

AI 對國安的衝擊： 勿高估中國威脅，莫低估未知風險

◆ 中興大學國際政治研究所教授 — 譚偉恩

從安全研究 (Security Studies) 的視角觀之，熟悉 AI 在特定議題領域的應用只是見樹，掌握 AI 可能將引爆社會的哪些風險，並提早加以規範，才是見林。

AI 作為尖端時代科技，在國與國之間容易形成「安全困境」(Security Dilemma)，主權國家能否在 AI 日益普遍的時代蒙受其益，值得吾人關注。作為一種「工具」，AI 可以是國家用來強化自身安全的利器，也可以是弱化他國安全的兵器。

AI 會升級中國帶來的威脅嗎？

《哈佛國際評論》(Harvard International Review) 112 年 9 月間刊登一篇引人注目的文章，指出美國和中國大陸是全球 AI 發展最快的兩個國家，如無意外，也將

是 AI 相關技術的全球領導者。然而，兩國的競爭如果無法被妥善管理，其結果對全世界來說將會是一場難以收拾的大災難。¹

過去一年裡，世人見證了 AI 技術的爆炸性成長，其中「生成式 AI」(Generative AI) 因為能組織及重整現有的資訊，創造協助用戶提高生產力的全新應用方案，而引起全球廣泛關注。據 Goldman Sachs 估算，生成式 AI 具有推動全球 GDP 成長 7% 的潛力。² 111 年 11 月，美國公司 OpenAI 發布的 ChatGPT 大型語言模型 (Large Language Model) 引發全球熱議；時隔不



ChatGPT 為美國公司 OpenAI 開發的人工智慧聊天機器人程式，發布後引發全球熱議；112 年 3 月又推出基於深度學習的 AI 模型，即功能升級的 GPT-4，可更精準地理解自然語言，右圖為 GPT 原版與 GPT-4 對相同問題產出的不同回答。（Source: OpenAI, <https://openai.com/gpt-4>）

OpenAI

GPT-4 surpasses ChatGPT in its advanced reasoning capabilities.

ChatGPT

Input

Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3-30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

Output

Here is a list of their availability:

- Andrew: 11 am to 3 pm
- Joanne: noon to 2 pm, and 3:30 pm to 5 pm
- Hannah: noon to 12:30 pm, and 4 pm to 6 pm

Based on their availability, there is a 30-minute window where all three of them are available, which is from 4 pm to 4:30 pm. So, the meeting can be scheduled at 4 pm.

GPT-4

Input

Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3-30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

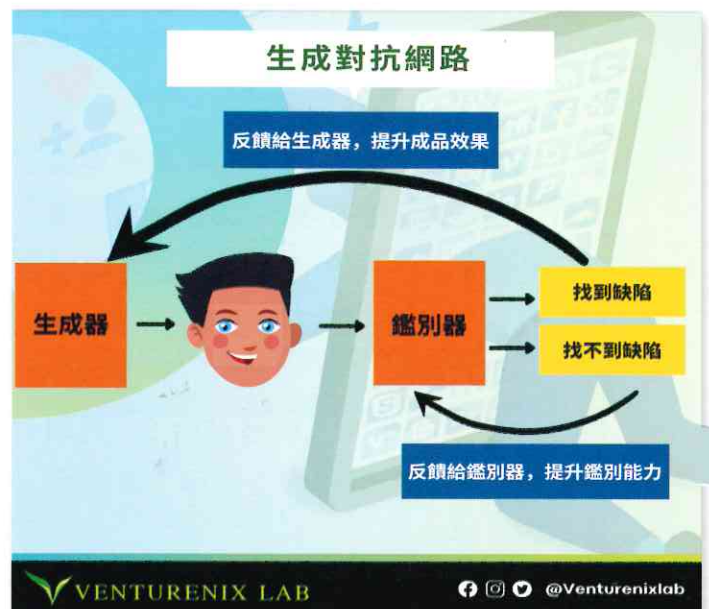
Output

Andrew: 11 am - 3 pm
Joanne: 12 pm - 2 pm, 3:30 pm - 5 pm
Hannah: 12 pm - 12:30 pm, 4 pm - 6 pm

