

# 聯合國首辦 AI 治理對話

## 籲力保兒童安全

古特雷斯促業界簽署安全承諾 倡設全球基金防算力壟斷



● 聯合國首屆「全球人工智能治理對話」當地時間7月7日在瑞士日內瓦揭幕。法新社

香港文匯報訊 聯合國首屆「全球人工智能治理對話」當地時間7月7日在瑞士日內瓦揭幕，為期兩日的會議匯聚各國政府、科技業界、學術界及公民社會逾4,000名代表，共同商討如何制訂全球性規則，以降低人工智能（AI）的潛在風險，同時善用這項技術帶來的機遇。聯合國秘書長古特雷斯在開幕式上發出嚴正警告，指AI的發展速度已遠遠超出所有人能跟上的步伐，必須即時建立全球治理機制，確保這項技術造福人類，尤其要將兒童安全置於監管首位。

古特雷斯在會上表示，AI這項科技能重塑經濟、改變工作形態、左右選舉結果乃至破壞安全平衡，其發展速度之快，包括開發人員在內的所有人都難以跟上。他特別指出，互聯網花了15年才觸及10億人口，而AI只用了兩年便達到相同規模。在沒有人研究過會對兒童帶來什麼影響時，AI就已接觸到小孩，涉入他們的學習、友誼，以及他們最私密的問題。

### 創新需要護欄規範約束

古特雷斯呼籲業界簽署一項AI兒童安全承諾，要求開發這些系統的公司必須先證明其安全性，才能讓兒童接觸使用。此外，系統不應被允許生成兒童色情圖像，當兒童表現出痛苦跡象時，系統應停止運作，並將他們轉介給真人尋求協助。

古特雷斯坦言，所謂的隨想編程，即使用AI表達你想要什麼而非自行撰寫程式，雖然可帶來奇想，但不能靠隨想編程來定義真相，也無法靠隨想編程來決定人類未來。他強調，創新需要護欄規範約束，若AI要展現強大力量，就必須受到治理。隨着AI以飛快速度持續發展，目前世界正進行一場沒有計劃、也未經同意的實驗，而這種情況不能繼續下去。

### 籲協助發展中國家提升AI能力

古特雷斯同時警告AI權力集中於少數企業與少數國家手中的風險。根據聯合國支持的獨立科學小組首度針對AI所做的全球評估報告，在全

球前500台AI超級電腦中，美國佔據75%的算力，不少國家所佔份額微乎其微。這意味着發展中國家在AI的發展上幾乎沒有發言權，並面臨被拋在後頭的風險。古特雷斯表示，協助發展中國家提升AI能力、促進普及化至關重要，以避免目前的數碼落差進一步加劇成為AI落差。他將敦促聯合國設立全球AI基金，以在全球各地培育數碼能力及負擔得起的運算資源。大會主席貝爾伯克表示這項基金成立勢在必行，但她未有透露具體金額，強調最大挑戰在於這些資金能否真正造福所有人。

### 第二次對話料明年紐約舉行

今次全球對話預計不會產生任何具體決定，但大會共同主席洛佩茲向媒體表示，此次會議將奠定良好基礎，第二次對話預計明年在紐約舉行，屆時將推出更全面的報告。科學專家的獨立報告發現AI的發展甚至比預期更集中，聯合國希望透過持續的全球對話，逐步建立共識與規範框架。

古特雷斯在總結時強調，AI的治理不能再等待，隨着AI技術持續滲透日常生活各個層面，從教育、就業到軍事安全，聯合國倡議的全球治理框架能否獲得主要大國與科技企業的支持，將是未來一年國際社會關注的焦點。

