利用」議題，請數發部於會後提供相關意見予國發會參考。

〈2〉資料為重要之數位戰略資源，應注意風險管控，保護隱私人權，及產業發展對經濟之衝擊。

(3) 113年9月30日：請數發部於下次AI法制分組會議說明「促進資料創新利用發展條例草案」、「AI 評估驗證工具」、「風險分類框架草案」推動進度。

4、承上，經查在「行政院智慧國家推動小組」下設之「數位治理分組」，係由數發部數位政府司擔

任召集單位，任務為「完備政府數位基礎、建構資料治理生態、規劃與研議提升智慧政府服務及數位治理所需之配套法制，促進國內數位治理生態結構之形成及落實等」；此與前開專案會議之AI法制分組似有功能重疊之虞，亦有請行政院進一步檢視釐清之必要。

(九)有關主權AI開發團隊是否得自行爬取政府機關網站資料作為訓練資料一節，經查，目前「政府網站資料開放宣告」¹²雖有宣告網站資料之授權情形，但普遍存在授權範圍及程度不明確、限制非商業用途或禁止再授權等情形，導致主權AI開發團隊仍需向各機關逐一確認；對此，數發部雖稱政府網站採取「政府網站資料開放宣告」CC0-1.0版本者，已宣告以無償、不主張著作權利之方式提供公眾使用，然經進一步查證，在186個機關網站中宣告為「CC0-1.0」者僅有4個網站，比率僅2%。數發部僅表示「將請各政府機關於網站頁面更清楚標示」，惟該措施恐無助於提高訓練資料獲取效率，仍有賴數發部檢討及研謀精進措施：

1、國科會查復¹³說明依數發部政府網站營運交流平臺¹⁴，對「政府網站資料開放宣告」訂有相關範本。據TAIDE計畫團隊表示，該宣告第2條「相關事項說明」項下第2、3項、第3條相關規定屬於通則性規範，網站頁面各項素材並未能逐一標示授權情形，爰TAIDE計畫團隊就模型所需之語料仍須向各單位逐一確認，以致影響語料取得的效

¹² 共有OGDL-Taiwan-1.0、CC BY 4.0、CC BY-SA 4.0、CC0-1.0等4個版本

¹³ 國科會113年12月4日履勘後補充資料(不備文)。

¹⁴ <https://www.webguide.nat.gov.tw/guidelines/442/show>

率。

- (1) 現今所有類型之LLM在取得資料及管理部分均無法完全符合相關規範與管理措施，因此勢必造成資料取得的困難。
- (2) 政府官網所提供之文章及影音，因多由外部廠商維運，其著作權可能不屬該機關，且資料公開並非等同於資料授權。以國科會為例，研究報告雖於國科會官網公開，惟著作權仍屬研究者。而研究計畫產出成果中，若包含國內外專利權、商標權、營業秘密、積體電路布局權、著作權或其他智慧財產權及成果，則歸屬各計畫執行機構所有。又以各政府觀光文獻資料為例，雖置於官網，但著作權仍屬相關撰寫者，並不在政府開放資料授權中。
- (3) TAIDE計畫團隊實際與機關洽談過程中，曾遇政府機關網站之委託案，在合約規範內僅取得著作之公開展示權、宣傳使用權，該著作或包含部分網站說明文字、圖片僅能公開瀏覽，無法供重製或再授權的案例。

2、數發部查復說明表示：

- (1) 雖然各機關建置的政府網站資料（如文本及多模態資料）屬於公開資料，但其使用過程中仍可能因授權條款而受到限制，例如限制非商業用途或禁止再授權。然而，依據「行政院及所屬各級機關政府資料開放作業原則」，政府資料開放係以政府資訊公開法為基礎，採取開放格式提供，並以無償、無地域及時間限制、不可撤回的方式授權利用，惟部分資料可能涉及著作權或第三方授權之複雜性，使機關在資料

篩選與釋出時需耗費大量時間與精力進行授權確認，進一步限制資料的範圍與應用潛力。

- (2) 政府網站資料開放宣告，係為利各界廣用網站資料，由機關對其網站刊載資訊受著作權保護之範圍，宣告其開放利用方式。數發部「政府網站服務管理規範」提供4個「政府網站資料開放宣告(OGDL-Taiwan-1.0、CC BY 4.0、CC BY-SA 4.0、CC0-1.0)版本」，供各機關依其網站屬性，挑選合適之版本後，於網站上進行資料開放宣告。現行CC0-1.0的版本已有宣告以無償、不主張著作權利之方式提供公眾使用，後續蒐集訓練資料時，可優先朝採用該版本宣告之政府網站進行資料蒐集。
- (3) 經該部於約詢後補充資料盤點總統府及五院所屬二、三級機關官網之政府網站資料開放宣告類型，共186個機關網站。其中宣告為「CC0-1.0」者，計數發部、數發部數位產業署、數發部資通安全署及雲林縣政府等4個網站，比率約2%。
- (4) 有關數發部於約詢前查復資料表示：「將請各政府機關於網站頁面更清楚標示」，經約詢後補充資料其做法為「規劃於資訊主管聯席會或是相關網站教育訓練課程進行宣導，提醒各政府機關須於特定網頁頁面上，清楚標示例外使用限制，避免誤用」等語。

