

「偏差」的人工智能能否尊重多元化呢？

— 余創豪 chonghoyu@gmail.com —

人工智能偏差是樣本問題

現在批評人工智能的聲音不絕於耳，其中最廣泛被提及的一點是人工智能的偏差，簡單地說，人工智能採用了偏頗的樣本，基於缺乏代表性的樣本去學習世界的事物，人工智能便會生產出狹隘的刻板形象。一個著名的例子就是 OpenAI 早期的圖像生成系統 DALL·E-2，當用戶要求系統製作出律師的圖像時，那麼全部顯示出來的人物都是中年白種男子；若果用戶要求空中小姐的照片，那麼輸出的結果都是年輕貌美的女子。

Stability AI 是一家總部位於倫敦的公司，這公司因為創建了人工智慧美術軟件 Stable Diffusion 而聲名大噪。但該公司亦面臨着與 DALL·E-2 相同的問題，這類機器學習系統主要是建立於西方的數據，如果用戶要求軟件製造出房屋、女人、男人的圖片，他們很可能只會見到西方房屋或白人的圖像。為了解決這個問題，最近 Stability AI 計劃發布一款基於本地化數據的日語人工智能工具，如是者，日本用戶便可以在生成圖片中見到多一點日本文化的元素。

其實，這並不是嶄新的現象，這種偏差無非是過去問題的延續和擴大。2010 年，心理學家亨里奇（Joseph Henrich）、海涅（Steven J. Heine）和諾倫扎揚（Ara Norenzayan）在一篇論文中指出：大多數心理學研究都是基於 W.E.I.R.D 樣本，該縮寫詞是指「白人」（White）、「受教育者」（Educated）、「工業化」（Industrialized）、「富有」（Rich）和「民主」（Democratic）的國家人口。據估計，所有已發表的研究中，有 99% 的資料源自從上述人口招募的參與者。不消說，這些研究的結論會有點偏差，並不一定適用於非西方社會和族群。

在互聯網和大數據時代，理論上 W.E.I.R.D 問題可以減輕，如今不同國家的信息可以輕易地、快捷地在互聯網上傳播，搜索引擎可以超越國界，包羅了源自不同文化、不同族群的資訊。然而，根據 Statista，直到目前為止，世界上 58.8% 的網頁都是用英文寫的。從這個角度來看，人工智能無非反映現實，而不是故意偏差。

此外，人們會倚賴自己熟悉的語言、文化背景去創作內容，這是無可厚非的，甚至可以說，這是自然和理性的。Stability AI 的總部設在英國，該公司的電腦程式編寫員當然會傾向於令系統顯示出西方建築和西方人物。

科幻片中的超級英雄

讓我用電影和電視劇來作為類比，筆者在香港長大，年少時，我很喜歡觀賞日本的科幻劇集，例如〈鹹蛋超人〉、〈幪面超人〉、〈鐵甲萬能俠〉、〈三一萬能俠〉、〈彩虹化身俠〉、〈太空三虎將〉、〈太空西遊記〉……等等。在這些科幻故事裏面，所有超級英雄都是日本人，或者是日本製造的機械人，即使是外星人，其面貌仍然是日本人的模樣，而且所有外星人都說日語。

不用說，在美國的科幻電影裏面，大部份超級英雄都是美國人，而且所有外星人都說英語。1996 年上映的美國科幻片〈天煞〉，敘述美國總統於獨立日帶領全世界打倒了侵略地球的外星人，當時一位中國影評人批評這是大美國主義。

然而，中國拍攝的科幻電影豈不是一樣嗎？例如〈流浪地球〉、〈上海堡壘〉、〈戰地：異種浩劫〉、〈功夫機械俠之南拳〉。上述後三部電影都描述地球軍隊在外星人的侵略下節節敗退，只有中國軍隊仍然能夠頑強抵抗，而最神奇的是〈功夫機械俠之南拳〉，在電影中，中國軍方使用時光機器，將一名機械人送返清朝，目的是去從武術大師身上學習失傳了的南拳，然後將招式輸入晶片。當人類的機械人下載了南拳招式的資料之後，便將外星人打至一敗塗地。我不禁聯想起在武俠電影〈東方不敗之風雲再起〉裏面，東方不敗以葵花寶典的武功抵擋西班牙軍人的槍炮，以中國功夫殲滅了西班牙艦隊。

為什麼機械人要回到清朝學習南拳呢？若果當時的武林高手真是如此了得，為什麼清軍會每戰必敗呢？蒙古人曾經發動三次西征，建立四大汗國，所向披靡。為什麼不將機械人送到元朝，讓他學習蒙古摔跤呢？我猜想原因是蒙古摔跤並不算是中國武術。

通過實踐去檢查真相

我並不是要存心抹黑任何人，我只是想通過這些例子去說明：大多數創作者都會參考自己熟悉的資料和自身的經驗，而且，為了迎合本地市場，他們必須要提供顧客喜歡的東西。

對不起，扯得太遠了！讓我們回到關於人工智能的討論。實驗是檢查真相的標準，到底人工智能圖像生成系統是否真的嚴重偏差呢？我親身在 Midjourney 上檢視過。如果我輸入「一名女子在一間屋子」，如你所料，所有顯示出來的圖片都是一個白人女人和一間西式房屋；不過，如果我特別指明：「一名穿着和服的日本女子在日式建築」，那麼所有圖像都準確顯示我輸入的內容。出乎我意料之外，假如我輸入「一個超級英雄與外星侵略者搏鬥」，所有圖片呈現的英雄並不是典型的白人，而是黑頭髮、膚色較深的中亞面孔；若果我將指令轉為「一個日本的超級英雄與外星侵略者搏鬥」，那麼圖片中的英雄則明顯

是日本人。經過檢測之後，我認為一些所謂人工智能的偏差並不是什麼大不了的事情，其實，只要你將指令說得更清楚（例如「19世紀墨西哥建築」、「中國明朝服飾」、「非洲傳統風格」.....等等），人工智能便能夠輸出多元化的產品。

2023年6月28日

「偏差」的人工智能能否尊重多元化呢？

原載於澳洲《同路人》

[更多資訊](#)

