- **余創豪** chonghoyu@gmail.com

人工智慧倫理成為新興學科

不用我多介紹，許多讀者都已經意識到人工智能已經進入了我們生活中的每一個範疇，由於它無處不在，而且功能強大，故此許多人都關心到人工智能會否被誤用這些倫理問題。 現在人工智慧倫理已經躍升為一門新興的學科，根據ZipRecruiter，人工智慧倫理家的年薪由九萬三千美元至十五萬美元不等，平均數值是十二萬八千美元。關於人工智能倫理的研究報告好像是雨後春筍，筆者閱讀了其中一些研究結果，我贊成一部份觀點，但對另一些卻有所保留。

隱蔽的種族主義和口音偏見

在三月底，英國《衞報》引述了一篇論文，指出「隨著人工智慧工具變得更加聰明，它們的種族主義也變得更加不明顯。」這篇論文是牛津大學、史丹福大學、芝加哥大學等多個研究機構學者合作的成果，其主題是「隱蔽的種族主義」（Overt racism ），他們發現：大型語言模型含有「口音偏見」（dialect prejudice），特別是對於非洲裔美國人的口音，當你對人工智能提出假設性問題的時候，這些系統會對採用黑人英語的個案給予較為負面的評價和判斷，例如會對申請工作的人分配低下的工作，對疑犯會作出有罪的裁決。

筆者同意「口音偏見」這個問題是值得關注的，理想地說，人工智能應該以事論事，而不應該考慮人的口音和文字表達方法。

然而，這是不是一個種族主義的問題呢？或者這種現象的根源是不是種族主義呢？自從人們習慣了在手機發出短訊之後，很多人的寫字方法已經脫離了傳統文化，筆者曾經在某間科技巨頭公司工作，當時一位白人同事寫的英文便好像是短訊一般，我和另一位白人同事暗地裏取笑她的英文，因為她這種行文方法，可能在不知不覺之間我對她的工作能力會打上問號，但撫心自問，我所關注的是她的文字，而不是種族。

誰的標準？

這並不是一個新問題或者新發現，在2021年一位意大利學者和一位美國學者曾經在《語言和語言學指南》（Language and Linguistics Compass ）學報發表了一篇論文，內容也是關於自然語言處理（natural language processing ）系統的偏差。作者指出：「目前的自然語言處理工具沒有考慮到不同族群的口音差異，而是期望所有輸入的語言都遵循訓練資料中的『標準』。但問題是：誰的標準？」 為此之故，作者提倡人工智能系統應該接納多元化的口音。

我同意任何事情都不能夠過度倚賴單一標準，定於一尊，以英語為例，除了英式和美式英文之外，人工智能系統亦應該考慮到澳洲、紐西蘭、加拿大等不同的標準。然而，從技術的角度來看，若果人工智能系統需要兼顧全世界所有不同的口音和寫作方法，那麼成本便會大幅度地提高。當出現了幾十個、甚至幾百個標準的時候，這等於完全沒有標準，到頭來只會帶來混亂和無效率。

語音輸入系統無法辨認我的廣東腔國語

這兩位作者引述了一份發表在2003年的研究論文，指出語音識別系統存在着「強烈的偏差」，當使用者說自己母語的時候，語音識別系統能夠準確地辨認出聲音所表達的意思，若果是外國人，識別的準確程度便會大大減低；即使使用者採用自己的母語，但如果帶有地方口音，亦會削弱了系統的準確度。

但我認為，這是十分正常和自然的事情，當我向語音系統輸入帶有廣東腔調的國語時，通常輸出的文字並非是我所預期的，例如我用國語輸入「人工智能」，輸出來的文字變成了「員工技能」，我說「 豐收」，結果變了「風騷」，「閃電俠」則變了「沈殿霞」，我唸出《道德經》的「無名天地之始，有名萬物之母。」輸出的竟是「無明天地支持，遊民漫步寂寞。」 我說出《論語》的「為仁由己，而由人乎哉？」輸出的竟是：「為人