(十)至於數發部及國科會說明後續將透過擬定AI Ready Data¹⁵、「促進資料創新利用發展條例(草案)」及

¹⁵根據113年8月26日國家發展委員會資料開放諮詢小組113年度第1次會議簡報，AI Ready DATA在確保資料的品質、資料結構化的程度、資料呈現樣態的多元性及可被搜尋查找的能見

「AI基本法」予以補強，則因尚在研擬階段，其成效有待後續驗證：

- 1、數發部查復¹⁶說明，目前國際尚無AI Ready Data標準，該部刻正觀測現有國際AI Ready Data評估原則，後續將與相關學者、專家及社群意見交流，評估試辦精選重點資料集，嘗試以國際AI預備資料評估原則，輔導部會準備AI模型訓練用之資料。
- 2、數發部林宜敬次長於113年12月27日本院約詢時說明：「數發部正在擬一個框架草案(按即「促進資料創新利用發展條例草案」)，將會鬆綁提供給大眾，不收費也不違法，公務員可以免責，此外會設立沙盒供研究創新使用」等語；惟林宜敬次長也坦承：「該條例主要是促進發展利用，所以沒有罰則」；換言之，該條例似乏有效工具促使公部門釋出資料，仍須仰賴行政院及數發部在行政上善盡督導、協調及協助之責。
- 3、國科會陳炳宇副主委則稱：「國科會負責『AI基本法』，已送行政院審議，希望透過本法讓各機關推動作用法的制定」等語。

(十一)另查審計部112年5月25日台審部一字第1120058546號函查核意見亦稱：「……TAIDE以提供基礎模型為目標，再供國內政府部門及民間使用，惟各界對政府開發國內版本生成式AI，多有科研經費及算力資源不足、營運模式未明及資料量與國外大型語言模型差距甚大等訾議」等語。

(十二)至於本案另有陳訴人陳訴中央氣象署及空軍氣象聯隊使用中國「盤谷天氣預報模式」，將引發重大

度，以隨時滿足AI應用的需求。

¹⁶ 113年6月13日數位策略字第1130010943號函及113年11月27日簡報

軍事、資訊、國家安全問題等情，經本院函詢中央氣象署及國防部說明如下，顯示「盤谷天氣預報模式」係屬開源模型，中央氣象署及國防部均未導入常規作業、對外連線及介接內部資料，爰尚無充分證據遽認兩機關有明確違失：

- 1、氣象署就其預報結果進行誤差特性分析評估，並與傳統數值天氣預報模式相比較，以了解AI天氣預報模式之預報能力特性及限制。
 - 2、該署並未且無須連線華為雲或盤谷大模型。該署係參用該開源軟體建置於內部網路進行研究。
 - 3、經過各國先進氣象作業及研發單位使用該模型並察考後，迄今並未發現任何資安疑慮警訊之消息
- (十三)末查，據國科函復資料顯示，TAIDE計畫團隊於洽辦訓練資料取得時，尚遭遇資料收費問題。例如，屬於公部門之中央社資料仍向計畫團隊索取費用，而私部門(ETtoday-新聞雲)反而願意無償提供資料，其中顯有值得檢討改善之處。數發部林宜敬次長於本院詢問時表示：「本部刻正規劃成立媒體的訓練資料庫，鼓勵民間媒體有版權的捐到資料庫中，這樣大家都可以沒有疑慮的拿去訓練。若權利人一定要收費，那就先不放資料庫」等語，或能部分紓解問題，惟仍待後續檢視其成效：
- (十四)綜上，AI及LLM等革命性技術及其運用潛力，除一般咸認將普遍對就業、社會、文化、人權等領域造成衝擊外，由本案所揭露之問題顯示，AI也成為當前資料治理之法制及行政機關整備情形的試金石，凸顯並迫使政府面對長期以來怠於積極建立資料生態系之違失；鑒於臺灣特色的訓練資料攸關國族文化及價值觀的延續，更與經濟發展息息相關，

