香港文匯報訊 「ChatGPT」等人工智能（AI）系統走紅全球，但過度使用AI技術的風險亦引發科技界擔憂。科技非牟利組織「未來生命研究所」28日就發表公開信，呼籲所有AI實驗室暫停訓練較AI聊天機械人「GPT-4」更強大的系統至少6個月。截至29日，公開信已獲得包括億萬富豪馬斯克在內逾1,000名科技界知名人士聯署，他們倡議加強對AI技術的監管，必要時更不排除政府介入。

馬斯克等逾千科企領袖 憂AI失控

公開信籲停訓練較GPT-4更強大系統半年 倡政府介入監管

未來生命研究所主張引導變革型技術，遠離極端擴張風險，轉向造福生活。今次公開信提到，研究表明具有與人類競爭智能的AI系統或會對社會和人類構成深遠風險，先進的AI技術更或代表地球生命史的深刻變化，需要詳盡的規劃和管理。「不幸的是，這並沒有發生，AI實驗室陷入了失控競賽，沒有人——甚至它們的創造者——能理解、預測或可靠地控制技術發展。」

假信息歧視偏見影響人類思維

公開信形容，現有AI技術在處理常規任務上已具有一定競爭力，但虛假信息、歧視偏見甚至影響人類思維等問題也隨之浮面，「我們應該冒險失去對我們文明的控制嗎？只有當我們確信它們有積極影響且風險可控，才應該開發強大的AI系統。這種信心必須有充分的理由，並隨着系統潛在影響的規模而不斷增加。」

科技界領袖在信中呼籲，全球所有AI實驗室應立即暫停訓練較GPT-4更強大的AI系統至少6個月，如果企業間無法協調，各國政府應適當介入。在此期間，AI實驗室和獨立專家應開發和實施一套設計具有普遍約束力的安全協議，由獨立的外部專家進行審計監督，「這並不意味完全暫停AI開發，只是緊急從奔向不可預測的危險競賽中收回腳步。」

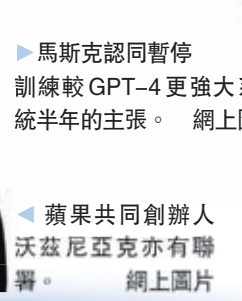
促盡快開發強大AI治理系統

信中還指出，AI開發人員應當與政策制定者合作，盡快開發強大的AI治理系統：包括專門負責AI事務的監管機構，監督跟蹤AI系統的計算能力；推出來源標識系統，幫助區分原創和合成信息源；加強認證系統運作，界定AI帶來的傷害以及對應責任方；以及為AI技術研發提供充裕資金，應對一旦濫用AI可能帶來的巨大經濟或政治破壞。

信中提到，GPT-4母公司OpenAI最近的聲明還指出，在開始訓練未來投入使用的AI系統之前，進行獨立審查或許非常重要，對於最先進的系統而言，業界或要達成共識，避免用於創建新模型的計算量無限增長，以至超出研發團隊的控制，「我們認為現時就是要加強監管的時候。」

這封信最後寫道，「人類可以享受AI帶來的繁榮未來。人類社會已經暫停其他可能對社會造成災難性影響的技術，那麼在AI領域，我們也可以這樣做。讓我們享受一個漫長的AI之夏，而不是毫無準備地陷入秋天。」

▶馬斯克認同暫停訓練較GPT-4更強大系統半年的主張。 網上圖片



◀蘋果共同創辦人沃茲尼亞克亦有聯署。 網上圖片

微軟網絡安全AI助手「Copilot」可同時處理千警報防駭客

香港文匯報訊 研發人工智能（AI）聊天機械人ChatGPT的OpenAI備受科企微軟重視，微軟還將採用OpenAI技術的網絡安全AI助手「Copilot」，應用到公司內置網絡安全系統中。該助手採用AI聊天機械人的語言系統，可同時處理多達1,000個警報，並迅速為用戶提供網絡攻擊情況報告，有助非專業人士操作。

微軟表示，Copilot使用OpenAI全新的「GPT-4」AI語言系統，加快分析潛在網絡安全漏洞的速度。在其協助下，網絡安全負責人可以更快了解黑客攻擊各個部分之間的聯繫，例如不同的可疑電子郵件、惡意軟件和被破壞的系統部分，會否指向同一個「幕後黑手」。Copilot只需接收簡單的英文指令，就能提供完整的網絡安全報告，即使員工並非AI領域專家也非常容易使用。

微軟負責安全、合規、身份和私隱事務副總裁亞凱爾表示，黑客的攻擊速度會愈來愈快，不少員工還缺乏應對網絡攻擊的相關技能，因此適當利用AI助手，對維護網絡安全非常重要。

裁員潮解散AI倫理團隊 演算法偏見問題難修復

香港文匯報訊 愈來愈多科企推出自家人工智能（AI）技術及產品，但負責評估AI相關倫理問題的團隊成員，卻多在科企裁員潮中被解僱。《金融時報》28日報道稱，包括微軟、Meta、Google、亞馬遜和Twitter等公司都削減了AI倫理團隊成員，讓人擔憂日新月異的AI技術缺乏監管，很可能遭到濫用。

