人工的智慧、信任與情感

王存國*

一、智慧系統的發展

在電腦發展的初期，電腦科學家對於「人工智慧」(Artificial Intelligence, AI) 充滿願景與期望，希望能夠發展出有模仿人類能力的人工智慧機器人，然而人工智慧的開展並不如預期。數十年來雖然發展出許多不同的技術支持人工智慧的持續發展，但由於人類擁有多方面與隨機應變的能力，無法單純地用程式語言和機械複製替代。遲至最近幾年方有較突破性的發展，而這些突破性的發展主要得利於電腦運算速度的加快（特別是利用繪圖晶片的平行運算能力）、大量資料的可取得性，以及演算法發展，容許系統在特定資料結構中快速的搜尋、運算，進行判斷，進而產生擬人的判斷思維。這樣的客觀條件發展，也造成需要大量運算的「人工類神經網路」(Artificial Neural Network) 成為一枝獨秀的人工智慧技術流派，也讓產官學界對人工智慧的應用重新燃起熱情，造成近年人工智慧風起雲湧的投資與發展。

當然在人工智慧重新成為電腦相關領域的顯學之前，個人與組織決策層級的技術和系統仍持續發展。當人們瞭解到電腦系統難以如人類有諸多方面及隨機應變的能力時，就將專注力集中在特定專業領域發展「專家系統」(Expert Systems)，冀望藉由系統達到快速複製並取代專家的效果。然而專家之所以為專家，在做決策判斷時，經驗往往比專業知識更為關鍵。如此難以言傳的推論、決策判斷過程，不但很難擷取也難以表達於系統，因此成就有限。

既然無法發展系統取代人類專家，把期望降低，轉而發展協助人類決策者分析資訊、協助判斷的系統也是合理的發展方向，因而產生出「決策支援系統」(Decision Support Systems)。對決策者而言，這樣的系統需要提供有效的資料分析工具，乃至模型，以協助 (assisting) 或擴增 (augmenting) 決策者的能力。然而，由於許多管理決策具有浮現、複雜與高度不確定的特性，這些特性讓決策

* 國立中央大學資訊管理學系講座教授

支援系統的發展也難有重大突破。縱使實務上一些由決策支援系統衍生出來的「高階主管資訊系統」（Executive Information Systems）和「群體決策支援系統」（Group Decision Support Systems），除了產生一些學術的成果外，對業界仍少有實質的助益。

在無法對專家和管理者的決策形成有充分的理解，並據以提供協助或甚至代行下，如能協助組織將相關的資料匯集、整理，再利用演算法從資料中挖掘出一些資料項間的關聯性（associations），以產生資料驅動（data-driven）的預測結果，或仍可有助於組織決策，這就形成近幾年「資料探勘」（data mining）的蓬勃發展。這樣的發展一定程度逆轉了學術界重解釋輕預測的傳統，也迎合實務界的實用主義及以試誤（trial and error）解決問題的傾向，同時也顯示了從人工「智慧」到專家「知識」，再到決策「資訊」，最終到最底層操弄「資料」的「發展」路徑。但在網際網路的發展，以及資料相關的成本變得微不足道下，特別是當資料大到取得來源已接近分析的母體時，資料驅動預測所可能產生的概化偏誤似乎也就不是太大的問題了，導致大數據分析（big data analytics）與資料科學（data science）的蓬勃發展。但這些科技快速發展是否就能解決組織內各層級與不同面向的作業與決策問題？前述的各種協助、擴增、自動化的基本問題，以及目前產業在數位轉型的困難，是否能在人工智慧的發展下得到相當程度的解決？