行政院必須督同數發部、國科會及其他相關機關本於職責積極解決AI所帶來的資料治理挑戰。

二、在AI發展所遭遇的挑戰中，著作權亦屬極為關鍵的議題；經查我國在AI訓練資料之合理使用規範，目前僅有經濟部智財局一項函釋可參，在法律保留及明確性方面，顯不足以因應AI高速發展所衍生的各種著作權問題，現況確有檢討之必要；而經濟部智財局所稱智財領域為高度國際調和事項，需持續觀察國內外之司法實務發展等語，固非無憑；惟考量AI發展刻不容緩，過度保守之法制環境未必有利於衡平「著作權法」的三個立法目的，爰行政院「數位政策法制協調專案會議」亦宜督同經濟部智財局積極辦理法規調適，以衡平AI發展需求及各方權益。

(一)由相關法令及函釋內容顯示，目前訓練資料是否符合「合理使用」除必須就個案判斷之外，智財局並認為尚須持續觀察國內外之司法實務發展，無法一概而論。因此，現況恐使訓練資料蒐集者既無法逐一確認授權，也難以預見蒐集行為是否構成侵權；司法體系也將承擔法令的不完備而投入寶貴的司法資源，在實務上也將面臨缺乏充分的法律依據以形成穩定的見解，終將演變成行政和司法體系間的灰色地帶。總之，現況已不足以因應AI高速發展所衍生的各種著作權問題而有檢討之必要，茲將AI涉及之著作權相關法律及函釋臚列如下：

1、「著作權法」第1條規定：「為保障著作人著作權益，調和社會公共利益，促進國家文化發展，特制定本法」。

2、「著作權法」第65條第1項及第2項規定：

(1)第65條第1項規定：「著作之合理使用，不構成

著作財產權之侵害」。

- (2) 第65條第2項規定：「著作之利用是否合於第44條至第63條所定之合理範圍或其他合理使用之情形，應審酌一切情狀，尤應注意下列事項，以為判斷之基準」。

〈1〉利用之目的及性質，包括係為商業目的或非營利教育目的。

〈2〉著作之性質。

〈3〉所利用之質量及其在整個著作所占之比例。

〈4〉利用結果對著作潛在市場與現在價值之影響。

3、智財局112年6月16日針對AI與智慧財產權疑問做成經授智字第11252800520號函釋，摘述如下：

- (1) 在蒐集資料訓練AI模型階段，訓練資料如受著作權法保護（下稱原始著作），會涉及「重製」原始著作之行為，除有著作權法第44條至第65條合理使用之情形外，應取得著作財產權人之同意或授權，始得為之。

- (2) 如未取得同意或授權，原始著作之著作權人（以所詢之「原作者」稱之）得依現行著作權法向司法機關舉證維權，……據了解我國尚未有類似訴訟案件，可持續觀察國內外之司法實務發展。

- (3) 由於著作權係屬私權，個案行為是否構成著作權侵害，仍須由司法機關就個案具體事實調查證據認定之。

4、對此，國科會查復說明，「TAIDE計畫」啟動之初，計畫團隊即洽詢經濟部智財局表示意見；而智財局表示，需在保護著作權人利益與AI產業發展間取得平衡，須審慎因應，未來仍須觀察各國著作

權議題發展，爰TAIDE模型使用之繁體中文訓練素材，由計畫團隊逐一向個別單位之原權利所有者之授權及再授權，以兼顧模型發展，並減少風險與爭議。

(二)依據國科會及智財局盤點其他先進國家如歐盟、日本等國在生成式AI訓練資料於著作權方面之發展如下，其中日本限制確實較為寬鬆，然而觀察其他國家制度，智財局所稱：「智財領域為高度國際調和事項，需持續觀察國內外之司法實務發展」等語，似非無憑。

1、各先進及亞鄰國家制度摘述如下：

- (1) 歐盟、德國：現行規定雖不限於商業目的，但允許權利人拒絕提供資料，生成式AI能否適用仍有爭論。
- (2) 日本：文化廳初步肯認生成式AI訓練階段利用他人著作之情形，有適用日本著作權法第30條之4「資料探勘」之著作權限制及例外規定之空間。惟日本插畫、漫畫、攝影、圖書等內容及出版業協會，向政府訴求主張AI生成圖像涉及侵害其著作權，要求更明確解釋資料探勘合理使用規定。
- (3) 新加坡：新加坡著作權法第244條，對資料探勘著作權限制有例外規定，允許「重製」的行為，重製後所得的重製物可以儲存或保留。惟限於利用人「有合法接觸權限的著作」，且權利人得以契約(使用者)條款排除該條合理使用規定之適用。
- (4) 英國：暫緩資料探勘例外規定放寬適用範圍之修法(現行資料探勘合理使用規定，須非商業

