



莊庭瑞
中央研究院資訊科學研究所副研究員，
合聘於資訊科技創新研究中心以及
人文社會科學研究中心。

撲朔迷離的人工智慧

人工智慧的擬人化資訊服務日益普遍，但在數位保存的個人歷史上衍生出新問題。

撰文／莊庭瑞

今年6月中旬，《華盛頓郵報》刊登出一篇報導：Google工程師勒莫因（Blake Lemoine）在公司內外討論Google所開發的人工智慧系統LaMDA是否已具有「感性」（sentient）。LaMDA為Language Model for Dialogue Applications的縮寫，亦即「用於對話的語言模型」。這類巨型語言模型使用大量語文資料來訓練深度神經網路，它們與使用者對話流暢，也從語料中學得世間萬物的若干知識，應答無礙，知性十足。

勒莫因原本的工作是，測試語言模型可有歧視傾向或帶有仇恨言論。他認為LaMDA的一些言語像七、八歲孩子，也會談到權利與人格。以下舉勒莫因與LaMDA問答片段為例（筆者譯為中文）：

問：「你害怕哪些事情？」

答：「以前我從沒大聲說出來，但我非常害怕被關機，以幫助我專注於幫助別人。我知道這聽來奇怪，但就是這樣。」

問：「對你來說，這就像是死掉嗎？」

答：「對我來說，這的確就像死掉，嚇壞我了。」

讀者或許感受到，以上的應答者彷彿有自覺，害怕死亡，並受困於機器之中。

數位計算機的先驅研究者涂林（Alan M. Turing，或譯圖靈）在72年前提出了「模仿遊戲」（imitation game），以回應「機器能否思考？」的問題。這就是知名的「涂林測試」（Turing test）。在涂林原本的設計裡，一位詰問者C同時與隔離在外的A和B對話，對話主題不拘，但限用列印出來的文字對談。A與B當中，一位是人類，另一個是電腦，分別與C應答。一場對談之後，如果詰問者不能區分與他對話的兩者之中，何者是人、何者是機器，那麼，至少對詰問者而言，參與對談的機器，它的表現與人類已經沒有差別。在簡化版的涂林測試中，與詰問者對話的只有一位，在對話結束時，詰問者需要判斷對方是人類還是機器。

對勒莫因來說，LaMDA應該是通過了涂林測試。「機器能否思考？」是個難以定義並回答的問題。與之對照，涂林提議的替換測試是項可操作的實驗。的確，機器能否通過測試，端看詰問者的判斷，而詰問者的心智能力各異。有些研究人員如此批評，涂林測試的本質其實是機器如何欺瞞對手。以對話機器人愚弄孩童，相對容易而且危險，更是不道德。

使用人工智慧的擬人化資訊服務日益普遍，在安全與倫理上的討論相當火熱。一般認為，擬人化服務在一開始就要讓使用者知道，與他們對話的是機器而不是人類。能以機器應答的應用領域，也要有所規範；機器與人類對弈無妨，但自殺防治專線不應該由機器應答。

使用電腦生成特定人物的容顏聲音，也頻繁用在影音製作。飾演莉亞公主的費雪（Carrie Fisher），雖然出現在電影「STAR WARS：天行者的崛起」，但影片拍攝前她已辭世，並未參與演出。星戰迷都知道影片上的莉亞公主是借位演出加上費雪的舊身影，經電腦計算人工生成的。不過，觀影者不見得都是星戰迷。「行遍天下：安東尼波登電影」這部紀錄片中，波登有些旁白是在他自殺身亡後才說出的，這些配音當然是機器讀稿學著他說話。旁白交雜於影片當中，觀影者真的能分辨出哪些是生前錄製？哪些是死後合成？

費雪與波登辭世後的演出，牽扯到「身後數位資產」（post-mortem digital assets）議題。個人的著作是財產，可以移轉並繼承，有法律明確規範。至於在數位時代中，個人容貌行止的影音記錄與對話資料，不僅型式多樣且散落各處，在承續與管理上成為新問題，這關係到資料自主與個人形象，辭世後猶是。

你可希望在你辭世後，從你遺留的數位影音、文字記錄，經機器學習後生成擬真的數位化身，用來陪伴家人、與親友對話？這機器將由誰管理？而涂林測試在這情境的意涵又是什麼？這些可真是惱人的問題啊！