

【講座紀實】

當 AI 成為人的那一天：2019 AI 嘉年華達人論壇紀實

吳睿恩¹

近年來，人工智慧（Artificial Intelligence，AI）研究快速成長，也逐漸應用在生活不同領域中，從廣告推播、理財顧問、自駕車，甚至是醫療及法律問題的解決，AI 的降臨，已成為必然。然而，AI 的發展固然帶來便利，卻也伴隨隱憂——已故的英國天體物理學家霍金（Stephen Hawking）生前便曾警告：人工智慧的發展，可能讓人類走向滅亡²；Tesla 主席、知名企業家馬斯克（Elon Musk）更曾指出：對於 AI 必須格外謹慎小心，其可能比核武更加危險³。

《基礎法學與人權研究通訊》過去亦曾介紹人工智慧的發展歷程與未來趨勢⁴，而在 2019 年 12 月 28 日，國立臺灣大學科技教育發展中心、國立臺灣大學人工智慧中心共同主辦「2019 AI 嘉年華」活動，其中的達人論壇則以「當 AI 成為人的那一天：未來世界的人與社會」為主題，旨在檢視：AI 對於社會、人類可能帶來哪些衝擊與挑戰？而我們又該如何看待與應對？

該場論壇由建置臺灣法實證研究資料庫第五期實施計畫共同主持人、臺灣大學資訊工程學系林守德教授主持，並邀請臺灣大學法律學系黃銘傑特聘教授、臺北科技大學智慧財產權研究所江雅綺副教授、Appier 企業人工智慧應用研發副總裁黃世昭博士、師範大學大眾傳播研究所

¹ 作者為臺灣大學法律研究所公法組一年級學生，臺灣法實證研究資料庫研究助理。

² Rory Cellan-Jones, *Stephen Hawking warns artificial intelligence could end mankind*, BBC NEWS (Dec. 2, 2014), <https://www.bbc.com/news/technology-30290540>

³ Matt McFarland, *Elon Musk: 'With artificial intelligence we are summoning the demon.'*, THE WASHINGTON POST (Oct. 25, 2014), <https://www.washingtonpost.com/news/innovations/wp/2014/10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/>

⁴ 蘇上雅（2018），〈機器-人-明日社會：「人工智慧、心靈與演算法社會」討論會紀實〉，《基礎法學與人權研究通訊》，21 期，頁 22-25，網址：http://tadels.law.ntu.edu.tw/getfile.php?db=edm&file=pdf_file&no=44（最後瀏覽日：2020 年 2 月 25 日）；蘇上雅（2019），〈圖靈沒有料到的 AI 歷史的今日：林守德教授「當人類智慧碰到人工智慧」講座紀實〉，《基礎法學與人權研究通訊》，22 期，頁 3-8，網址：http://tadels.law.ntu.edu.tw/getfile.php?db=edm&file=pdf_file&no=48（最後瀏覽日：2020 年 2 月 25 日）

王維菁教授與清華大學動力機械工程學系丁川康教授，共六位不同領域的學者與業界專家，從不同面向切入，探討 AI 對於倫理、法律與社會所帶來的衝擊與挑戰。

AI 與倫理：誰為危害負責？應否限制 AI 發展？

本場論壇聚焦四大議題，第一個議題是：人工智慧是否會對人類產生危害？對於此一危害，該由誰來負責？是否應該因此限制人工智慧的發展？

黃銘傑教授首先指出，再怎麼聰明的人工智慧，只要會與人類產生互動，仍可能對社會產生負面影響。但這樣的負面影響，一般而言，應非是設計者惡意所致，因此，當 AI 的學習能力越來越強，以至於其所行為、決策已遠遠超出設計者當時的想像時，是否可以要求設計者為 AI 所帶來的危害負責，便成問題。

若是無法要求設計者負責，可以直接對 AI 咎責嗎？黃教授認為，這涉及法律上「人的本質到底是什麼」——具有法人格，才能在法律上享受權利、負擔義務。近年來，固然有論者（如歐盟⁵）開始提倡應賦予 AI 法人格，然而黃教授也提出質疑：若 AI 不具有感性、也沒有自由意志，僅僅因為其可能造成不當影響，就承認其具有法人格，毋寧是一種弔詭的說法。黃教授指出，人之所以遵守法律，是因為違法會受罰，負面的誘因迫使人類遵守法律的行為準則；未來或許可以思考，如何同時賦予 AI 正面與負面的誘因，讓 AI 也想要守法。又或者可以成立 AI 基金、AI 保險，由整體社會一起分攤 AI 可能帶來的負面效果。