使用之目的)。

(5) 美國：

〈1〉國科會稱，美國著作權法第107條訂有合理使用(Fair Use)概括合理使用規定，該國國內已有多起以此為爭點提起生成式AI相關訴訟。另在著作權法保護部分，美國著作權局於西元2023年3月發布「包含AI生成內容之著作權登記指南」，指出在「沒有人類參與」的情況下自主生成作品，或透過人類下提示詞所生成之作品，不受著作權保護；利用AI輔助創作之情形，著作權註冊申請人應在申請書中揭露作品是否包含AI生成之內容，並說明人類對該作品所投入的創意貢獻之部分，就該部分可主張著作權。

〈2〉智財局則稱，未來最重要之參考指標美國，亦仍在制定AI智慧財產權政策和法規的初期階段，美國著作權局於2023年就AI相關著作權議題徵求公眾意見，評估是否須針對AI所引起的著作權法及政策問題採取立法或監管措施，尚待2025年始會提出評估。

(6) 至於韓國、香港，目前並無資料探勘合理使用規定，韓國在2021年係國會議員提案，但僅止於提案階段而無後續進展，且針對如何因應生成式AI仍在觀察，尚未實際採取立法行動。

(三)惟查，綜整本院諮詢中研院法律所資訊法中心劉靜怡合聘研究員之意見及智財局自辦諮詢會議中專家所提意見顯示，為兼顧著作權及AI發展，法制工具的強化有其必要；而我國既然要「成為全球AI影

響力中心」¹⁷，政府就不應停滯在觀望國際的階段。況且，若回歸「著作權法」第1條所揭櫫的三個立法目的，當可發現在強勢文化所建構的AI革命性技術浪潮下，若政府僅以保障著作權益為考量，實有損害社會公共利益及國家文化發展之虞，爰智財局允宜積極透過立法程序與社會對話，始能在AI技術衝擊下，衡平「著作權法」的三個立法目的。

- 1、中研院法律所資訊法中心劉靜怡合聘研究員：政府說美國還在打官司，歐盟也不知道，要等到有結果。人家是基礎建設都正在使用，才有這些問題浮現，結果政府說要等有結果？那怎麼來得及。
- 2、達文西個資暨高科技法律事務所葉奇鑫律師：
 - (1) AI需要用到的資料非常龐大，……要向全世界這麼多的著作取得授權，是無法操作的事。
 - (2) 立法茲事體大，等立完法後，產業又輸人家更多。
- 3、北辰著作權事務所幸秋妙律師：……如果是全新的生成、學習資料的重製是短暫的，應該可以主張第65條的合理使用，不然AI產業無法發展。
- 4、東吳大學法律學系章忠信助理教授：
 - (1) ……可以考慮用法定授權，不必經過同意，可以利用，但使用報酬要拿出分配……。
 - (2) 智財局作為著作權專責機關，同時面對利用人及著作權人，不宜輕易透過行政解釋將利用他人著作訓練生成式AI的行為，直接認定是合理使用……。
 - (3) ……透過資料探勘來解決訓練階段的問題，已

¹⁷ 賴總統113年5月20日就職演說。

經是難以阻擋的趨勢。所以，我國真的要考慮立法。

5、益思科技法律事務所賴文智律師：

- (1) 如果單純只就訓練，智財局可以放寬一點，立法是一種民主的程序，因為AI的時代是**不能避免的，不應該造成業者在做訓練時都不敢做。**
- (2) ……臺灣如果不開放一點，讓臺灣自己的政府機關或業者可以多訓練一點繁體中文的資料，事實上影響力就會愈來愈小。
- (3) 到時候在人工智慧的影響下，可能會變成臺灣的東西要送給人家訓練，人家都不想拿去訓練，這個時候是連什麼利益都分配不到，結果是一樣的。
- (4) 基於文化的考量，希望智財局對於此議題不要太保守，其實不是只有產業面的問題，還有整個繁體中文(正體中文)資料的能見度問題。

6、東海大學法律學院胡心蘭教授：……**如果今天要扶植AI產業，需要一個明確的規定，第一個好處是可減少訴訟，第二個好處是可以解決前述「轉化性」的不確定性問題，如都以個案討論的話，是非常不經濟的。**

7、翰廷法律事務所黃秀蘭律師：

- (1) ……如果還是依過去智財局解釋令函，個案須由司法機關認定，處理的時間會拖更長。
- (2) ……可以用立法訂定法定授權、強制授權或報酬請求權的方式解決，因為AI趨勢既然不可擋，我們的立場就是大家都可以用，用了之後要補償權利人。

8、北辰著作權事務所蕭雄淋律師：

(1) ……應以立法解決，因為如果只用行政解釋，首先是智財局恐難輕易作出解釋，再者是行政解釋無法如同立法一般詳細明確，交給司法機關認定導致個案適用之不確定性，變成司法及行政機關都不負責的情形。

