

AI藏危機如「地雷陣」 中美對話裝「安全閥」

填補安全應急領域空白無礙產業競爭 為全球管控前沿技術提供範例

人工智能（AI）浪潮席捲全球，已然成為新時代大國博弈與全球發展的核心賽道。中美兩國作為全球AI領域的兩大核心力量，競爭與共生並存，博弈與協作交織，雙方圍繞AI展開的競爭與合作，不僅關乎兩國科技未來，更牽動全球數字化進程與人類科技福祉。 ●香港文匯報特約記者 張浩



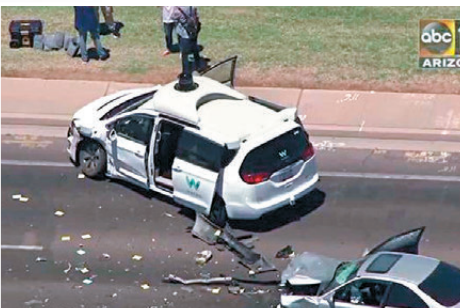
●美國抗議者呼籲暫停人工智能開發。資料圖片

在9月23日舉行的聯合國安理會「人工智能與國際安全」高級別會議上，聯合國秘書長古特雷斯呼籲中美兩國在AI領域保持對話，加強合作。他說，「AI正飛速發展，其潛在風險超出人類充分理解的能力。作為AI領域領先國家，中美肩負特殊責任，應當保持持續對話，增強透明度，共同管控風險，避免技術分裂。」

古特雷斯的呼籲代表世界上大多數國家的焦慮和期待。長期以來，中美兩國作為全球公認的AI領域兩大強國，處於「競爭多於溝通、博弈大於合作」的狀態。為維持技術領先優勢，美國持續收緊高端算力芯片、核心EDA軟件、先進製造設備的對華出口管制，試圖從源頭遏制中國大模型迭代與智能產業升級。面對技術封鎖，中國堅持自主創新、深耕產業落地，快速完善本土算力體系、基礎模型矩陣與行業應用生態。這種技術壁壘與產業博弈，也被多國政界與學界視為中美AI關係失序的核心癥結。

美國前代理助理國務卿桑賴直言，一味將中國視作AI領域的競爭對手、通過封鎖遏制對方技術發展，是典型的戰略誤判，「我們應確保彼此都能使用這些技術，將其用於造福人類、確保科技安全。中國恰好是其中的關鍵一環，我們需要與中國合作。」一些中國學者也指出，中美AI領域長期處於「有競爭、無疏導，有博弈、無管控」的失衡狀態，單邊技術遏制只會加劇對立，無法從根本上解決AI全球性風險治理難題。

在此背景下，一旦出現AI深度偽造與情跨境擴散、智能系統異常風險、AI工具惡意濫用等跨國突發事件，雙方沒有



●Waymo自動駕駛汽車發生車禍。網上圖片

專屬對接機制，容易誘發戰略誤判，讓單純的技術風險升級為雙邊對立隱患。這種失衡狀態，長期困擾着中美AI關係，也給全球AI安全埋下隱患。

避免技術意外引發戰略誤判

9月24日舉行的中美元首會晤上，習近平主席指出：「中美都是人工智能大國，雙方有競爭，更可以合作，應該各展所長而不是相互設防。雙方可以繼續開展人工智能對話，就其風險和收益進行交流，共同防範人工智能被濫用惡用。要保持人類對人工智能技術的管控，秉持以人為本、向上向善理念，讓人工智能為人類的進步服務。」特朗普總統也就AI議題表態：「AI關乎人類未來，美中應就此保持對話，加強合作。」依託元首會晤達成的共識，中美雙方正式承諾用好中美AI對話機制，並且建立AI事件溝通渠道，一旦發生重大AI相關突發事件，雙方能及時互通信息，避免技術意外或惡意使用引發戰略誤判，雙方商定下一輪中美AI政府間對話將於11月舉行。