不少學者對於這些相關的問題提出他們的看法與分析（Berente et al., 2021; Brock & von Wangenheim, 2019; Gregory et al., 2021; Murray et al., 2021; Raisch & Krakowski, 2021）。無論學者們的論述如何，組織本身就是個社會技術系統（a socio-technical system），在其中人類與各種科技、設備互動、一起運作以達到組織的目標。這就如同歐洲 High-level Expert Group（HLEG）on AI 認為探討人工智慧系統應該涵蓋 “all actors and processes that are part of the system’s socio-technical context through its entire life cycle”（HLEG, 2022, p.5），這樣的認知自然產生了各種需要持續研討的問題。本文所探討的人工智慧系統限制在狹義的人工智慧（narrow AI），這種系統的分析智能被運用在特定的功能與應用上，而非具有近似人類認知能力的「人工一般智慧」（Artificial General Intelligence）。

二、人機互動與演化

一般來說，我們不稱人機協同合作（collaboration），而只說是「互動、一起運作」，主因在於學者經常提出的意向性（intentionality）。人類本身可以設定目

標，因而有達成目標的意向與動力，但人工智慧自身沒有意向，其對特定目標的能動性（agency）仍得來自系統的設計者或使用者。既然人工智慧沒有意向性，其本身自然也無法承擔責任（Raisch & Krokowski, 2021），這也衍生出在設計人工智慧時一些似乎無解的倫理或意識形態的議題，如自駕系統設計與篩選徵聘員工等。或許如 Raisch & Krokowski（2021）主張，人工智慧在組織的應用會持續的演化，特別是複雜的管理問題與決策程序會經由人機持續彼此的學習，形成人類能力擴增與決策被自動化交替出現。例如，依據「世界汽車工程師協會」（SAE）對於自駕車分級標準就從駕駛輔助到完全自動化分為五級，目前量產汽車多只達到第二級。但人機共同運作中，機器的部分會隨科技進步而逐漸增加，但在何情境、技術進步到何種程度下，哪些決策可交給人工智慧，哪些決策人類須保有最終的決策權，又有哪些理論可用來分析這些議題，這些在科技持續進步下需要更多、更廣泛的研究。

三、機器的智慧來自於學習

相較於早期的專家系統和決策支援系統，人工智慧擁有之前系統沒有的學習能力（Berente et al., 2021）。目前人工智慧藉由快速搜尋、運算與處理大量資料的能力，以機器學習方而言，無論是運用監督式（supervised）或非監督式（unsupervised）的學習，理想上都應可以幫助人類打破仰賴鄰近搜尋（myopic search）與有限理性（bounded rationality）（Murray et al., 2021），或最基本決策時間上的限制，而能幫助人類做出較佳的決策。而人類從系統所得的結果，如能與其原有之經驗與直覺結合，強化對決策問題的理解，再更進一步地將之回饋給機器學習，而能達到彼此強化的效果。監督式學習的效果取決於機器學習用的標注訓練資料（training data）的品質，畢竟監督式學習在本質上就如同將一個統計模型盡量地與樣本資料擬合，因而可能產生過適（overfitting）或樣本相依（sample dependent）的問題，造成機器的學習只能反應所學習的資料（包括資料可能包含的專屬性雜訊）而無法概化，甚至產生偏誤。因此，以實務上複雜、動態又具高度不確定性的管理問題或商業判斷而言，如何決定監督式機器學習所需的資料與如何標注資料、評估訓練資料的周延性與精確性，以及建構有效訓練資料的成本，在學理與實務上都需要更多的探討。

非監督式學習的系統會依照設計者提供的規則自我（深度）學習，例如 AlphaGo Zero 就是依據設計者所提供的圍棋規則，藉由自己與自己對弈來學習。當然，非監督式學習有許多不同的模型與演算法，但相較監督式學習，其

主要的優點之一在於可大量減少或完全避免標注大量訓練資料的要求。特別是當問題相關的資料可以被大量取得時，更進一步的深度學習（deep learning）或強化學習（reinforcement learning）的演算法就有更大的發揮空間，可以處理更複雜的問題，例如多媒體與自然語言的處理（Berente et al., 2021）。但當所需要搜尋的資料空間極大，問題的複雜性極高時，非監督式學習就像是個黑盒子，如何得到特定的結果往往令人無法理解，自然也無法產生有意義的解讀。縱使實務上這樣的系統或能有助於組織解決問題，但如只是知其然而不知其所以然，仍會令實用導向的管理者產生疑慮，無法對人工智慧有充分的信任（trust），因此研究發展能解釋的人工智慧（explainable AI），以及能學習、理解如何學習的系統，自然是人工智慧接下來持續重要的研究課題。