(2) 我國既以發展AI產業為目標，相關問題應設法解決，建議參考我國著作權法一直以來主要沿襲參考之日本立法例，不僅針對AI，連同網路與高科技時代的所有著作權問題都一起解決。

(四)另查，行政院113年9月5日第3919次院會簡報，主權AI在「相關法規調適推動」之部會分工，已指定由經濟部進行「著作權法」之法規調適作業。此外，行政院「數位政策法制協調專案會議」於113年9月30日會議亦決議：「請經濟部就AI所涉著作權議題，持續觀察國際規範趨勢，並適時採取相應措施，以利我國AI之創新應用。至於應用智慧財產權之合理使用及回饋予創作者之機制，請經濟部協調相關部會共同討論是否有兼顧二者之可行作法」。爰「著作權法」之法規調適作業由行政院督同經濟部辦理，應無疑義。

(五)綜上，AI的革命性技術對於我國著作權之法制環境已形成新的挑戰，現況已不足以因應，政府不應坐以待斃，爰行政院允宜督同「著作權法」主管機關經濟部積極研謀法規調適，以衡平「著作權法」所揭櫫的保障著作人著作權益、調和社會公共利益及促進國家文化發展等三個立法目的。

三、主權AI及LLM之推廣應用成效乃推動產品生命週期及永續發展之重要動能，行政院推動「五大信賴產業推動方案」期待百工百業導入AI帶動全面升級，惟相較

於商業資服模式，主權AI在任務及市場雖有一定區隔，但在整合服務、導入成本、推廣強度及模型能力等仍有精進空間，國科會及數發部允宜密切合作並持續支持計畫推動，以避免主權AI因使用率低落而遭到邊緣化。

- (一)依據國家發展委員會發表「運用AI賦能 擴展產業發展」¹⁸闡述「五大信賴產業」在促進AI智慧應用之推動策略包括推動百工百業導入AI帶動政府及產業升級之目標，已顯示未來無論通用型或專用型AI模型的應用潛力及市場將十分可觀。
- (二)根據國科會函復顯示，目前TAIDE模型作為主權AI的示範性計畫及象徵，已有初步的推廣應用，其模型能力亦與同期主流模型概同，在臺灣文化及用語方面則有明顯優勢；惟其全職工程師已由6名縮減為2名，人力資源不增反減，是否能持續維持開發動能而保持競爭力，不無疑問；學者專家亦已提出警語，認為「AI發展是一場持久戰，不能一年有，一年沒有，否則就沒有了」、「TAIDE發展有些緩慢有必要加速」等語，非常值得主管機關重視並採取因應作為。

1、國科會函復重點如下：

- (1)有許多業者以TAIDE為基礎發展應用服務的具體案例，例如：群聯aiDAPTV+平臺、高雄大學Kuwa GenAI OS、叡揚資訊整合TAIDE模型、中興大學農業知識檢索系統「神農 TAIDE」、陽明交大臺客語TAIDE大語言模型結合語音AI介面、教育部AI CUP競賽報告檢查系統、臺南大

¹⁸ 國家發展委員會2024台灣經濟論衡9月號
(https://www.ndc.gov.tw/Content_List.aspx?n=E44AF0AAC3907F4C)

學臺語對話生成式AI機器人等，對TAIDE作為主權AI應用的擴散有所助益。

- (2) 根據該會113年11月27日簡報，TAIDE模型主要著眼於中翻英、英翻中、寫文章、寫信、摘要等五大任務，TAIDE 8B模型與Meta最新的Llama3.1 70B模型比較，兩者表現差距不大，顯見TAIDE模型在5大任務之能力已達預設目標，在準確性方面分數則遠高於同等級 Llama3 8B分數，顯見TAIDE仍具有臺灣文化及用語方面的優勢。
- (3) 在人力資源部分，TAIDE計畫第1階段計畫團隊主要仰賴國內研究LLM學者，帶領其研究室團隊，加入模型訓練工作。第1階段至計畫執行第6個月，已招募6名全職工程師，其餘兼任人力計32名；第2階段計畫則僅有2名全職工程師，兼任研究人力計22名。
- (4) 資料蒐集及處理需專業人力進行資料清理、標註，同時需具備瞭解知識架構、兼具資料管理、模型訓練專長之人才，並有長期的團隊合作。
- (5) 資料授權需專責人力進行外部洽商，並與資料擁有者、模型訓練與清理三方溝通，經驗與知識累積不易，徵才困難。
- (6) 日、星兩國政府各自支持開發之Fugaku-LLM、SEA-LION模型，與以英語為主的跨國企業LLM服務相較，更理解當地的文化脈絡、使用需求。因該等模型均以開源釋出，業者因而可據以推出當地市場所需的應用服務，例如：新加坡國家電腦系統公司(NCS)應用SEA-LION模型，將其法律合規流程中使用的內容翻譯成其他東