◇Google倫理團隊聯合負責人格布魯早前被炒。 網上圖片

加劇虛假資訊傳播

報道指出，微軟今年1月已解散所有AI倫理與社會團隊，約10名員工被解僱。Twitter一個負責修復演算法中種族偏見的AI倫理團隊成員也被全數炒魷。Meta負責評估名下社媒平台道德規範問題的創新團隊同被解散。亞馬遜名下串流平台Twitch上周裁撤AI倫理團隊，據報要由研發團隊兼顧處理演算法偏見問題。

報道指出，「ChatGPT」等AI聊天機械人問世後仍問題不斷，包括不時引述虛假消息或給出邏輯錯誤的答案，如果沒有AI倫理團隊監督，不排除會被用於傳播虛假資訊等。科企Alphabet名下AI公司DeepMind前倫理政策研究員斯特雷特直言，「如此多AI倫理團隊員工被解僱令人震驚，尤其科企現時相較以往任何時候，都需要更多這樣的團隊。」

倫敦國王學院AI研究所主任拉克也認為，科企應設法保留AI倫理團隊，「我們無法完全預測新技術會帶來的東西，認真關注這些問題至關重要。」

「平庸AI」遭惡意利用 或助長恐怖主義

香港文匯報訊 美國紐約大學心理學教授、人工智能（AI）初創企業Geometric Intelligence創辦人馬庫斯聯署了「未來生命研究所」的公開信。馬庫斯解釋，科技界現時並非擔心人類無法控制的「超級AI」問世，反而擔憂技術仍存在缺陷的「平庸AI」（Mediocre AI）過快被大規模部署，不成熟的技術被惡意利用，反而會帶來更大破壞。

模仿特定聊天風格作欺詐

馬庫斯指出，現時的AI聊天機械人利用大型語言模型，幾乎可以訪問海量數據，卻沒有完善的監管機制，若被惡意利用，後果或不堪設想。歐洲刑警組織近日一份報告就提醒，有犯罪團體嘗試利用AI聊天機械人模仿特定聊天風格，用於網絡欺詐誤導受害人。AI技術還或被犯罪集團用於收集可能促進恐怖主義活動的信息，包括籌集資金或分享匿名文件等。

馬庫斯還提醒，部分AI技術可以短期炮製大量宣傳話術，不排除導致極端思想快速蔓延，現時人們應專注思考這些迫在眉睫的風險，關注加強監管，「這種恐怖主義甚至可能引發核戰爭，也許人類不會真的從地球上消失，但事態發展會急轉直下，文明都或會被破壞。」

△▽網民用AI偽造方濟各「時尚照」幾可亂真。 網上圖片

AI生成教宗穿羽絨照 網民當眞激讚「潮流教主」

香港文匯報訊 數張天主教教宗方濟各的「時尚穿搭照」在Twitter上引起熱議，網民紛紛讚賞教宗的品味夠「潮」，也有人覺得教宗這身衣着十足十饒舌歌手。不過真相很快曝光，原來相片全都是AI生成的「假照片」，令網民大吃一驚。

大部分人初看這些相都信以為真，留言「教宗要準備發布新專輯了」、「感覺教宗準備上台給音樂節表演收尾」、「教宗原來是潮流教主，向他的穿搭致敬」、「果真是不一樣的教宗」等。

建築工「Midjourney」生成影像

不過有眼利網民很快發現相片被Twitter官方標上「AI製作的虛假圖像」提醒，紛紛震撼表示「怎麼是假的呢？還想問一下是哪個品牌」、「如果不標註的話根本看不出是假的」、「以後人類鑑別信息真偽的難度提高了不止一個級別」、「以後『沒圖沒真相』這話也不能說了，有影片也可能不是真相」、「AI進步的速度驚人」、「破綻將會越來越少」等等。

據美國新聞網BuzzFeed News報道，照片的作者是31歲的芝加哥建築工人帕布羅，他用AI藝術工具「Midjourney」生成影像，創作只因「覺得看到教宗穿上羽絨很有趣」。

監管談判跑輸技術發展 歐盟AI法案或難產

香港文匯報訊 歐盟早有提出立法監管人工智能（AI）技術，不過伴隨ChatGPT等AI聊天機械人快速興起，歐盟內部就監管事務的談判愈發複雜，原計劃本月底在歐洲議會投票的《人工智能法案》也可能「難產」。據報在上月歐洲議會全體大會上，各成員國議員就因分歧眾多，未能達成任何協議，預計各國議員達成共識的難度將超出預期。

歐盟委員會早在前年就提出《人工智能法案》，旨在加強對AI技術監管，在草案成為正式法律之前，需要經過歐盟委員會、歐盟理事會和歐洲議會的三方會談。主導該法案談判的意大利籍歐洲議會議員貝尼菲披露，想要在未來數月展開三方會談困難重重，「因為法案涉及到許多不同的委員會，談判相當複雜，單是在歐洲議會，你每次就要與約20名議員交談。」

短時間無法判斷ChatGPT等風險

還有報道披露，歐盟原本希望平衡鼓勵創新和保護民眾權益，將不同的AI技術根據風險水平進行分類，要求提供高風險技術的企業保持透明運作。然而ChatGPT迅速冒起，歐盟立法者此前並未過多關注這項技術，短時間無法判斷其風險水平，自然難以確認監管力度。

科技業界估計歐洲議會最快今年年底可就《人工智能法案》達成協議，不過也有擔憂認為法案內容複雜，立法工作恐會拖延至明年。屆時將是歐洲議會選舉年，若關注其他優先事項的議員上任，法案仍會有被擱置的風險。