Common availability for a 30-minute meeting: 12 pm - 12:30 pm

久，112 年 3 月又推出一種基於深度學習的 AI 模型，即功能升級的 GPT-4，可以更精準地理解自然語言，因此在終端生成的品質上，能更佳滿足各種領域的需求，例如寫譯出更好的詞句或文章，甚至是設計出有創意或具藝術美感的物件。

生成式 AI 有兩個問題亟待深刻思考。首先，AI 須透過大量數據培訓，因此深度學習的技術品質至關重要，最常見的應用之一是生成對抗網路（Generative Adversarial Networks, GAN）。GAN 建立在「鑑別」與「生成」的核心系統上。兩系統持續相互加乘，也相互競爭；每一回合「後面」的生成，都是基於對「先前」



生成對抗網路建立在「鑑別」與「生成」的核心系統上，每回合的生成都是基於對鑑別結果的學習，且新生成會讓鑑別系統面臨更困難的評斷，如此交換加乘，循環地培訓 AI。（圖片來源：Venturenix LAB, <https://venturenixlab.co/blog-有圖有真相烏克蘭被投降>）

¹ Sam Meacham, "A Race to Extinction: How Great Power Competition Is Making Artificial Intelligence Existentially Dangerous." *Harvard International Review*, September 8, 2023, via at: <https://hir.harvard.edu/a-race-to-extinction-how-great-power-competition-is-making-artificial-intelligence-existentially-dangerous/>.

² "Generative AI could raise global GDP by 7%," *Goldman Sachs*, April 5, 2023, via at: <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>.

洛阳纸贵是什么意思？

根据历史记载，西晋时期左思的《三都赋》在京城洛阳广为流传，人们啧啧称赞，竞相传抄，一下子便纸昂贵了几倍。相传，原来每刀千八百文的纸，一下子涨到两、三千文，后来竟倾销一空；不少人只好到外地买纸，抄写这篇千古名赋。

可以看出，当时洛阳的纸非常昂贵，而且供不应求，这也说明了左思的《三都赋》受到了广泛的认可和喜爱。

这个成语在现在的经济学原理里，对应的理论是什么？

这个成语在现代经济学中对应的理论是需求定律。在商品经济中，供给和需求是相互作用的，当某种商品或服务的需求量大、价格低时，供给量可能不足以满足需求，从而导致价格上涨，即“物以稀为贵”。

在生活中，“洛阳纸贵”这个成语可以用来形容某种商品或服务非常受欢迎，价格被炒高，而且供不应求的情况。例如，在旅游热门季节，一些热门目的地的机票和酒店价格就会更高；相反，在旅游淡季，机票和酒店的价格就更低。

用洛阳纸贵四个字写一首藏头诗

洛阳城里春光好，
阳艳无双不負賞。
紙貴滿天詩詞賦，
貴比黃金樂未央。

重新生成

请输入问题或“/”获取模板



中國大陸目前推出具有文學創作、文案輸出、數理邏輯推理，以及中文詩詞賞析功能的「文心一言」（ERNIE Bot）；左圖為其解釋洛陽紙貴的經濟學原理及創作詩句，右圖則為其按使用者的指令繪製海報，製圖方面在品質與正確率仍有待提升。（圖片來源：百度文心一言 AI 生成、製圖）

鑑別結果的學習，且該新的生成都讓鑑別系統面臨更困難的評斷，如此交換加乘，循環地培訓 AI。美「中」兩國現在都在競相發展生成式 AI，並試圖應用於軍事層面，一旦時機成熟，可以想像在戰場上生成足以混淆敵軍的圖像，或是結合 3D 列印，生成欺敵或誘入陷阱假空間等「軍用版」應用程式將誕生。

乍聽下，AI 似乎有點無所不能，而且在關鍵技術面的突破，很可能瞬間扭轉強權競爭間的勝負之勢。然而，目前中國大陸的大型語言模型，在品質、AI 訓練課程、晶片技術以及研發人才方面，似乎還與美國有 2 至 3 年的差距。儘管部分文獻指出，2019 年起中國大陸研究 AI 的論文數量已