南亞語言；又例如日本富士通以Fugaku-LLM為基礎，結合自身Fujitsu Kozuchi(AI平臺)提供LLM相關應用服務。

2、次查，台灣人工智慧學校秘書長侯宜秀於114年2月5日演說¹⁹指出：

- (1) 主權AI的技術發展是一場長期且持續的戰鬥，臺灣AI已落後太多，需急起直追發展臺灣主權AI。
- (2) AI發展是一場持久戰，不能一年有，一年沒有，否則就沒有了，AI對話引擎3個月就有一個新的版本，沒有完美模型，國內應對研究人員端出臺灣所屬的TAIDE模型成果多所包容。要給空間去回饋研究人員。
- (3) 臺灣沒有選擇的空間，一定要往前加速發展AI，目前國科會主導的TAIDE發展有些緩慢，有必要加速。

(三)再查，目前市場指標性之AI商業模式除以強大的模型能力為核心以外，同時也以解決用戶痛點為導向，與作業系統、資料庫、雲端服務、資通安全、著作權疑慮及套裝軟體等深度整合，用戶亦無需建維硬體環境，形成開箱即用的資服模式，對於缺乏自行維運及開發能力的公私部門極具吸引力；再者，對於資服產業稍有涉獵者均知，組織一旦導入資服形成慣性後，轉換系統之成本極鉅，故一般均不會輕易轉換資服方案；因此，國科會對於主權AI的邊緣化風險雖稱：「TAIDE模型以發展5大任務為發展重點，並非以取代業界LLM或與業界LLM競爭市場為目

¹⁹ 網址：<https://www.ctee.com.tw/news/20250205701828-430104>

標」云云，但外部現況對TAIDE模型的推廣仍是真實存在的壓力，過去許多政府APP因低使用率而遭人詬病更是前車之鑑，爰行政院允宜督同國科會、數發部及相關部會積極布建主權AI的有利條件，檢討TAIDE是否仍以計畫形式推動，同時降低導入門檻並持續挹注資源。

- 1、據悉目前市場指標性之AI資服方案，國內至少已有教育部、永豐銀行、光寶科技、中華開發金控、遠傳電信、聯發科、台積電tGenie、研華科技、奇美醫院、中國醫學大學附屬醫院、三軍總醫院、中山醫學院附屬醫院、家樂福、民視、KPMG安侯建業聯合會計師事務所等公私部門導入，顯見毋須建維硬體環境、開箱即用以及適當著作權保障承諾等有利大幅降低初始導入門檻。
- 2、國科會表示，跨國企業的LLM服務對TAIDE可能有影響，但不至於到「威脅」程度，主要係因TAIDE模型繁中比例高，且採用「高品質」的在地繁中語料進行訓練，故TAIDE較其他跨國企業LLM服務「更懂臺灣」。在五大任務功能(中翻英、英翻中、寫文章、寫信、摘要)上，亦有助國內使用者提升生產力，故TAIDE仍具優勢。
- 3、數發部則稱，113年已提供60家資服業者與新創業者使用算力平臺，發展多元應用如多國語言直譯器(如台語直譯印尼話)、提升防洪預警能力(預測淹水模型)、環保科技執法輔助工具(辨識空汙模型)等，其中共有22家業者選用TAIDE模型進行微調訓練。

(四)綜上，公私部門及社會各層面普遍導入AI將是難以逆轉之趨勢，主權AI若欲維持開發動能並永續發展，

達成主權AI保留臺灣特色的用語、圖像、社會、文化及價值觀的目的，就必須展現足夠的產品價值並降低導入門檻；惟查，在商業資服模式積極拓展市場及各國無不傾力推動主權AI發展的情形下，TAIDE計畫之人力竟不增反減，難稱合理，爰行政院允宜督同國科會、數發部及相關部會積極布建主權AI的有利條件，檢討TAIDE是否仍以計畫形式推動，同時降低導入門檻並持續挹注資源。

四、算力及電力屬於AI發展的重要配套，截至113年11月，我國算力達103.540 PFlops，在世界排行第11名，仍須持續積極建置並爭取國際合作，並同時設法節約算力；而與算力息息相關之電力配套，經濟部預估117年AI科技之用電需求約增加200萬瓩，較112年成長約8倍，惟以目前電網特性及輸配規劃等，仍有影響區域發展及廠商意願之虞，尚有進一步檢討之空間；爰此，有關算力及電力等重要配套，尚待行政院督同相關部會積極規劃辦理。

