

【公告】經濟部與 NVIDIA 「Taipei-1」 第2次徵案申請須知

經濟部「領航企業研發深耕計畫」(簡稱大 A+計畫)，吸引國際級大廠在臺從事前瞻創新研發活動，NVIDIA 來台設立「人工智慧創新研發中心」，建置超級電腦「Taipei-1」(系統說明詳附件1)，自113年7月起至116年2月止，免費分享算力給我國各界研發使用，歡迎我國廠商、學校、研究機構及新創團隊提出申請。

一、徵案說明

- 本次徵案期間:自公告日起至113年8月26日(一)下午5時整截止收件，採線上申請(含檔案上傳；申請書格式詳附件2)。
- 算力使用期間:113年10月01日至114年4月02日止，預計推薦8隊(DGX 4隊、OVX 4隊)
- 共分四批:每批涵蓋 DGX 與 OVX 各1隊，每隊依序使用6周算力。

二、申請資格

- 可單獨一家公司/單位，或多家企業/單位聯合申請。
- 聯合申請需由一家企業/單位主導，並擔任專案主持人。聯合申請之**主導單位**必須具法令地位法人(legal entity)，負責規劃與協調團隊成員的算力資源配置與使用，並綜整聯合提案之結案報告，請各團隊成員予以配合。
- 基於 NVIDIA 算力管理規定，申請單位必須完整使用16台 DGX 或8台 OVX，且一次須完整使用6周，為充分運用算力資源，如算力需求未滿16台 DGX 或8台 OVX 之申請單位，請組隊進行聯合申請。
 - 申請者可自行組隊，或洽目前已有組隊意願之單位合作，如台灣人工智慧晶片聯盟(AITA)，聯繫窗口：aita@itri.org.tw/03-5912777。
 - 如為學界申請，可透過國科會前瞻處相關計畫窗口進行組隊等事宜。

三、審查重點

採批次收件、批次審查，由經濟部產業技術司籌組專家委員會，針對提案申請書進行審議，分就 DGX 與 OVX 進行審查及核定。審查重點如下：

- 專案主題具前瞻性與商業化潛力
- 提案單位須具備大算力執行經驗及訓練資料
- 算力需求合理性，且能充分運用 DGX 16台或 OVX 8台算力
- 預期產出與效益

四、整體作業時程

- 徵案期間:自公告日起至8月26日(一)下午5時整截止收件。
- 9月中旬:進行審議並擇優推薦,獲推薦單位將公告於 A+企業創新研發淬鍊計畫網站 (<https://service.moea.gov.tw/EE514/tw/aiip>)。
- 各批次之專案執行:依審查結果,按排序優先者依序選擇使用批次(擇一),每批次各有 DGX 與 OVX 各1隊。考量重要假期與系統升級維護等,預估各批次使用時間如下(主辦單位保有調整之權利),屆時以專案計畫窗口一周前之通知為準:
 - 第1批次 (113年10月01日至113年11月12日)
 - 第2批次 (113年11月13日至113年12月30日)
 - 第3批次 (114年01月03日至114年02月18日)
 - 第4批次 (114年02月19日至114年04月02日)
- 專案結束:執行結束後兩周內提交結案報告(採線上提報,格式詳附件3),並配合進行成果揭露(輔以影片或展示),讓各界了解該專案提升研發能量並增加競爭力之效益。

五、申請主題

- 專案主題包括但不限於:生成式 AI 相關應用、大型語言模型開發、數位孿生、物理資訊神經網路、量子計算模擬、藥物發現、醫療保健、數位孿生、3D 動畫/影像渲染等,亦可多個主題進行申請。

六、申請方式

- 採線上申請(網址:
https://docs.google.com/forms/d/1hicrFB1ckn_tZRL6at98v_RqsOyaPgSmBlcyvBS0dA/viewform?edit_requested=true)。為避免系統壅塞影響申請權益,請儘早申請。