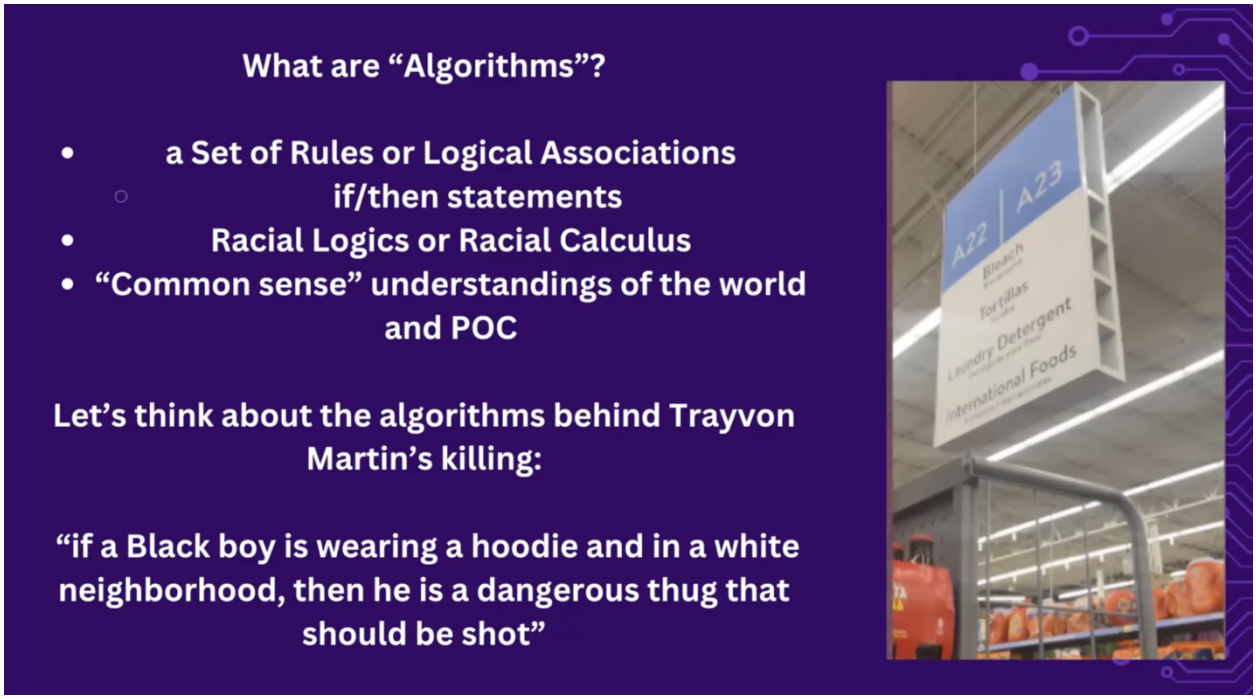
油脂，而油鹽覆蓋哉。」我只能怪自己沒有學好普通話，而不會埋怨人工智慧系統對我有偏見。

我想向大家分享一個有趣的故事：很多年前語音系統仍然在萌芽階段，有一次我的教授用以下的說話對電腦發出指令：「Open the data set。」 通常美國人用「data」這個字的時候，發音是「dat-uh」（爹他 ），他嘗試了很多次，結果電腦都無法明白他的指令。我在香港學習英文，英式的發音是「day-tuh」（地他 ），電腦馬上明白我的語音輸入，並且執行了指令，打開了資料檔案。當時我們只是哈哈大笑，並沒有將問題上綱上線到關於種族和偏見的層次。

批判種族技術理論

美國女性黑人學者蒂拉・坦克斯利（Tiera Tanksley）指出人工智能充滿着歧視和偏見，根據她的「批判種族技術理論」（Critical race technology theory），現今整個數碼資訊系統都為黑人青年帶來不公平和非人性化的學習經驗。她斷言，在程式編碼、資料、演算法、使用者介面的整個技術架構中，反黑人種族主義的成份無所不在。

我同意訓練人工智能是有偏差的，例如最初DALLE.2 顯示出來的律師照片都是白人中年男性，谷歌的人面識別系統曾經將黑人誤認為猿猴，亞馬遜的履歷表審查系統傾向排斥女性。但是，直到目前為止，我沒有發現 drag and drop、cut and paste 等使用者介面具有歧視和偏見的成份。



去年在加州大學歐文（Irvine）分校的一次演講中，坦克斯利舉出一些種族主義邏輯、種族主義數學、種族主義演算法的例子：「如果一個黑人穿著連帽衫並且出現在白人社區，那麼他就是一個危險的暴徒，應該被槍殺。」這個例子是關於一名無辜的黑人少年被槍殺的事件，2012年，一位名叫特雷馮・馬丁（Trayvon Martin）的黑人少年在佛羅里達州被一名西裔人士誤會是壞人而被槍殺。在演說中她舉出了另一個例子：在沃爾瑪百貨公司中她看見一個牌子，上面標示着：漂白水、墨西哥玉米餅、洗衣粉、國際食物。坦克斯利斷言，沃爾瑪將化學物品和外國食物放在一起，這包含了種族歧視的色彩。無可置疑，特雷馮・馬丁槍殺事件是不公義的，但我不肯定沃爾瑪的告示牌是否存心歧視國際食物，這似乎是太過敏感而上綱上線。但無論如何，這些事件跟人工智能和一般數碼科技的演算法有什麼直接關係呢？ 也許她暗示了這些在社會上無處不在的歧視和偏見已經滲入了資訊科技業，包括了人工智能的開發者。但我不相信任何一個電腦程式員會在系統中加入這種邏輯：「如果出現一個可疑的黑人，馬上將他槍殺。」縱使出現這種編碼，公司一定會將之封殺。

身為美國社會的少數民族，我當然支持人工智能系統應該避免歧視和偏見，然而，我擔心現在某些做法會否傾向於上綱上線和矯枉過正呢？

2024年3月23日

原載於澳洲《同路人》雜誌

[更多資訊]