

2026：有關通用人工智能的爭辯

最近不斷又聽到人工智能將會智力超越人類，從而替代人類，甚而統治人類。這當然比上周本欄說的人工智能“代替人類學習”更駭人聽聞。進一步說，人類因此將會滅亡。

這些言論的起點，是美國非常流行的觀點：目前是人工智能，下一步是通用人工智能，也就是 AGI。什麼是 AGI？基本上，AGI 要具有人腦的全部功能。目前的 AI，都是有特定功能的機器設計。像引起全球熱議的中國春晚武術機器人，可以把人體的肢體活動全面模仿；最令人慨嘆的是打筋斗，之後站穩，重心、平衡、筋絡、節奏；慨嘆之餘，這些機器人，卻不會做其他的事情。德國總理也看到了機器手寫毛筆字，的確比人手還要靈活；這雙手卻不一定會做其他的事情。網絡上看到工廠的機器工人，可以撿起不經意丟在地上的東西，也令人慨嘆；不是很有人的智力嗎？但是這機器人，卻並不一定能做其他的任務。電視上還看到球型的機器“人”，已經幫忙警察可以追蹤賊人、進入火場、… 令人贊嘆；但是這樣的機器，卻不能負擔其他任務。

上周提到的腦機接口，就是負責撿起微弱的腦電波，用來驅動電腦；這是比較高級的功能，因為它連接的是人類的智力中樞 - 腦。它能幫助人類彌補缺陷，過近似常人的生活。理論上，它能让那位受益者做任何事情 - 可以通過電腦開車、唱歌、…。它能負擔很多任務 -- 凡是電腦能驅動的。但是到底它是建築在“病人”原來的知識基礎上，並不是為“病人”注入知識。

李飛飛：AI 不是爲了取代人類！

這裏有一個根本的倫理問題：爲什麼要製造替代人類的機器？也就是，發展 AGI，理由何在？這個問題史丹福的李飛飛，早就感到了。李飛飛原來是史丹福大學的電腦科學的教師，2015 年才晉升為教授。她鑽研而發明了 AI 的“視覺”，因此被譽為是“AI 教母”。在 Google 度進修假的時候，做了其中一個負責 AI 的首席科學家。2019 年回來就覺得要成立一個研究 AI 爲人類服務的機構，就拉了當時史丹福的首席副校長 John Etchemendy（哲學家），成立了 Stanford Institute for Human-Centered Artificial Intelligence（史丹福以本人工智能研究院），簡稱 HAI。當時李飛飛就認爲，要從根本上改變 AI 的發展方向，保證它以人文價值和跨學科視野為導向。她的主張，首先覺得當時從事 AI 研究的，都是自然科學家和工程師，她覺得應該有哲學家、社會學家、和人文學科的專家參與。



■ 史丹福大學的電腦科學教師 - 李飛飛

HAI 有三條原則。第一、AI 的發展與研究，必須以對人類個人及社會的影響為依據。第二、AI 的應用，應該以提升和輔助人類能力為目的，而不是取代人類。第三、機器智能，模仿人類智能的矛盾和複雜，必須向腦科學和心理學學習。

HAI 提出今年 2026 年 AI 的任務，一、是從“佈道”轉為“評估”。不再問 AI “能做什麼？”而是問“做得怎樣？有什麼代價？爲誰做的？”二、各國發展各自的 AI 獨立“主權”，擺脫研發商和美國政府。三、HAI 另一個領導人 James

Landay 直言：“今年不會有 ARI！”美國的輿論說，這使一些不切實際的幻想受到了冷卻（tempered）。可見 ARI 是一個熱點。AI 應不應該取代人類，是時候要冷靜頭腦了。

Landay: 今年不會有 ARI!