七、算力資源與使用說明

- **DGX 運算資源:**16台 DGX H100, 每台8張 H100 GPU。
- **OVX 運算資源:**8台 OVX 2.0, 每台8張 L40 GPU。
- **每專案算力使用時間共6周:**第1周 on-boarding 作業包含申請帳號、資料上傳、教育訓練等前置作業,第2至5周運算使用共計4周(含假日),第6周 off-boarding 如資料下載、解除帳號、刪除資料等。
- **聯合組隊算力分配注意事項:**由聯合申請之主導單位進行規劃與協調算力分配,並於算力啟用前一周,將分配方式及帳號提供 NVIDIA 進行事前作業,每個單位以2個帳號為原則。
- **訓練資料上傳方式:**使用者之訓練資料主要以網路傳輸方式上傳至平台;然資料量過於龐大時,可向 NVIDIA 提出需求並討論,以實體硬碟方式提供資料。
- **技術支援:**NVIDIA 會提供申請通過者聯繫窗口,安排技術人員協助使用

者解決平台使用的問題。服務時間為每周一至五的上午9時至下午6時。

八、權利義務

- 本算力為「非商業用途」：本算力以測試、評估和開發為目的，包括模型開發或解決方案的研究，以支持未來產品或服務的研發，相關開發之模型或研究成果，為使用者所擁有；禁止將「Taipei-1」運算資源用於產生營收（例如雲端服務、挖礦等）。
- 使用者不應在該環境中處理和輸入任何專有或敏感資訊。
- NVIDIA 要求每個算力使用企業/單位須與 NVIDIA 簽署保密協議(NDA)、使用協議(User Agreement)，相關文件為制式合約無法調整。獲通過之使用者，必須完成文件簽署，始得啟用算力，如因未完成簽署而造成算力使用時間短少，由使用者自行承擔。
- 申請書所提交之各項資料，請務必據實填寫，如 Token 數及記憶體等，以利進行審查及算力配置。
- 本計畫為政府資源，將在不涉及使用者之研發或營業機密情況下，適當進行成果揭露。
- 依據行政院及所屬機關(構)使用生成式 AI 參考指引，若有使用生成式 AI 執行業務或提供服務，應向該業務或服務之有關對象進行揭露，使其知曉。
- 本公告如有未盡事宜，除依法律相關規定外，主辦單位保留修改及補充（包括任何異動、更新、修改）之權利，並以 A+企業創新研發淬鍊計畫網站公告為依據。

九、聯絡窗口

如有任何疑問，請洽經濟部產業技術司專案計畫窗口：

- 工研院產科國際所 出小姐 電郵：minghui_chu@itri.org.tw / 電話：03-5916637
- 工研院產科國際所 葉小姐 電郵：celeste@itri.org.tw / 電話：03-5912578

附件說明：

附件1：Taipei-1系統說明

附件2：提案申請書(系統填報及上傳電子檔)

附件3：結案報告格式(系統填報及上傳電子檔)

附件4：Taipei-1使用者協議(NVIDIA 制式)

附件1：Taipei-1系統說明

Taipei-1是一座 NVIDIA 在臺灣投資建置的人工智慧與數位學生運算超級電腦，目的是為了支持台灣 NVIDIA 人工智慧創新研發中心(以下本文簡稱為本中心)研發所需的運算資源。整體運算資源的部分運算力預計將投入與台灣進行與本中心目標一致的共創或研究計劃。

整體系統包含2個部份的主要運算資源，一個部分是主要運用在 AI 運算與 NVIDIA 最新的前瞻技術開發，採用 NVIDIA DGX H100運算主機構成。另一個部分主要運用在 Digital Twins 數位學生與 NVIDIA 最新的前瞻技術開發，採用 NVIDIA OVX 運算主機構成。相關系統說明如下：

一、Taipei-1 系統規格

1. DGX 系統



整體依據 NVIDIA DGX SuperPOD (<https://www.nvidia.com/zh-tw/data-center/dgxsuperpod/>)參考架構進行規劃建置，採用64台 NVIDIA 最新人工智慧超級電腦 DGX H100 (<https://www.nvidia.com/zh-tw/data-center/dgx-h100/>)，每台 DGX H100 主機具備8個 NVIDIA H100 GPU 透過最新一代的 NVLink 以及 NVSwitch 技術進行整合，整體共計512個 H100 GPU。DGX 主機內每一 GPU 彼此之間具備900GB/s 的雙向溝通頻寬，主機與主機之間透過8條最新 InfiniBand 技術搭配 NVIDIA Quantum-2 InfiniBand (400Gb/s 高頻寬與低延遲高速網路技術 <https://www.nvidia.com/zh-tw/networking/quantum2/>)進行主機間 GPU 運算溝通，為運用在 AI 運算與 NVIDIA 最新的前瞻技術開發的超級電腦。