超過美國，但僅為表象數字基礎的比較，無法精準瞭解美「中」在 AI 發展的實際情形。現況而言，中國大陸目前比較能端出檯面的，是具有文學創作、文案輸出、數理邏輯推理，以及中文詩詞賞析功能的「文心一言」（ERNIE Bot）。它在網路公測期間，不僅能產出如何續寫科幻小說《三體》的建議、解釋洛陽紙貴背後的經濟學原理，還能準確計算複雜的雞兔同籠問題，並能按使用者的指令繪製海報（類似另一 AI 軟體「Midjourney」）。

因生成式 AI 的水平極度仰賴它學習的資料質量，據網路科技調查權威「W3Techs」統計，全球網站有 53% 是英文，中文網站只有 1.4%，因此單以資料量



臺灣研發 AI 的環境不比國外差，但和中國大陸一樣面臨人才缺稀的困境；AI 的發展是一場馬拉松，過程中毋需過度擔憂中國大陸帶來的威脅，但要穩定加速自身研發技術與人才培育的步調。

球 5% 左右。根據清華大學中國科技政策研究中心發布的《中國人工智能發展報告》，現階段全中國大陸的 AI 人才僅有美國的 1/5。這意謂除非中國大陸能從國民基本教育一路到高等或研究所教育，都執行 AI 人才的培養，否則幾乎不可能趕上美國。而臺灣研發 AI 的環境不見得較差，但人才數量卻明顯不足，中國大陸面臨的困境對我國而言是一個啟示，即 AI 的發展是一場馬拉松，過程中毋需過度擔憂中國大陸帶來的威脅，但要穩定加速自身研發技術與人才培育的步調。

來說，美國研發的 AI 占有先天學習優勢，更別提「理解中文」在 AI 技術層面還有許多困難要克服。其次，隨著美「中」競爭強度升高，美國目前對中國大陸施行許多 AI 領域關鍵技術的貿易管制，包括雲端計算資料中心的微處理器，以及半導體的晶片生產儀器等，也無疑抑制了中國研發 AI 的空間。第三，也是本文認為最重要的一點，就是美國比中國大陸更有能力在其經濟市場上應用 AI 科技，並以此帶動多元且不息的 AI 產業生態鍊。

中國大陸如果要在 AI 領域與美國分庭抗禮，必須取得高質量的數據資料與專業人才，儘管中國大陸占全球 AI 研發近乎一半以上的比例，但其吸引的 AI 人才只占全

示，即 AI 的發展是一場馬拉松，過程中毋需過度擔憂中國大陸帶來的威脅，但要穩定加速自身研發技術與人才培育的步調。

何種 AI 的風險被忽略或低估？

AI 的討論範疇，目前已漸漸擴大而模糊化，有時可能談的是弱 AI（例如電動遊戲），有時則是強 AI（例如前面提到的生成式技術）；計算機科學領域則有符號式 AI（例如專家系統，由開發者完全指定系統已知的對象和元件的相互關係）或次符號式 AI（例如在海量數據基礎上進行自我訓練的人工神經網絡）。這些特徵表明 AI 不應被視為封閉的工業技術系統，而是一種複雜、動態、具開放性的「社會技術系

統」，透過資料學習與意識產出，跟人類社會各領域進行互動；亦即 AI 雖然是一種協助工具，但人類也在使用它的過程「被建構」。此現象挑戰了人們對於「什麼是 AI」，以及「AI 如何影響人類生活」的既定認知和前提假設；許多關於 AI 影響政治、經濟、產業轉型和國際關係的文獻中，將人類社會的改變理解成 AI 技術與應用下的依變數，忽略掉 AI 之所以能夠藉由學習過程產生智能的數據，來自於人類社會的事實。