常態化對話+突發事件溝通

這套新機制包含兩大核心。第一是常態化AI對話機制。對話的討論範圍聚焦AI本身：包括前沿大模型的安全風險評估方法、社會收益、倫理規範、防範AI惡意濫用等議題。雙方可以定期交換風險的研判，探討AI向善發展路徑。第二是AI突發事件溝通渠道，被不少國際媒體類比為AI時代的「危機熱線」。這條渠道的核心功能是應急處置，一旦出現由AI引發的重大跨國突發事件，例如大規模AI深度偽造信息跨境傳播、AI系統失控或被用於重大網絡攻擊，雙方快速通報、交換信息，防止因信息不對稱導致事態升級，避免單純的技術事件演變為大國之間的戰略衝突。

需要明確，這一機制的定位是風險管控工具，不是技術合作協議。機制並不涉及雙方共用大模型源代碼、交換核心AI演算法、開放前沿技術成果。面對AI這種具備強外溢效應的新技術，大國之間最需要防範的是突發AI事件帶來的意外誤判。對話的目標不是消除兩國在AI產業上的競爭，而是增設安全護欄，這一機制的設立，填補中美在AI安全應急領域的空白，也為全球大國管控前沿技術風險提供可參考範例。

《華爾街日報》等媒體分析指出，中美達成的AI共識屬有限合作。兩國願意在風險管控領域保持溝通，但底

層算力、芯片產業鏈依舊是大國科技競爭的主戰場。AI領域如同當前中美關係的一個縮影，呈現出「競爭與管控並存」的狀態：一方面設立危機熱線防止誤判；另一方面算力、大模型產業競爭持續推進，並行不悖。



●一台半身人形機器人在進行包裹分拣展示。資料圖片

世界樂見中美攜手完善AI治理

當前全球AI發展早已形成深度交融的生態，任何技術壁壘、陣營割裂都會產生全球性負面效應。中美在AI領域既有競爭，更有共同責任。國際社會呼籲兩國以對話疏導分歧、以合作管控風險，攜手完善全球AI治理規則，讓AI真正造福世界各國人民。中美元首會晤達成AI對話、風險管控等共識，為全球AI發展注入穩定性。

絕大多數國家均樂見中美跳出零和競爭，通過常態化對話與良性合作管控風險，共同守護全球AI產業健康發展格局。中小國家與廣大發展中國家，並無意願捲入AI競爭博弈，反而普遍擔憂技術脫鉤帶來的發展危機。新加坡總統尚達曼曾指出：「AI領域的重大風險，必須由中國和美國共同參與處理。全球AI治理離不開中美兩國，如果中美走向AI技術割裂，小國將承受最大代價。」南非總統拉馬福薩直言：「如果中美AI競爭走向兩套互不相容的技術體系，全球南方國家將被迫承擔高昂成本。」長期以來，發展中國家依託全球通用科技體系推進民生、醫療、農業領域的AI應用，中美若陷入技術對峙、標準割裂，不僅會加劇全球數字鴻溝，更會讓全球AI治理陷入碎片化困境，阻礙AI普惠價值的落地。

順應世界各國的普遍期待，中美唯有跳出對抗桎梏、摒棄技術割裂，在競爭中提質增效，在合作中守護大局，才能破解全球AI治理難題。此次中美元首會晤搭建起中美AI對話的全新橋樑，為科技大國博弈裝上「安全閥」。以此為起點，新時代的AI賽道，不是非贏即輸的零和棋局，而是攜手共進、共治共贏的廣闊藍海。

浩海
觀瀾

良性競爭激發創新活力 務實對話穩全局防風險

張浩

進入數字時代後，傳統的大國競爭範式被徹底打破，不僅局限於領土、軍事、工業等傳統賽道，而主要由大國關係和先進技術雙重主導，博弈的邊界也不斷拓寬：科技賽道競爭、國際規則制定、數字空間治理、產業鏈話語權爭奪等。