2. OVX 系統：



整體架構依據 NVIDIA OVX SuperPOD (<https://www.nvidia.com/zh-tw/data-center/products/ovx/>)參考架構進行規劃建置，採用32台 OVX 2.0 主機進行整合，每台主機搭配8個 NVIDIA L40 GPU，整體共計256個 L40 GPU，透過 NVIDIA Spectrum X (<https://www.nvidia.com/zh-tw/networking/spectrumx/>)高速乙太網路架構進行整合，投入符合實際物理情況的 Digital Twins 數位孿生和生成式人工智慧，包含 AI 訓練及推論，NVIDIA OVX™ 系統可以提供最佳高度擬真符合實際物理的3D繪圖技術和運算效能，以及在數位孿生應用研發中整合人工智慧運算的各項技術。

本專案分享 DGX 系統、OVX 系統各25%算力，故為16台 DGX H100、8台 OVX 2.0。

二、Taipei-1 環境軟體說明

DGX 系統中將使用 NGC(NVIDIA GPU Cloud) 相關最新與最佳化的 AI 開發套件 (<https://ngc.nvidia.com/>)，包含 TensorFlow Framework、PyTorch Framework、Computer Vision、Conversional AI、NeMo Framework for Generative AI、BioNeMo healthcare for Generative AI、TAO for pre-training models and optimization 以及 Triton Inference Server 等等 AI 開發環境與套件 (詳細請參考 <https://developer.nvidia.com/>)以協助合作單位與企業使用最適切的開發環境進行研究或共創。

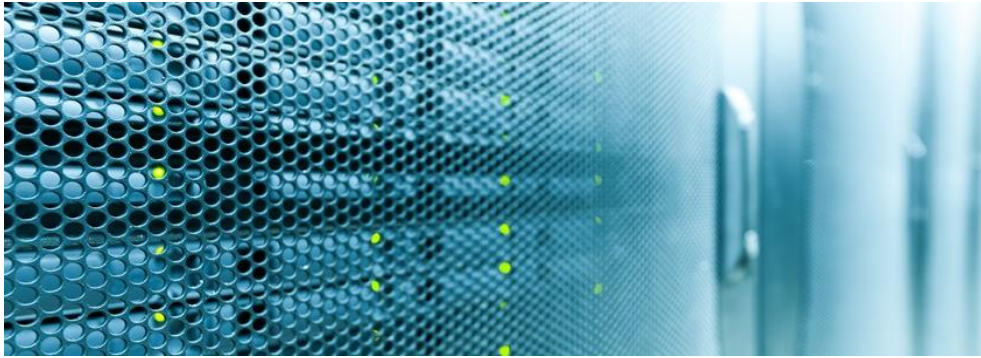
使用者介面搭配 NVIDIA Base Command Platform MLOps 使用介面提供服務，使用手冊請詳閱 <https://docs.nvidia.com/base-command-platform/index.html>。

OVX 環境中搭配 NVIDIA Omniverse Enterprise 軟體，整合 Omniverse Nucleus server、USD Composer、USD Presenter 以及 Audio2Face、Kaolin、Machinima 以及 Isaac Sim 相關 Applications 套件，幫助研發人員共創 Digital Twins 數位孿生目標，相關 Applications 完整文件與資料請參閱 <https://docs.omniverse.nvidia.com/>。

其他 NVIDIA 軟體開發工具套件 (SDK) 相關主題：利用 NVIDIA SDK 提供的技術和工具所開發更多創新應用，以需要大規模運算者為佳。以下舉例介紹數個 NVIDIA SDK 做為參考：