原來，就在今年，而且是二月與三月，在 AI 的研發者之間，爆發了針鋒相對的大辯論。辯論的一方，以 OpenAI 的 Samuel Altman 為代表，認為 AI 的下一步，就是 ARI，而且就快出現。他相信“放大”（Scaling）是一個方向 — 更多的數據、更大的運算能力，AI 就會變成更大的“腦”，現在要遇到的問題就會迎刃而解。

辯論的另一方，除了上述的 HAI 的 Landay，還有另一位科學家也是企業家吳恩達，講得更直率：“AGI 已失去明確的定義，目前更多被當做行銷術語使用，甚至誤導了公眾。”“AGI 在明確定義下，2026 年不可能實現，真正實現 AGI 可能還需要幾十年甚至更久。”作為 Scaling 的專家，他卻認為 Scaling 碰到了“牆”。



“Scaling 單純依靠獲取資料和增大模型愈來愈難，必須在 Scaling 的基礎上注入其他類型的知識。”這一點與李飛飛的觀點一致，特別需要社會科學和人文學科的參與。

■ 科學家兼企業家 - 吳恩達

引起筆者興趣的是他說：“AI 在持續學習和記憶系統方面與人類仍有巨大差距，目前尚未找到解決這些問題的正確路徑。”吳恩達的話，也許只

說了一半。本欄在 2023 年曾經引用過 Noam Chomsky 的話：“機器的學習靠統計學的關聯性(correlation)，人類的學習靠理解(understanding)。”機器可以運用大量的數據，迅速的運算，通過關聯性，回答人類需要解決的問題；但是機器不需要理解。因此，機器可以根據巨量的心理學家的經驗，做出世界上最佳的心理輔導，但是它不必理解其中的意義。要完成 ARI，用 AI 代替人腦，就要機器也能夠“理解”。這也許是吳恩達的“尚未找到解決這些問題的正確路徑”。筆者等待的另一半話：這些方面，能夠靠數據與運算解決嗎？

吳恩達：與其追求 ARI，不如建構有現實意義的智能體！

我們在“人話”裏經常說的，如“學習”、“理解”。可以有比較明確的描述。例如，“學習”是“是人類對外界事物賦予意義的過程”；又如“理解”可以是“找到因果關係，而且可以舉一反三的過程”；等等。但是要把這些過程化成可供機器運作的數據，就不只是困難，而是不同層面的概念。但是不解決這些問題，就難以要機器替代人腦。

在 Altman 看來，人腦，不外是儲藏一堆信息（數據），讓 AI 這個“餅”做大，總有一天可以超越人腦。且不說其中的倫理問題（應不應該？）不能靠數據與運算解決的，還有很多人類生活中不可或缺的。翻查了一下，例如味道。已經有科學家創造了“人工舌頭”；可以分辨出咖啡與可樂，還可以精準定出“甜度”。但是一無法生成“味道”；二無法定出“太甜”還是“正好”，那是人的主觀感覺（官感），並沒有精準的概念，機器如何模仿？又例如疼痛。已經有科學家通過種種辦法，精準地定義“痛度”，但也一

樣，主觀的感覺，不可能制定精準的量度。不是說要科學家研究如何改進“精準度”，而是這根本是一個“莫須有”的課題。

也有科學家用不同的辦法，測量人類的喜怒哀樂 - 通過人的表情、血壓、腦電波、內分泌、等等。這作為科學研究，無可厚非，但如何讓機器分享？一切主觀範疇的東西，如何要機器掌握？

目前科學家的討論，是技術上的可能性：可能嗎？但是這些討論，卻對於機器的定位，有一個理論上的爭辯：AI 真的可以替代人腦嗎？不是技術上有沒有可能，而是理論上：可以嗎？因此現在輿論上已經開始把 ARI 稱為假設性的（hypothetical）目標。

因此吳恩達說：相比於追求通用智能（AGI），通過構建可靠的 AI 智能體來完成法律、醫療、客服等具體領域的任務更具現實意義。

吳恩達的另一句話，發人深省：“AI 不會取代所有人，但使用 AI 的人會取代不使用 AI 的人。”