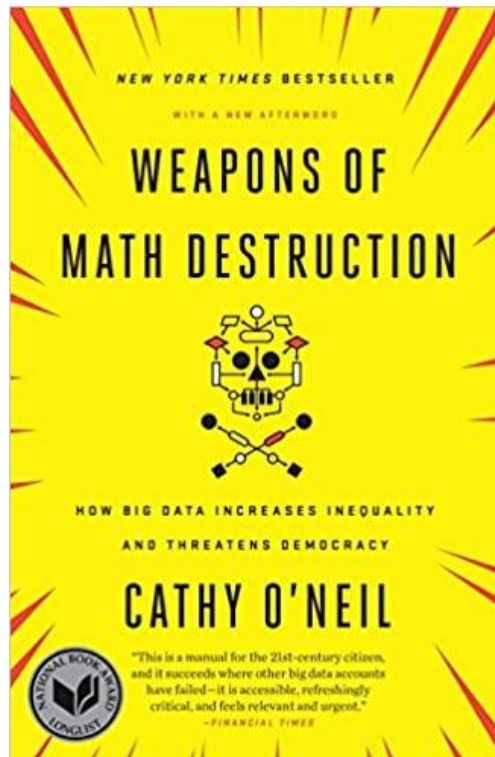

大數據是毀滅性武器嗎？

余創豪 chonghoyu@gmail.com



無可置疑，現在大數據和人工智能已經大幅度地改變了社會的運作模式和人民的生活方式，不過，不少人仍然對這種變遷持保留態度，數學奇才凱茜·奧尼爾（Cathy O'Neil）所著的《數學毀滅性武器：大數據如何增加不平等並威脅民主》

（Weapons of math destruction）就是對大數據提出質疑的經典著作。

在這本書中，奧尼爾提出了許多採用大數據而造成負面影響的例子，例如大學排行榜促使家長和學生拼命地擠進名牌大學，學校利用評分制度將所謂「不稱職」的老師辭退，但往往低分者才是嚴謹的老師，這些都是十分複雜的問題，我將會在其他文章詳細地討論。在這篇文章我只是集中討論奧尼爾對「背景審查」（background check）的批評，所謂「背景審查」，就是香港人所指的「起底」。奧

尼爾指出：許多人在電腦系統中的資料並不齊全或者錯誤，結果不少無辜的人因而失去工作機會，或者無法向銀行借貸，她舉出了一個例子：某君在申請工作時被拒絕，原因是背景審查的資料顯示他曾經犯下刑事罪行，其實他是無辜的，真正犯案的人和他同名同姓，並且同年同月同日出生，但資料庫卻張冠李戴。

筆者的回應是：真正的問題並不在於現代社會過於電腦化和數據化，而是不夠電腦化和數據化！其實，用姓名去識別人是非常落後和容易出錯的方法，天下間同名同姓的人實在太多了，如果我在街上大喊一聲「陳志強」，我相信會有十幾人回應我。故此，熟悉電腦資料分析的人都會採用獨特的身份證明號碼。美國的電腦科技已經十分成熟，而且人工智能之目的正是要講求精準，但為什麼美國的大數據仍然會出現這些基本的錯誤呢？

全世界大部分國家都有國民身份證號碼，而美國是例外，美國公民自由聯盟（ACLU）認為編配身份證號碼是違反人權和侵犯私隱，但他們所舉出的論據，卻令我有

如丈八金剛，摸不著頭腦。他們指出，恐怖分子和罪犯有辦法偽造證件，所以國民身份證並不能起到核實真正身份的作用。按照這種邏輯來推論，護照、駕駛執照、工作證、學生證.....都可以偽造，那麼為何不乾脆取消所有證件？

第二個理據是「滑波理論」（slippery slope theory），政府今天可以用身份證來獲取你的個人資料，明天便可以進一步限制你的自由和侵犯你的私隱。「滑波理論」只是一種激起人恐懼心理的手法，人們幾乎可以用同樣的說法，去支持或者反對任何東西，例如反對槍械管制的人說，今天政府可以收去你的半自動衝鋒槍，明天便會更進一步限制你其他自由；但反對槍械氾濫的人亦可以說，今天政府可以容許你買半自動衝鋒槍，明天便會進一步放鬆，你可以買全自動機關槍、火箭炮、坦克車.....。「滑波理論」是基於這個比喻：一輛沒有煞車掣的車子一旦進入斜坡，便會失控地繼續下滑，直至車毀人亡。但我們並不是沒有煞車掣的汽車！

他們又指出，全國性的身份數據庫令政府有能力監控人民，並且造成對人民的壓迫和歧視。上面提過，世界上大部分國家都有國民身份證，包括了歐洲的民主國家，但並不見得那些歐洲國家因此而變成專制政體，或者人民的自由大幅度地被削弱。

另一方面，保守派基督徒亦反對政府編配身份證號碼，他們認為這就是《啟示錄》所說的獸的印記：666，這種做法令美國更加接近納粹德國。

奧尼爾並沒有主張廢除數據化，她只是呼籲人們不應該盲目地相信自動化電腦系統和大數據，我完全同意，但我認為問題的根本是在於另一種盲目。倘若美國人仍然教條化地相信某些意識形態，以為身份證號碼會造成極權政府，反對建立統一和精準的資料庫，我恐怕奧尼爾所說的「數學毀滅性武器」會繼續做成張冠李戴和其他不同形式的錯誤。最近筆者承接了一個數據分析計劃，我需要將兩個數據集的資料接連起來。如你所料，兩個數據集都沒有獨特的身份證明號碼，唯一的方法是採用姓名來連接，我不敢保證結果是百分之百正確。倘若出錯的時候，我希望不會有人指責我製造「統計學毀滅性武器」！

2021年3月10日

[**更多資訊**](#)