中美雙邊博弈構成當下全球政治、經濟、地緣戰略的基調。兩國的競爭節奏與合作態勢，直接左右全球產供鏈布局、國際經貿規則，以及全球力量平衡。其中，AI作為兩國競爭核心技術，正向各領域滲透，重塑全球生產力底層邏輯、國家安全防禦體系和社會治理模式，重構各國核心競爭力評價體系。

縱觀世界近現代史，唯一能與當下場景相提並論的歷史範式，便是冷戰時期的美蘇雙邊關係，以及兩國圍繞戰略核武器進行的長達數十年的對話、博弈與管控。冷戰時期，美蘇關係構築起全球兩極格局的主基調，戰略核技術則作為當時的

頂級顛覆性技術，牢牢掌控全球安全底線，核均勢也成為那個時代全球戰略穩定的基石。

當下中美關係疊加AI數字革命，正推動人類從「核均勢穩定」轉向「數字智能戰略穩定」。AI已超越單純技術賽道，成為重塑大國權力結構、改寫國際秩序底層邏輯，主導廿一世紀人類文明走向的「核心戰略變數」。中美圍繞AI領域展開的戰略博弈，不僅是算力、模型與產業之爭，更關乎全球規則、技術倫理與風險管控體系的重塑。這場博弈具有雙重屬性，競爭會持續驅動創新突破，合作則是穩住全局、防範災難風險的壓艙石。各國已看清，一旦AI賽道走向陣營分裂，代價終將由全球所有國家共同承擔。就像巴西總統盧拉所說，「我們不希望AI變成大國博弈工具。中美之間保持對話、管控分歧，才能防止科技壁壘不斷加高，讓南方國家有機會共享人工智能紅利。」

美蘇時代的戰略穩定，依託核二次打擊能力、固定的危機溝通機制實現風險管控，維持數十年的平衡。但長期的全面對抗、零和博弈，最終導致陣營割裂、大國崩潰，以及全球秩序重塑。數

字時代的戰略風險比核時代更複雜，AI具備決策速度快、誤判風險高、技術迭代快、供應鏈可武器化的特點，極易引發新型大國衝突，成為全新的全球不穩定因素。

此次中美元首會晤，在技術競爭之外搭建起兩國常態化對話渠道。這表明，相對於互相確保摧毀的「核平衡」，新的戰略穩定，更多表現為雙方深度嵌套的技術競爭，更注重及時溝通、風險預警與底線共識。良性競爭能激發創新活力，務實對話則可對沖演算法誤判、技術脫鉤等潛在危機。

AI將持續推動全球戰略格局洗牌

未來，人工智能將持續推動全球戰略格局深度洗牌，打破傳統地緣力量邊界，重塑新舊大國的優勢位次。在這一歷史性變革中，中美唯有以競爭啟動創新、以合作管控風險，摒棄陣營對抗與技術分裂，順應全球多數國家對穩定、普惠、有序AI治理的共同期待，方能駕馭智能時代的變局，引領廿一世紀大國戰略格局朝着更包容、均衡、穩定的方向演進。

世界潮流，浩浩湯湯。大國關係與先進生產力，始終是驅動世界潮流向前演進的主要動力。

這個時代的大國關係，正如美國國務卿魯比奧所言，「中美關係的走向，將直接定義整個廿一世紀的全球格局。」而這個時代的先進生產力，則像聯合國秘書長古特雷斯指出的，人工智能（AI）「處於我們共同未來的中心」，且將成為「廿一世紀的偉大均衡器」。二者相互交織，深度綁定，在本世紀餘下的74年裏持續博弈、協同演進，成為貫穿全球發展的主線。

回望過去一百年，全球格局競爭邏輯依然是依託地緣疆域、工業產能，以及軍事霸權來展開。大國戰略平衡，亦遵循國際關係均勢理論，依靠軍力、版圖、工業規模等硬實力實現對沖。大國關係穩定的核心就是力量相互制衡，邊界明確、規則清晰，競爭維度也相對單一。