(一)根據國科會統計，截至2024年11月，我國算力為103.540 PFlops²⁰，在世界排行第11名，略為落後韓國，而在人均算力則為0.0045TFlops²¹，略為落後日本及新加坡；國科會雖然表示我國正急起直追並期於118年達成480PFlops之算力；惟若單靠政府資源建置算力，短期內恐難以滿足公私部門百工百業全面導入AI之需求，政府仍宜積極透過與私部門協力，並思考導入算力節約技術，使算力成長跟上需求擴張，而該等工作至少涉及經濟部、國科會及數發部等機關權責，仍有賴行政院善盡跨部會督導協

²⁰ Peta Floating-point operations per second，每秒千萬億次浮點運算次數。

²¹ Tera Floating-point operations per second，每秒兆次浮點運算次數。

調之職責。

- 1、依全球超級電腦統計專區113年11月最新統計排名如下表1：

表1 截至2024年11月之各國算力排名(國科會國網中心提供)

算力 排名	GDP 排名	人口數 排名	國家	算 力 (PFlops)	計 算 核 心 數 (Cores)
1	1	2	美國	6,475.558	56,801,360
2	4	12	日本	940.710	11,736,944
10	13	28	韓國	213.090	2,438,772
11	22	57	臺灣	103.540	712,336
27	32	115	新加坡	23.237	280,832

- 2、另據國科會113年11月27日簡報整理，我國在亞鄰國家之人均算力排名如下表2：

表2 我國在亞鄰國家之人均算力排名(國科會國網中心提供)

國家	人口數(萬人) (註)	算力(PFlops)	人 均 算 力 (TFlops)
日本	12,494	940.710	0.0075
新加坡	407	23.237	0.0057
臺灣	2,323	103.540	0.0045
韓國	5,163	213.090	0.0041

註：依據2022年OECD數據

- 3、據國科會表示，我國正急起直追建置算力，將藉由晶創與大南方計畫建置超級電腦，其中晶創主機建置時程為113至117年，將取代目前台灣杉二號及三號主機，目標為280PFlops，而大南方主機建置時程則為115至118年，目標為200PFlops。換言之，我國至118年可望建置480PFlops之算力，預計可大幅提升排名。
- 4、另據行政院113年9月5日第3919次院會簡報，在「鞏固主權AI基盤」項下，已將「爭取國際合作等擴大算力」納為經濟部之任務；嗣依國科會113

年11月27日簡報以經濟部與輝達合作之Taipei-1超級電腦為代表案例，其算力達22.3PF，為全球排名第38位，並已完成兩次算力徵件專案。

5、惟查，學者專家仍認為，在各領域都要導入AI之情形下，算力仍是我國發展AI必須精進之重要議題，若由開源節流之概念視之，除經濟部推動類似輝達Taipei-1主機建置合作，數發部建置算力池等措施外，如DeepSeek模型之節約算力技術等，均應持續檢討精進。

- (1) 中研院資訊科技創新研究中心研究員李育杰²²：雖然要建立基本算力，但不該一次投入所有資金。一半用於購買設備，一半用於雲端服務，才能跟上最新的技術。我們資源有限，投資時需要有明確的策略，……積累了數據和人才，而不是盲目追求超大規模的算力。
- (2) 台灣人工智慧學校秘書長侯宜秀：……臺灣甚至各國將DeepSeek視為洪水猛獸，畢竟它有資安上疑慮，但它用腦力補足算力，提供一條新的路徑，尤其開源的技術方法，我國的研發人員有必要研究且學習²³。
- (3) 工業技術研究院產業科技國際策略發展研究所研究總監楊瑞臨：臺灣AI算力目前就是「供不應求」，不管是新藥探勘、新材料開發，都需要算力，國內製造業、學校、醫院等需求都很大。²⁴

(二)次查，經濟部在「112年度全國電力資源供需報告」

²² <https://futurecity.cw.com.tw/article/3548>

²³ 呂雪慧 2025年2月5日。TAIDE發展緩慢 侯宜秀：台灣主權AI落後太多需急起直追。工商時報

²⁴ 柳名耕、王玉樹。2024年7月7日。台算力供不應求 長期仍需自建，中時新聞網。

指出，113至122年全國電力需求年均成長率約為2.8%，其中，AI科技之用電需求，預估至117年約增加200萬瓩，較112年成長約8倍(按換算結果：112年需求約為24萬瓩，117年需求約成長至228萬瓩)，並稱已樂觀估計；惟按其他專家學者或機構對AI用電需求之評估，雖非針對臺灣估計，但多有超過經濟部所預估者，例如美國能源部估計2023年到2028年間成長率將介於13%到27%之間，而國際能源總署(IEA)則預估至2026年，AI產業電力需量至少是2023年的10倍等，爰經濟部對於未來AI發展之電力需求，似可進一步就研究方法加以研析。