王維菁教授則認為，對於 AI 所造成的社會影響，不能流於主觀及想像，而應該建立起嚴謹、科學的制度，來具體評估社會影響。王教授並指出，這樣的影響評估，至少可以包含四個層面：首先，是「利害關係人」的評估，必須考慮 AI 的發展，造成哪些人獲利及哪些人受害；其次，是「系統性偏見與歧視」的評估，若 AI 是奠基於既有的資料與事實來進行運算，將可能合理化既存的偏見與歧視，從而產生「肯認現實」的偏差；再者，是對於「自主性」的影響，當人類越來越依賴 AI 而做決策，是否會因此讓渡決策的權力予機器及程式？又如我們現在在網路社群平台能看到什麼，都是由演算法所決定，人們接收資訊的權利，其實也受到限制；最

⁵ 參見：葉雲卿（2017），〈新型態的法律權利責任主體的誕生——由 2017 年歐洲議會提案看機器人擁有著作權之可能性〉，《北美智權報》，190 期。網址：

http://www.naipo.com/Portals/1/web_tw/Knowledge_Center/Laws/IPNC_170726_0201.htm（最後瀏覽日：2020 年 2 月 25 日）

後，AI 的影響並不是單一存在的，「科技連動性」的影響，也必須被考慮，以工業革命為例，鐵路與都市化的發展、娛樂產業的興起，都是相互關聯的，因此，AI 與其他科技的相互連動，及其對於社會的整體影響，都必須納入評估。

最後，丁川康教授也對上述觀點加以回應。丁教授認為，AI 固然威力強大，但技術本身是中性的，正如同牛頓力學、質能轉換公式也同樣可能帶來巨大的影響，我們也不會因此主張限制其技術的發展。誠然，AI 可能擴大資料裡所潛藏的問題，進而帶來負面的影響，從而在應用上必須謹慎小心，但不宜因此就捨棄、限制 AI 技術的發展。

AI 與隱私：隱私保護與促進研究之難

第二個議題與前一議題密切相關。林守德教授提到，訓練 AI 時，需要大量的資料，但取得資料的同時，卻可能侵犯人的隱私，從而有一種說法是：中國近年來之所以在 AI 技術快速發展，是因為其欠缺嚴謹的隱私保護規範，而使得研究者易於取得各式資料。在隱私保護與促進研究發展之間，應該如何權衡，確實是一個難題。

對此，江雅綺教授進一步提出她的擔心：是否在科技與商業的力量之下，所有的資料將越來越集中在少數領域的菁英手中，而一般大眾卻沒有清楚認知到這樣的發展？不過江教授也表示，自己抱持相對樂觀的態度，雖然目前在保護個人隱私權、人格權，以及促進相關研究發展之間，不同的價值看似有所矛盾、衝突，但這也是一個科技或法律架構發展的契機，屆時，或許人們可以找出一個新的路線，使此一問題不再是兩難的選擇題，而能夠被更好的解決。

黃世昭博士則從業界的角度澄清此一議題。他指出，不管是哪個國家，其實近年來法規都是朝著尊重隱私的方向發展；對於業界而言，所在乎的倒不是法律「會不會太嚴格」，而是法律是否「足夠清楚明確」，及希望法律「不要變動太快」。

AI 與社會：對於未來的影響為何？

第三個議題，是 AI 與人類社會的關係：隨著 AI 逐漸應用在生活中的各個領域，其會如何影響人的未來與社會？王維菁教授從傳播、新聞的領域切入，她認為 AI 對於新聞產業，已經帶來了一些顯著的影響：第一，新聞的生產開始朝向自動化，不論新聞寫作或編輯皆然，儘管

對於科技的角色究竟係「輔助性」或「取代性」仍有爭論，但王教授認為，科技確實會產生新聞人力取代以及失業的問題；第二，資訊與程式工程師漸漸取代記者，過去記者在工作時，會受到新聞專業倫理規範的拘束，而當工程師、程式取代記者時，是否還能夠符合人們對於新聞的想像、價值與要求，毋寧值得觀察。

AI 對於資訊的傳播也會帶來影響。王教授指出，演算法基於使用者自身、所屬群體或互動對象的資訊偏好來進行篩選，結果是造成「資訊偏食」，進而產生意見極化、同溫層的現象。過去，人們所接收的資訊相對全面，還能夠想像不同意見者的想法，但隨著資訊的接收越來越偏食，對話將變得更為困難，也將危及公共領域的存在。