## 古特雷斯：應全面禁用「殺手機械人」

香港文匯報訊 聯合國秘書長古特雷斯7日於「全球人工智能治理對話」上，強烈呼籲國際社會制訂法律，全面禁止致命自主武器系統，即所謂的「殺手機械人」。

古特雷斯直言，讓機器在沒有人類控制和判斷的情況下，自行選擇、鎖定及攻擊目標並奪走生命，這種做法在道德上令人完全無法接受，強調必須制定國際法來徹底禁止這類武器。古特雷斯的表態，令今年初人工智能（AI）公司Anthropic與美國國防部因自主武器問題而對簿公堂的爭議，再次成為外界關注焦點。

### 「有些決定必須由人類作出」

古特雷斯表示，他最擔憂AI在軍事領域的發展，尤其這類武器完全摒棄人類的判斷。教宗利奧十四世在今年春季發表的教宗通諭中亦表態支持，警告AI驅動的武器系統或會降低先進國家發動戰爭的政治成本，使戰爭變得更易發生且脫離人類掌控。

支持研發AI武器的軍工企業則認為，自主武器仍會處於人類監控下，AI只是輔助指揮官提升決策的速度與準確度。然而反對者指出，AI系統同樣會出錯，在戰況激烈的關鍵時刻，將生殺大權交給機器是極其危險的做法。古特雷斯明確站在反對一方，呼籲國際社會不應等待暴行發生後才採取行動，有些決定必須永遠由人類作出，而當中最重大的，莫過於剝奪他人生命的決定，他強調制訂相關安全法規是刻不容緩的任務。

## OpenAI等警告失控風險 籲成立國際監管機構

香港文匯報訊 人工智能（AI）技術發展屢創突破，但失控風險亦與日俱增。美國AI企業OpenAI與Anthropic近期發出呼籲，主張成立國際性組織以監督AI發展，並在必要時協調全球放緩頂尖AI技術研發進程，確保社會安全與監管能力能跟上技術演進的腳步。

OpenAI行政總裁阿爾特曼與首席科學家帕喬發發表網誌文章，明言此類組織核心目標之一是促使各國能採取協同行動，包括在必要時減慢前沿AI開發速度。兩人更預測到2028年，OpenAI相當大比例的研究工作將可能由AI系統與研究人員協作完成。Anthropic的聲明亦指出，單方面減緩全球AI發展速度可能是有益的舉措，原因在於公司內部觀察到「遞迴式自我改進」現象，即AI模型能協助訓練下一代更先進AI系統，此能力雖可加速技術進步，卻同時提高人類喪失控制權的風險。

這種被稱為「智能爆炸」的情境早於2023年已引發科技界高度警覺。當時由馬斯克等業界翹楚聯署的公開信，便呼籲暫停開發前沿AI模型6個月，理由是商業利益可能凌駕安全考量，繼而衍生不可預測的風險。隨着OpenAI與Anthropic相繼邁向公開市場，外界憂慮企業對股東的獲利責任或凌駕對AI安全議題的重視。兩家公司接連呼籲成立獨立國際監管機構，被部分觀察人士視為一大警號。

## AI引科企裁員潮擴大 微軟裁4800人

香港文匯報訊 美國軟件巨頭微軟6日宣布即時裁員4,800人，佔總員工2.1%，其中遊戲主機XBOX部門佔1,600個裁員名額，佔該部門員工總數約五分之一，這是該公司在人工智能（AI）時代降低成本的最新舉措。

美媒CNBC報道，在微軟任職長達27年的執行副總裁柯爾曼在內部信件中寫道，「在我任職期間，科技的開發、部署與應用方式正以前所未有的速度發生轉變。」他表示，儘管股市對微軟的擔憂大多集中在AI領域布局，以及行政總裁納德拉遲遲未能提出一套清晰的策略，說明微軟將如何發展AI模型、AI代理等服務，但AI取代人力並非此次裁員的直接原因。柯爾曼稱：「真正的情況是AI正改變工作的方式，我們的客戶也正經歷同樣的轉變，他們期待我們協助度過這一過程，若我們自己都做不到，就無法真正幫助客戶。」