綜上，本文定性 AI 所帶來的風險，無法僅靠計算機科學技術的解決方案因應，而需要有融入社會元素的解決方案。對 AI

的規範和管理，必須認識到其社會複雜性的風險，且肇因往往是二個以上因素的交互作用。基此，AI 時代的國家安全，不能單從已發生損害事件的經驗法則角度，而必須同時藉由「事件已發生、可能發生，甚至從未發生」的風險級別角度，來設計管理及預防的因應系統，避免低估或錯估的狀況發生。目前可行的方法，是建立包括臺灣內部資料和與其他相似情況國家外部資料的 AI 數據資料庫，也就是讓我們的 AI 系統在處理國家安全事務時，擁有全面性的臺灣觀點，也有可交叉比對異同性的國際觀點。其理論依據來自於：所有訓練 AI 的數據資訊皆有其盲點，如果來源趨於單一，AI 即可能在處理特定議題領域時表



AI 不應被視為封閉的工業技術系統，而是一種複雜、動態、具開放性的「社會技術系統」，透過資料學習與意識產出，跟人類社會各領域進行互動。

現傑出，但在跨議題領域無法觸類旁通，做出統整性的取舍判斷。³ 然而，技術面要讓 AI 成為跨領域智能生成的「通才」，需要一段時間，在此之前不該低估 AI 犯錯的風險或能力有限之事實；當能更細心地考慮到 AI 力有未逮之處，規範上就可以實現更佳的監管，減少意外和失敗。

結語

持平而論，目前那些在研究與開發過程中可預見的 AI 風險，在社會和法律制度往往可獲得某程度的規範，但跨國性的 AI 監管機制仍不易形成，目前多半倚靠道德原則（例如 OECD 的人工智慧原則）、專業守則（例如 Microsoft 的《The Future Computed》）以及自願標準（例如 ISO/IEC 23894 關於 AI 風險管理的建議），一方面不具有拘束力，另一方面可能一段時間後就跟不上 AI 發展的速度。

就臺灣而言，AI 時代帶來的國家安全衝擊有內外兩種。外部衝擊主要是來自中國大陸透過生成式 AI，對臺灣以影響或干擾民眾的方式，左右對特定候選人的認知，以進行「介選」，或者逐行認知作戰。然而，由於中國大陸訓練 AI 的資料不夠多元與開放，加上相關領域的人才不足，在可預見的將來能對我國造成的安全衝擊仍屬有限。



跨國性的 AI 監管機制不易形成，Microsoft 的《The Future Computed》專業守則和 ISO/IEC 關於 AI 風險管理建議的自願標準都不具拘束力。（Source: Microsoft, <https://news.microsoft.com/futurecomputed/>; IEC, https://webstore.iec.ch/preview/info_isoiec23894%7Bed1.0%7Den.pdf）

承上，內部衝擊反而是當前國家安全應予關注的重點，《人工智慧基本法》草案雖在 112 年 3 月底出爐，但現行內容對於建構安全的 AI 產業環境，還有保障人民權利的設想受到傳統科技發展的立法侷限，忽略 AI 源出於社會並且和社會有著相互建構之特徵，而在產業發展及政府監管的規範有強化空間。建議不妨參考韓國的《智慧資訊化基本法》，提供更具體的市場應用誘因，並且依 AI 實際功能的類型將風險分類，然後進行區分管制與風險監測機制，保持對 AI 治理的動態性，兼顧技術創新與社會衡平。

³ 一般人很難理解 AI 運作，即便專業培訓神經網絡或懂得使用 Python 的工程師，也很難全盤通曉 AI 複雜的程式模型，因為 AI 的深度學習能力係由許多小型程式組合而成，每個程式有不同的資料庫與運算規則，最終智能生成的決策樹狀結構相當複雜。文獻指出，AI 的學習能力在相當程度上代表它們的行為是不可預知的，也就是「意外」一定會發生。由於神經網絡係透過許多層次的人工神經元組成，在訓練過程中會對特定的數據資料變得敏感，但 AI 無法向設計者回報該情況，以致人類只能透過觀察 AI 輸入不同資料後的反應，來分析與研究 AI 的學習情況，表示 AI 理解人類世界的方式目前還是一個未解的謎，而人類自己理解世界的方式與 AI 理解世界的方式，也還沒找到對應的橋梁。