人工智能聊天機械人 ChatGPT 對整個社會帶來極大的震盪，現在電腦表現出幾乎與人類無法區分的智能行為，科幻小說所描述的噩夢會變成現實嗎？四月初一群著名的人工智慧研究人員簽署了一封公開信，呼籲放緩開發人工智能的步伐。約書亞・本吉奧（Yoshua Bengio ） 是聯署人之一，本吉奧是人工智能領域的泰山北斗，他與楊立昆（Yann LeCun ）、杰弗里・欣頓（Geoffrey Hinton） 合稱為「人工智能三大教父」。

本吉奧表示簽署這封信是應該的，因為我們無法保證在可見將來不會有人製造出危險的自動化人工智能，這系統的行為可能會偏離人類的目標和價值觀，我們需要基於「預防原則」（precautionary principle）去規管和減慢 人工智能的發展。

本吉奧的警告並不新鮮，過去特斯拉總裁馬斯克曾經多次說過類似的話， 2014 年，他指出谷歌「深層思想」（DeepMind ） 的發展程度已經到達危險的程度，他擔心谷歌會製造出毀滅人類的人工智能機械人；2018 年，特斯拉總裁馬斯克曾經說：「人工智能比核子武器更加危險。」 2019 年，美國未來學者馬田・福特 （ Martin Ford ） 出版了一部訪問集，書名是《智慧建築師》，這本書收錄了十多位專家對人工智能發展趨勢的意見，其中一條問題就是要回應馬斯克的警告，接受訪問的專家有不同的意見，有些人認為應該要慎重地面對這個潛在危機，但有些人認為這是杞人憂天。這本書是在 ChatGPT 面世之前出版的，我不知道現在主流意見會否改變方向。馬斯克經常口出狂言，而且人工智能並非他的專長，但現在聯署公開信的人包括了人工智能專家，看來我們不能夠掉以輕心。

一般來說，本吉奧所提出的「預防原則」是應用在公共衛生政策和環境措施方面的，根據這項原則，如果一項行動可能對公眾或生態造成潛在危害，而它是否無害還沒有達到

科學共識，那麼證明其無害的責任要由採取行動的一方承擔。從某個角度來看，「預防原則」與「無罪假定」（Presumed innocent until proven guilty ）是相違背的，基於無罪假定，如果你認為人家的產品或者服務可能對公眾或者環境構成危害， 而是否有害還沒有得到科學證據的確認，那麼舉證責任應該是落在提出指控的一方。

事實上「預防原則」 曾經令不少無辜者受到損害，舉例說，道康寧公司（Dow Corning ） 一直以來都生產隆胸手術所採用的矽膠。但有些做過隆胸手術的女性抱怨在手術之後患上重病，她們懷疑元兇可能是矽膠。基於預防原則，美國聯邦藥物監管局從 1992 年起至 2006 年禁止矽膠隆胸手術，聯邦藥物監管局並沒有證據顯示矽膠不安全，但當局要求道康寧公司去證明這種物料無害。隨後如雪片飛來的法律訴訟迫使這公司申請破產保護， 雖然最後許多醫學研究都顯示出矽膠隆胸手術並不會致癌，或者引發其他致命的疾病，但道康寧公司的聲譽已經一蹶不振。

如果將「預防原則」 應用在人工智能，那麼人工智能的研發者便有責任證明自己不會製造出「魔鬼終結者」（terminator ）、「 梅根」（M3gan）、《2001 太空漫遊》的 HAL9000、《復仇者聯盟》的奧創（Ultron ）。然而，若要實際執行起來，這是相當困難的！一部電腦或者一台機械人需要具有自我意識，方可不理會人類的指令而做出傷害人類的事情。不過，現在心靈哲學（philosopher of mind ）和認知科學（cognitive science ） 對自我意識的結構與來源還未充分理解，更加重要的是，自我意識是主觀的、第一身的，旁觀者很難去驗證一部人工智能電腦是否真的有自我意識。

不過，即使人工智能還未具備自我意識，它也可以傷害人類，這是因為研發者的指令沒有經過周全考慮，結果正如本吉奧所說，人工智能的行動與人類的目標並不一致，這就是所謂「對準難題」（Alignment problem ）， 舉例說，若果聯合國吩咐人工智能尋找出解決人類衝突的方法，從而達致世界和平，那麼人工智能的可能方案就是：地球上再沒有人類！馬斯克用過類似的例子：若果你吩咐人工智能去杜絕垃圾電子郵件，那麼「他」想出來的辦法可能就是消滅所有發出電子郵件的人！這些都是誇張的假設性情況。在真實世界中，2018 年一部自動駕駛汽車在亞利桑那州撞死了一名路人，因為汽車人工智能的設計者沒有考慮到會有人亂過馬路！在這種情況下，「預防原則」 是值得參考的，自駕車還未在市

場中，正是因為未有足夠證據顯示它十分安全。然而，我恐怕無論研究人員怎樣小心，錯漏仍然是難以避免的。事實上，無論傳統汽車設置了幾多安全措施，在路面上仍然有許多交通意外而導致傷亡。絕對無害是沒有可能的，要達到什麼程度的無害才算是通過「預防原則」的考驗呢？

平心而論，本吉奧並沒有叫停人工智能的發展，他只是呼籲不要盲目地向前推進，而是要檢討其潛在風險。我同意，然而，「預防原則」並不是仙丹靈藥，有時候過於謹慎、過多箝制會窒礙了科技和經濟的發展，甚至造成無辜者受到傷害。對研究人員來說，這是一個沉重的舉證責任，無怪乎 一些人工智能團隊招攬了倫理學家、 心理學家、社會學家、法律專家……。

2023 年 4 月 15 日

原載於澳洲《同路人》雜誌

[更多資訊]