- 1、根據經濟部113年7月15日公布「112年度全國電力資源供需報告」關於未來國內算力成長所需電力部分指出，考量經濟情勢、未來AI科技潮帶動之半導體產業擴廠及電動車推動政策等因素，預估113至122年全國電力需求年均成長率約為2.8%，其中，AI科技之用電需求預估至117年約增加200萬瓩，較112年成長約8倍(按換算結果：112年需求約為24萬瓩，117年需求約成長至228萬瓩)。
- 2、嗣據經濟部能源署沈慧佳科長於本院詢問時說明略以：「能源署執掌全國電力供需評估，是與技術司、產發署及台電橫向聯繫協作，只要有跟台電申請我們就會知道。但我們較難掌握未來AI應用的滲透率，或是從數據資料庫轉換成算力的情形。所以『112年度全國電力資源供需報告』預估的200萬瓩是兩台大型燃氣機組的發電量，已經寬估了」等語。
- 3、另據經濟部能源署補充資料說明，其AI用電分析方式主要有「以申請用電量評估」及「以設備需

求評估」兩種，其中「以設備需求評估」係依據經濟部統計處及國際智庫對國際伺服器成長量評估臺灣內銷量，年成長率約25%來計算。能源署考量業者潛在需求恐有重複計算之疑慮，故以用電需求較高者為採用基準，至2027年以設備需求評估之用電量較高，共200萬瓩，後續將視情形滾動檢討。

- 4、另查，根據台灣能源技術服務產業發展協會張世平監事於113年10月17日根據國際能源總署（IEA）-電力2024(IEA-Electricity 2024)所發表之「AI浪潮下，資料中心(DC)之全球電力需求趨勢²⁵」一文對AI產業電力需求評估指出，AI用電需求，可以AI伺服器未來預估銷售數量及其額定功率來預測。這AI伺服器市場目前由科技公司輝達(NVIDIA)主導，約95%的市場占有率，2023年輝達出貨量為10萬台，每年平均用電7.3TWh，至2026年，AI產業將呈指數式成長(Exponentially)，至少是2023年電力需量的10倍。
- 5、另查，美國能源部2024年12月20日公布之委託研究報告「DOE Releases New Report Evaluating Increase in Electricity Demand from Data Centers」針對美國之評估則顯示，美國之AI用電成長率明顯較能源署針對臺灣所估計情形為高。
 - (1) 從2014年到2018年，美國資料中心的電力使用量複合年均成長率約為7%。

²⁵ 張世平。113年10月27日。AI浪潮下，資料中心(DC)之全球電力需求趨勢。台灣能源技術服務產業發展協會專題報導，第77期(<https://www.taesco.org.tw/MeetingNews/Detail/183>)

(2) 2018年到2023年間，成長率加速至18%。

(3)預計2023年到2028年間，成長率將介於13%到27%之間。

(三)再查，台灣電力股份有限公司曾文生董事長曾於113年8月9日對外表示²⁶：「為減緩輸電瓶頸，桃園以北不核供大型AI資料中心」等語，經經濟部查復表示：「主要是區域供需的考量，不希望無人駐守且用電量大的資料中心一直往北部設置，造成北部用電需求大於供給的狀況持續擴大」等語，且能源署沈慧佶科長於本院詢問時補充：「輸配電工程需要時間，其中的時間成本就需要廠商抉擇」等語，已凸顯電網南北供需失衡之長期問題，況據悉部分廠商選址係考量AI用戶群落特性及需求，恐有囿於輸配電工程進度而有影響設置意願之虞，爰經濟部仍應就電網結構及配置適時檢討。

(四)綜上，充足之算力及電力為AI發展的重要配套，針對現況算力需求，政府在資源有限之情形下，除持續建置算力外，亦宜循公私協力持續積極建置，並思考導入算力節約技術；而在電力部分，則可進一步研析用電需求估計方法，並持續檢討電網結構長期失衡或國家能源政策，以充分支援AI及數位經濟產業。

²⁶ <https://www.business today. com. tw/article/category/183027/post/202408100003/>

參、處理辦法：

- 一、調查意見一至四，函請行政院督同數位發展部、國家科學及技術委員會及經濟部確實檢討改進見復。
- 二、調查意見一至四，函請審計部參處並函復陳訴人。
- 三、檢附派查函及相關附件，送請交通及採購委員會、教育及文化委員會、財政及經濟委員會聯席會議處理。

調查委員：賴鼎銘

林郁容

葉宜津