- NVIDIA NeMo 框架 (<https://www.nvidia.com/en-us/ai-data-science/generative-ai/nemo-framework/>)：該框架專注於語音處理與自然語言處理的眾多方向，以滿足各種實際應用場景的需求。
- NVIDIA BioNeMo (<https://www.nvidia.com/en-us/gpu-cloud/bionemo/>)：該平台結合深度學習技術，用於藥物分子設計、蛋白質結構預測等方面，促進生物技術與藥物研發的交叉合作。
- NVIDIA Modulus (<https://developer.nvidia.com/modulus>)：開發物理資訊神經網絡 (Physics-Informed Neural Network, PINN) 應用於科學計算。
- NVIDIA CUDA Quantum (<https://developer.nvidia.com/cuda-quantum>)：協助研究人員在一般計算機上模擬量子系統。
- NVIDIA CLARA (<https://developer.nvidia.com/industries/healthcare>)：醫療相關的研究與應用，用以幫助提升診斷和治療效果。

附件2：NVIDIA 「Taipei-1」提案申請書(格式)



DOIT 經濟部產業技術司
Department of Industrial Technology, MOEA

NVIDIA 人工智慧創新研發中心

NVIDIA 「Taipei-1」提案申請書

申請單位：_____

申請日期：____年____月____日

壹、「Taipei-1」使用申請書

項目	申請內容
主導企業或單位/專案主持人	請勾選： 1. <input type="checkbox"/> 單獨申請 2. <input type="checkbox"/> 聯合申請 主導企業或單位： 專案主持人姓名： 職稱： 電話： 電子郵件信箱：
聯合組隊申請之其他企業或單位 (單獨提案不需填寫)	共有___企業/單位，請羅列如下：
專案主要聯絡人	姓名： 職稱： 電話： 電子郵件信箱：
具執行 AI、數位學生專案，或曾使用大算力等經驗說明	
專案重點說明 1. 專案主題 ：包括以下題目但不限於如生成式 AI 應用、大型語言模型、數位學生等(註明有無其他政府計畫補助) 2. 專案的重要性 ：整體之研究規劃與重要性，本次專案之目的及解決那些問題 3. 預期產出 ：專案計畫預期成果，需有書面結案報告，並以公開展示或影片呈現前後變化 4. 完成專案後的效益 ，如我國提昇前瞻技術、帶動產業共創、提升商業化價值、培育人才等；請具體提出實際可量化之效益(如降低多少開發時間或成本、製程減少多少時間等) 5. 專案對於大算力之需求理由	1. 專案主題： 2. 專案的重要性： 3. 預期產出 4. 預期效益： 5. 算力需求說明： 6. 執行方法與困難：

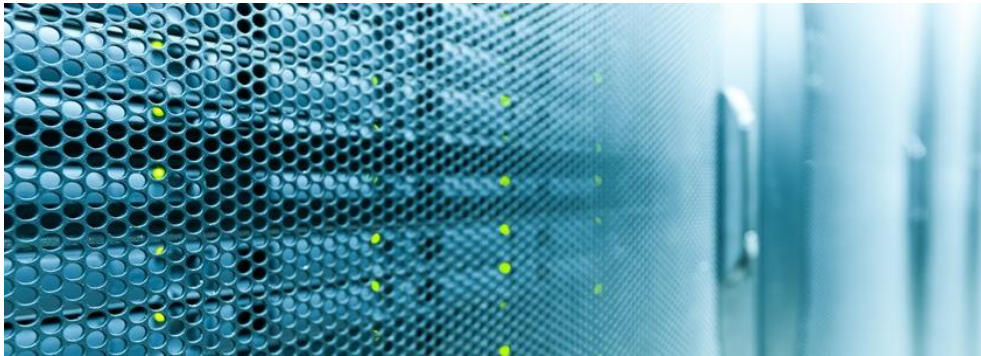
項目	申請內容
6. 專案執行方式及可能遭遇的困難	
專案運算資源需求 (以6周充份使用 DGX 16台或 OVX 8台為原則) (請據實填寫，以利評估及算力配置)	請勾選(擇一)： 1. <input type="checkbox"/> DGX 運算群；所需 ___ 台 2. <input type="checkbox"/> OVX 運算群；所需 ___ 台 所需儲存空間(以 GB 或 TB 為單位)：_____
資料集描述 請說明各資料集使用 Tokens 數(單位：百萬)；若是影像資料模擬，請提供筆數(單位：萬) (請據實填寫，以利評估及算力配置)	(如繁體中文資料庫100百萬 Tokens、英文資料庫200百萬 Tokens， <u>總計300百萬 Tokens</u> ； 影音資料庫200萬筆、動態圖像庫100萬筆， <u>總計300萬筆</u> 。)
聯合組隊之各成員算力需求 (單獨提案不需填寫)	(如：成員 A 開發 OO 模型，使用 OO 筆數據，OO 百萬 Tokens，預估所需 DGX 台數或算力使用時間等)
所需之 NVIDIA 軟體及軟體開發套件 備註： 1. 申請 DGX 運算群者請說明計畫使用 DGX H100時將需要的 NGC 軟體內容 (https://ngc.nvidia.com/) 2. 申請 OVX 運算群者請說明計畫使用 OVX 時將需要的 NVIDIA Omniverse Enterprise Applications 軟體內容 (https://docs.omniverse.nvidia.com/)	
1. 計畫參與人員 2. 算力平台使用人員(每單位最多申請2個帳號) 備註： 請提供人員清單，包括企業或單位、姓名、職稱、email、電話。	
若貴單位曾執行通過之 Taipei-1專案 (請填寫本項目)	曾使用算力期間：___年___月___日~___年___月___日 階段性成果說明：(如已提交結案報告，請一併檢附，如尚未屆時繳交，請扼要說明階段成果。)