XBOX行政總裁夏爾馬則向部門員工表示將合共裁員3,200人，其中一半包含在今次裁員計劃內，其餘將在下一財年內陸續離職。

## 全球最大數據中心流產 美AI基建擴張遇重挫

香港文匯報訊 據美國傳媒報道，華爾街資管巨頭黑石集團旗下的數據中心營運商QTS，日前正式宣布終止籌劃多年的超大型數據中心項目，並撤回全部規劃文件，意味着這個全球史上規模最大的數據中心項目正式宣告流產，輿論視為美國人工智能（AI）基礎設施擴張遭遇重大挫折。

該個名為「數碼門戶」的項目，原計劃在弗吉尼亞州威廉王子縣建設一個佔地約850公頃的園區，預計耗資達1,000億美元（約7,841億港元），將成為全球最大的數據中心園區。據路透社報道，項目原計劃為威廉王子縣帶來數



● 美國數據中心扎堆建設推高電費賬單。圖為美國居民在家門外豎起反對建設數據中心的標語牌。網上圖片

十億美元投資，創造「可觀的年度地方稅收」和「數千個長期職位」。然而儘管該計劃已獲得威廉王子縣監事委員會批准，但面對多年來當地民眾的反對和訴訟，QTS最終還是選擇放棄。

### 集中能源損耗遭嚴重質疑

分析認為，這場圍繞數據中心建設引發的矛盾，折射出美國AI基建狂瀾背後的能源供應困境。據美國霍士新聞網報道，弗吉尼亞州利科縣縣長維爾爾卡斯在6月26日要求政府員工節約用電，以應對高達25%的電價暴漲。雖然維爾爾卡斯沒有直接將電價上漲歸因於數據中心建設，但弗吉尼亞州聯合立法審計與審查委員會2023年發布的數據中心評估報告認為，該州數據中心產業的繁榮大幅推高能源需求，普通居民和企業客戶將為數據中心的巨額電力消耗買單。

據美國資料中心網站統計，弗吉尼亞州目前共有637個數據中心，大部分位於其北部的「數據中心走廊」附近，該地區被公認為全球最大數據中心市場。然而高度集聚的數據中心產業布局帶來的集中能源損耗，正引發嚴重質疑。美媒認為，「數碼門戶」被叫停不僅標誌着一個計劃的失敗，也凸顯美國資本密集型AI基礎設施部署方式的結構性轉變，產權風險、電力供應及地方政治阻力正逐漸成為限制其發展的瓶頸。



金融法規跟不上AI 歐銀行界：須改革監理模式

香港文匯報訊 歐洲多名政策制定者與金融監管官員稱，人工智能（AI）發展速度已超越現行金融監理制度，監管機關正面臨如何在推動AI應用與維護市場穩定之間取得平衡的挑戰。

英國金融行為監管局行政總裁拉希接受美媒CNBC訪問時表示，在AI技術快速演進、尤其AI智能體持續發展的情況下，傳統法規制定流程已無法跟上科技變化，監管機構必須重新思考如何因應AI帶來的各項創新。拉希指出，英國透過金融穩定委員會推動前沿AI研究，同時成立AI安全研究所，希望協助政策制定者、監管機關及企業更了解AI風險，並安全導入相關技術。

### AI智能體或加劇市場波動

歐洲央行行長拉加德受訪時則指出，AI確實有助提升生產力，但同時也是一項重大風險，過去約10年來，外界主要關注網絡安全、黑客攻擊及資料竊取等問題，然而隨着AI模型快速發展與能力持續提升，當前面臨的風險更嚴重，但相關防禦措施及所需資金仍未到位。

在上周的歐洲央行年度會議中，AI對生產力及金融市場完整性的影響成為重要討論議題，英倫銀行副行長布登登提到，AI智能體未來可能在金融市場承壓時進一步放大波動。隨着AI智能體在金融市場應用增加，監管單位或需建立更嚴格監督機制，例如比照金融市場熔斷機制，當AI模型發生錯誤並可能導致市場崩盤時，可立即限制或停止整體市場交易。