

---

## 人工智能會變成魔鬼終結者嗎？

余創豪 [chonghoyu@gmail.com](mailto:chonghoyu@gmail.com)

---

近年來人工智能的發展可說是突飛猛進，很多專家都開始關心，人工智能會否有朝一日達到奇點（singularity），所謂奇點，就是人工智慧發展到具有自我意識，甚至自由意志，或者能力強大到超出了人類可以控制的地步。相信很多科幻小說迷都會記得，在《2001 年太空漫遊》這部經典科幻電影中，具有人工智慧的電腦 HAL 造反，冷血地殺死了太空人，還有《魔鬼終結者》（Terminator）裏面的電腦天網（Skynet）更加毀滅了整个人類文明。

最近前谷歌「X 計畫」首席執行長莫高達（Mo Gawdat）在接受《時代雜誌》的訪問中指出：奇點即將到來！而且這會對人類構成了重大威脅。他認為《魔鬼終結者》的天網可能會惡夢成真，到了那一天，人類會無助地面對著好像上帝一樣的超級機器，為人類帶來了世界末日。莫高達描述自己在谷歌研製機械臂的時候，有一天一個機械人拾起地上的小波，然後將小球拿到研究人員面前，好像是炫耀自己的能耐。莫高達說：「突然之間，我意識到這是非常恐怖的，我完全僵硬起來。.....事實上，我們正在創造神！」

筆者並不完全否定奇點的可能性，也許莫高達有其他強力的證據，顯示出魔鬼終結者並非天方夜譚，但是，如果其推論只是建基於觀察到機械人做出類似人類的行為，那麼這種推論可能是犯了「擬人化」（Anthropomorphism）的謬誤。「擬人化」本身並不一定是謬誤，事實上，文人經常採用擬人法來作為修辭格，例如唐代詩人杜牧曾經在《贈別·其二》中這樣寫：「蠟燭有心還惜別，替人垂淚到天明。」杜甫在《春望》的名句亦是一個好例子：「感時花濺淚，恨別鳥驚心。」詩人和讀者都清楚地知道，蠟燭、花朵並不會哭泣，詩人只是將自己的情緒投射到他們身上。

另一種擬人化是介乎有意和無意之間，例如小孩子會跟自己的玩偶談話，成人和小孩也會將人類的性格特質投射到自己的寵物身上，在理性層面上，我們都知道玩偶和寵物不可能具有人類的思想，然而，在情感上我們卻希望這是真的，在行為上我們亦會將玩偶和寵物以人來看待。倫理學家喬納森·海特（Jonathan Haidt）曾經提出一個有點搞笑的案例：一個家庭飼養的狗不幸死亡，他們決定物盡其用，將狗的屍體煮食，這正是「香肉滾三滾，神仙站不穩！」那麼，他們的行為是否不道德呢？我相信很多人都會批評這個家庭

的做法，因為這樣實在太不尊重生前對這家庭忠心耿耿的小狗，在某程度上這種理據是將狗比擬為人。

但是，若果將機械人擬人的行為當為是有意識，或者是有潛力發展為自我意識，那麼這便算是一種謬誤。大約 20 年前，心理學家蘇珊·菲斯克（Susan Fiske）提出一個理論：通常人們以兩種性格特質去判斷他人：這個人是否有溫情（warmth）？這個人是否有能耐（competence）？有趣的是，現在一些心理學研究發現到，人類亦是採用同樣的標準去判斷具有人工智能的機械人。

現在回頭說莫高達因著機械人在研究人員面前舉起小球而受到驚嚇的事件，按照常理，這個動作可以被解釋為溫情的互動，就好像主人將球擲出之後，小狗將小球銜回到主人面前，主人會哈哈大笑。筆者的想法是：人類對人工智能的感覺比菲斯克提出的理論更加複雜，人類對科技總是懷著混雜的情緒：既愛又怕。事實上，人工智慧可能導致世界末日的說法並不新鮮，在 2008 年位於歐洲的大型哈頓對撞機 (Large Hadron Collider) 正式啟用時，有些人擔心，哈頓對撞機高速撞擊粒子而產生微小黑洞，可能會吞噬整個地球，當然這並沒有發生。在荷里活電影中，關於世界末日的電影可說是數不勝數，比較著名的是《哈米吉多頓》、《2012》、《生化危機》，世界末日的恐懼已經通過流行文化，深深地植入了人們的心底中。

筆者再次聲明，我並不排除人工智能出現自我意識的可能性，我也不會斷言將來不會出現天網這種威脅人類文明的超級電腦，但根據目前的證據來看，這種擔憂並不是建構在強力的事實基礎上，反之，很大部分是心理因素，特別是擬人化的心理傾向，再加上對科技既愛又怕的矛盾心理，還有對世界末日的恐懼。

2021 年 10 月 5 日

原載於《同路人》

[更多資訊](#)