四、人工智慧的可被信任或只是可靠的

對人工智慧信任相關的文獻已有不少，Glikson & Woolley（2020）對相關實徵研究的回顧，將人工智慧分為機器人（robot）、虛擬（virtual）和內嵌（embedded）三種形式，分析比較人類對這三種形式人工智慧在認知性信任與情感性信任的差異性。雖然擬人的機器人較易產生情感性信任，但要對人工智慧產生認知性信任，則需要人工智慧展現確實性（tangibility）、透明度（transparency）、可靠度（reliability）、即時性（immediacy）等特性。換言之，人工智慧不應是個不透明的黑盒子，可解釋性（explainability）與發現偏見（bias）並加以消弭，對於建立使用者對人工智慧的信任（human-AI trust）或許就相當重要。然而，信任本身就是個複雜的概念，有不同的類型與理論（Ferrario et al., 2020; Ryan, 2020）。另一方面，由於人工智慧本身並無能對其決策或行動負責，一些學者因此認為探討對人工智慧本身的信任或可信任性（trustworthiness）並無意義，至多探討人工智慧的可靠性（reliability），而應把信任議題只放在人工智慧的使用者與組織上（Ryan, 2020），如此議題自然也就應涵蓋到決策影響者、人工智慧的使用者以及技術提供者之間是否能有足夠的人際間信任（interpersonal trust）。例如，IBM 作為人工智慧的開發與使用者，對人工智慧的發展就環繞著可解釋性、公平性（fairness）與可追溯性（traceability）的三項核心原則（Rossi, 2018）。而歐洲對可信賴人工智慧（trustworthy AI）的要求也更為詳細，其認為從人類的監督、技術的穩定與安全性、資料的隱私與治理、透明度、多樣與公平性、社會和環境的福祉，到當責性，都應被廣為考慮（Jacovi et al., 2021）。

人工智慧系統依設計者設定的模型與演算法運作，這或也反應了設計者的意識形態與倫理觀，這對組織應用人工智慧系統在公平、代理、治理等問題上會有何影響，仍須隨著科技的演進與應用的普及而持續地被關注。雖然 Siau & Wang（2018）簡要地提出建立對人工智慧產生初始信任的因素，以及如何持續維繫信任的作法，但這些建議仍相當基礎與片面，因此在系統、開發者與使用者三方面，甚至產業與社會面的信任問題，仍需要更多的研究。

五、服務系統與情感性人工智慧

隨著服務業的發展，以及人工智慧處理自然語言的進步，人工智慧在與人類互動、問題解決、服務提供上也持續進步，因而可被較普遍地應用在與人類有高接觸（high contact）的工作。除了一些電商與客服等已相當普遍的應用，也有學者主張服務業與服務系統的益形重要形成了所謂的感覺經濟（feeling economy）（Huang et al., 2019），而人工智慧協助人類從事與感覺（feeling）相關工作的任務需求也就會越來越多，這在製造產業逐漸自動化下也更為明顯。學者也據以提出應發展具有感覺智慧（feeling intelligence）的人工智慧，或所謂的情感性人工智慧（Emotional AI）。文獻中探討情感性人工智慧已相當多，但無論是何種類型的人工智慧，當它做出自動化的機械性動作，分析問題提出對策，或與人類進行似乎有情感性的互動，其背後仍必須有一個「認知引擎」（cognitive engine）來驅動。因此，這也將產生與信任類似的議題，如人工智慧是否能跟人類一樣擁有感覺、情感、同理心，其行為的背後是否有感情的驅動，是否能（學習）與人類建立起關係，還是單純地依設計做出特定的行為或反應？至少到目前為止答案應該無疑是後者，更不用說情感的基礎往往在於互信，這讓問題變得更為困難與複雜，無法單以技術解決。