最後，AI 也可能影響人們對於「真實」的認知：社交機器人（social bot）加上帳號控制系統，便可以介入網路資訊的傳播、甚至帶動輿論風向；AI 對於產生虛假內容也越來越在行，而且不只是文本，近來非常流行的 deepfake 技術，甚至能夠生產虛假的影音內容。

而黃銘傑教授則擔心，AI 可能對於民主自由的生活帶來衝擊：AI 透過將人「分類」來進行決策，如假釋聲請、信用卡的身分照會等，但這樣的做法，可能與憲法對於人民自由權、平等權的保障有所扞格；而在 2016 年英國脫歐公投、美國總統大選中，也透過分眾的資訊投放，來影響人們的投票選擇。凡此，都是未來社會可能面臨的挑戰。

當 AI 通過圖靈測試的那天到來……

最後一個議題，呼應本場論壇的主題「當 AI 成為人的那一天」，當 AI 能過通過圖靈測試（Turing test）時，人們將無法分辨與自己對話的他方，究竟是人是 AI；也無法分辨眼前的圖畫、文字作品，是出自人或 AI。這天或許很快到來、或許還要一陣子，當其來臨時，有何需要注意之處？

黃世昭博士認為，很可能在五年內，AI 就能夠通過圖靈測試，但其所產生的影響，則必須視情境而定——以聊天機器人（chatbot）的開發為例，它可以快速解決客戶提出的問題、滿足其需要，就此部分來說，其實是一件好事；但如同先前王維菁教授所指出的，當網路內容由 AI 快速生成，過去一個人花十分鐘才寫得出一篇假新聞，現在 AI 十分鐘可以做出一千篇文章、語氣都不同的假新聞，此時所帶來的影響，可能就需要進一步的評估、甚至介入。

丁川康教授則認為，通過圖靈測試的那天已經到來。丁教授舉自己指導的研究⁶為例，他們藉由 AI 創作蒙德里安 (Piet Cornelies Mondrian) 風格的畫作，並請受試者判斷眼前的兩幅畫，哪一幅是 AI 的創作、哪一幅則是蒙德里安本人所繪——結果顯示，受試者的答對率僅有 48%，比擲銅板亂猜還要糟！因此，AI 所生產的音樂、畫作，人類已經無法判斷、也已經應用在許多領域中。面對這樣的發展，丁教授指出，版權收益的歸屬將是問題：AI 算是創作者嗎？其作品是否也受到智慧財產權規範的保護？若答案為是，收益又應該歸屬於何人？

總結

本場論壇最後，有聽眾詢問：應如何避免 AI 的偏見與歧視？對此，江雅綺教授認為，即使是人為的決策，也充斥著偏見，因此對於 AI 的偏見，比較重要的，應當是建立「演算法的監督機制」，了解 AI 為何之所以做成這樣的決策，進而區辨其中有無偏見與歧視存在。不過，丁川康教授也說明，目前學界普遍的想法，是 AI 的效能與透明性、可解釋性之間存在取捨關係 (trade off) ——若希望 AI 做出很好的決策，往往無法解釋它為何做出這樣的決定；但若使用較典型的方法，一步一腳印，雖然可以知道 AI 的決策過程，但決策結果可能不盡理想。而林守德教授也補充，若 AI 的偏見來自於資料的偏差，確實需要更大量的正確資料，才能把問題矯正回來，而這相當困難；因此，晚近許多研究開始轉向，改由演算法或機器學習的方法本身修正，直接改善偏見的問題。在可見的將來，此一眾人關注的問題，或許將能逐漸被解決。

總結本次論壇內容，六位專家均認為 AI 對於人類社會的影響已是進行式，在帶來便利同時卻也伴隨隱憂；而對於是否要透過法律加以限制、規範 AI 的發展，大家似乎則抱持比較保留的態度，也未見比較具體的共識。或許正如同上述偏見問題所揭示的，在持續關注、監督 AI 發展的同時，人們也期待科技可以自行修正、改善問題，屆時，我們不再需要在兩難的價值中作出權衡，而能夠找到兼顧並行的第三條路。

⁶ 吳家榮 (2015), 《以遺傳演算法進行蒙德里安風格創作之研究》, 國立中正大學資訊工程研究所碩士論文, <https://hdl.handle.net/11296/y4p9sb>。