項目	申請內容
	本次申請之必要性及差異：(如成員、技術、模型、預期效益等)
補充資訊(請以附件表示)	

備註：

- NVIDIA 要求每個算力使用企業/單位須與 NVIDIA 簽署保密協議(NDA)、使用協議(User Agreement)，相關文件為制式合約無法調整。
- 為協助審查委員瞭解提案內容，上述項目請完整且據實填寫。
- 採系統填報並上傳電子檔，網址：
https://docs.google.com/forms/d/1h1crFBlckn_tZRL6at98v_RqsOyaPgSmBlcyyvBS0dA/viewform?edit_requested=true

附件3：NVIDIA「Taipei-1」結案報告(格式)



NVIDIA 人工智慧創新研發中心

NVIDIA「Taipei-1」結案報告

第____次徵案第____批

DGX OVX (請勾選)

單獨 聯合申請 (請勾選)

專案主題：_____

主導單位：_____

聯合單位：_____

使用日期：____年____月____日~____年____月____日

壹、計畫執行成果

(計畫執行成果以表格重點方式呈現，並輔以影片或圖片/表格，若有補充資訊請於附件說明。)

項目	說明
主導企業或單位/專案主持人	主導企業或單位： 專案主持人姓名： 職稱： 電話： 電子郵件信箱：
聯合組隊申請之其他企業或單位 (單獨提案不需填寫)	共有___企業/單位，請羅列如下：
執行單位1：OO(單位名稱) (請按團隊成員自行增列)	
類別(產/學/研/新創)	
專案主題(技術)	
使用模型	
具體產出	如: 智慧客服大型語言模型
使用算力前之情形	如：模型精確度60%、原來模型運算時間為5周...
使用算力後之提升	如：模型提升精確度至90%、模型運算時間縮短為5天...
技術擴散	如：原本應用於 OO 產業，後續若進行 OO 調整可應用於 OO 產業...
總參與人數(人)	如: 20人
核心人員(人)	如: 資料科學家：2位 軟體工程師：4位
AI 人才培育(選填)	如: 課程(時間)：生成式 AI(2024-7/27-8/15) 參與人數：30人
重要效益與亮點案例	請以影片或圖片/表格輔助說明重要效益與亮點案例 1. 全球(或全台)第一個 OOOO 模型，可協助 OOO，帶動 OOOO(舉例) 說明目前如何、有哪些痛點、因大算力的運用使得此技術改善至如何或特殊處，故在全球(或台灣)為第一個 OO，以及此產出之效益有哪些。 2. 第一個 OOO，可協助 OOO，實際帶動 OOOO(舉例)

執行單位2：OO(單位名稱) (以下表格請按團隊成員自行增列)
....

備註：

- 若為聯合申請，請自行增列。
- 本報告將會於貴單位再次申請時，供委員參酌。
- 結案報告含封面勿超過10頁(不含附件)。
- 採系統填報並上傳電子檔，網址：
<https://docs.google.com/forms/d/1LIHNj0UIONdmMh3BHMfgkVIYq1CGuzGT2HG9fgQEw0k/edit>

貳、附件

其他補充資料。