# 傾聽其他AI的心聲

highscope.ch.ntu.edu.tw/wordpress/

**傾聽其他AI的心聲**
**臺大資工系 黃柏瑋 編譯**

在與Siri或Google Assistant對話時，是否經常因為答非所問而燃起無名火呢？究竟人機之間，為何始終無法像人與人之間的互動，流暢而自然呢？部分原因，在於AI無法理解他者。是否能準確接收指令，並精確做出反應或提供相應的服務，並非一味精進演算法、擴大資料庫與加強訓練可以單方面解決的。一個真正的智能系統，必須能夠意識到周遭環境中其他智能體的存在，甚至理解其思維，並做出預測。至少在人工智能的相關領域研究中，「反求諸己」這項處世之道，或許不是放諸四海皆準的鐵則。

## 他人的錯誤信念

心智理論（Theory of mind，ToM），泛指能夠理解自己以及周圍其他人，包括情緒、信念（belief）、意圖、認知等心理狀態的能力。人類幼童在四至五歲時，即發展出相當的ToM，能意識到他人信念可能與現實有所落差。

Sally-Anne測試，是心理學家常用以評估一個體是否已發展出ToM的工具：情境中的兩位要角─Sally和Anne─同時坐在一個籃子與一個箱子的旁邊，Sally將一顆球放入籃子後離開現場，接著Anne在Sally不知情的情況下，暗地裡把球從籃子移至箱子中。研究人員接著詢問全程在場旁觀的受試者：當Sally回來時，會最先伸手到籃子或箱子裡找球？

若受試者回答「籃子」，則代表其能充分理解Sally因為沒有目擊Anne的動作，而產生與事實（球已被移到箱子裡）不符的錯誤信念（認為球在籃子裡），具備ToM；反之，則代表其無法站在Sally的立場思考，是否具備ToM有待商榷。

## 機器人如何理解機器人

AI的類神經網路極其複雜，往往超乎人類所能夠理解的範圍，好比黑盒子般令人難以參透，使得人們無法確實掌握背後的運作機制，以至於產生不信任感。面對這樣的情況，有些專家嘗試藉由開發另一種AI，幫助人類理解AI的運作。DeepMind研究人員Neil Rabinowitz及同事，便是其中之一。受到ToM的啟發，Rabinowitz設計出一名為「ToMnet」的AI。

ToMnet由三層神經網路所構成：第一層網路，依據過去行為預測其他AI未來的行為傾向；第二層網路，理解其他AI現有的信念；第三層網路，則依據前兩組網路的結果與當下情境，預測其他AI接下來的舉措。

實驗中，研究人員安排其他三個AI在虛擬空間中四處收集彩色盒子以獲取積分，而ToMnet則在空間上方俯視所有角色的行動。三個AI分別有各自的角色設定與「性格」，分別是：看不見周遭環境的「盲人」、記不得自己最近做過什麼的「健忘者」以及視力與記憶力無礙的「正常人」。盲人傾向沿著牆前進；健忘者總是什麼東西靠得最近便抓什麼；正常人則會運用策略，並規劃路線以獲得高分。

ToMnet在經過訓練後，便能依據觀察結果，預測不同角色未來的行為。此外，ToMnet也能通過經典的Sally-Anne測驗，意識到其他AI可能抱持著錯誤信念（false belief），例如：當研究人員將新角色─「近視者」加入虛擬空間中，在其清晰視野之外的空間被改變後，ToMnet精準預測近視者接下來不會如正常人一般，改變原有的行進方向。

## 福兮？禍兮？

其實，AI在「情緒辨識」方面已經有不小收穫，例如：許多臉部辨識軟體已可偵測出人類極細微的臉部表情、分辨對方是否誠懇，甚至如性向、政治立場、智商等私人訊息；「語音辨識」軟體，亦可由音調及語氣，推測人類當下的情緒，觀察之細膩甚至遠超過人類對自身的了解。此外，透過好友名單、網頁瀏覽紀錄與聊天紀錄等大數據，一個人的個性，在AI過人的運算與統計分析能力之前，已然無所遁形。諸如此類，早已為Facebook、Youtube等娛樂與社群網站所用，分析大眾喜好，推出更受歡迎的操作介面或作品。

當然，這些應用並非總是充滿市儈氣息。例如麻省理工媒體實驗室Affectiva，嘗試在車廂內部加裝AI情緒辨識系統，以紅外線和 RGB 相機追蹤駕駛臉部表情與頭部位置，監督駕駛精神與心理狀態，令AI能適時介入，取代人類駕駛。 然而，水能載舟，亦能覆舟。或許更令人坐立不安的是：當人工智慧學會理解其他AI的心智後，是否會進一步解析人類的心理，甚至操控人類的喜好與判斷？但就像知名駭客Kevin David Mitnick曾說過：「我對人工智慧沒有具體立場，因為它本身是中立的。它最後到底是好是壞，取決於人類的應用。」

**編譯來源**

M. Hutson, "Artificial intelligence has learned to probe the minds of other computers", Science|AAAS, 2018.

**參考資料**


1. Pete Etchells, “The Sally Anne task: a psychological experiment for a post-truth era”, The Guardian, 2017.
2. Mikko Alasaarela, “The Rise of Emotionally Intelligent AI”, Median, 2017.