六、人工智慧與管理工作

人工智慧除了取代人類從事一些較為例行的作業，以及資料處理分析的認知性工作外，也對管理工作與管理者所需的知能產生影響。人工智慧能讓管理者從耗時最多的協調與控制性工作中跳脫，以有更多的時間依據對組織歷史與文化的瞭解，從事判斷、創新思考、策略發展等較高價值的管理任務（Kolbjørnsrud et al., 2016）。管理工作者應將人工智慧視為能擴增自身能力的同

儕，讓自己的決策做的更好、使自己有更多有意義的創新想法，而不是會取代自己的威脅。當然這又回到人工智慧可信任或可靠的問題，畢竟管理者仍須對決策或行動承擔最終的責任。管理者要能有效地運用、掌握數位科技，自然需要隨著科技的發展，持續強化相關的知識與技能。而發展、維繫關係仍是人類的強項，縱有學者提出所謂的情感性人工智慧，但目前至多侷限在人機互動與問題解決的層面上。在可預見的將來，人工智慧應難有能與人類匹敵的人際關係能力，恐也難與人類發展出共同意向性以有效協作，管理者因此應持續發展人際相關的技能，以充分發揮人類這方面的優勢能力。

人工智慧的導入，以技術與資料持續的驅動對管理的變革。組織的運作包含大量、複雜的作業流程，組織內特定作業一部分的人工智能化，也會對該作業其他的部分，甚至其他的作業流程產生影響，造成組織運作持續地調整、演化（Murray et al., 2021）。這樣的議題在近期數位轉型的風潮下益形重要，需要時時地被關注。電腦運算速度仍不斷地加快、資料量仍大量地累積、演算法也持續地推陳出新，這場人工智慧的復興大業，還只是個開端，人工智慧對人類在各種層面的可能影響力還難以評估。但是，當產官學都全力投入人工智慧技術與應用發展的同時，我們期待，政府與相關單位對人工智慧的管制與治理上，也能投入相對等的關注與資源。

參考文獻

Berente, N., B. Gu, J. Recker, & R. Santhanam. (2021). Managing Artificial Intelligence. *MIS Quarterly,* 45(3), 1433-1450.

Brock, J.K-U., & F. von Wangenheim. (2019). Demystifying AI: What Digital Transformation Leaders Can Teach You about Realistic Artificial Intelligence. *California Management Review,* 61(4), 110-134.

Ferrario, A., M. Loi, & E. Vigano. (2020). In AI We Trust Incrementally: A Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions. *Philosophy & Technology,* 33, 523-539.

Glikson, E., & A.W. Woolley. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals,* 14(2), 598-626.

Gregory, R.W., O. Henfridsson, E. Kaganer, & S.H. Kyriakou. (2021). The Role of Artificial Intelligence and Data Network Effects for Creating User Value. *Academy of Management Review,* 46(3), 534-551.

Huang, M.H., R. Rust, & V. Maksimovic. (2019). The Feeling Economy: Managing in the Next Generation of Artificial Intelligence (AI). *California Management Review,* 61(4), 43-65.

Jacovi, A., A. Marasović, T. Miller, & Y. Goldberg. (2021). Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,* March 2021, 624-635.

Kolbjørnsrud, V., R. Amico, & R.J. Thomas. (2016). How Artificial Intelligence Will Redefine Management. *Harvard Business Review,* Reprint H0380Z.

Murray, A., J. Rhymer, & D.G. Sirmon. (2021). Humans and Technology: Forms of Conjoined Agency in Organizations. *Academy of Management Review,* 46(3), 552-571.

Raisch, S., & S. Krakowski. (2021). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *Academy of Management Review,* 46(1), 192-210.

Rossi, F. (2018). Building Trust in Artificial Intelligence. *Journal International Affair,* 71(1), 127-133.

Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics,* Published online: June 10, 2020.

Siau, K., & W. Wang. (2018). Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal,* 31(2), 47-53.