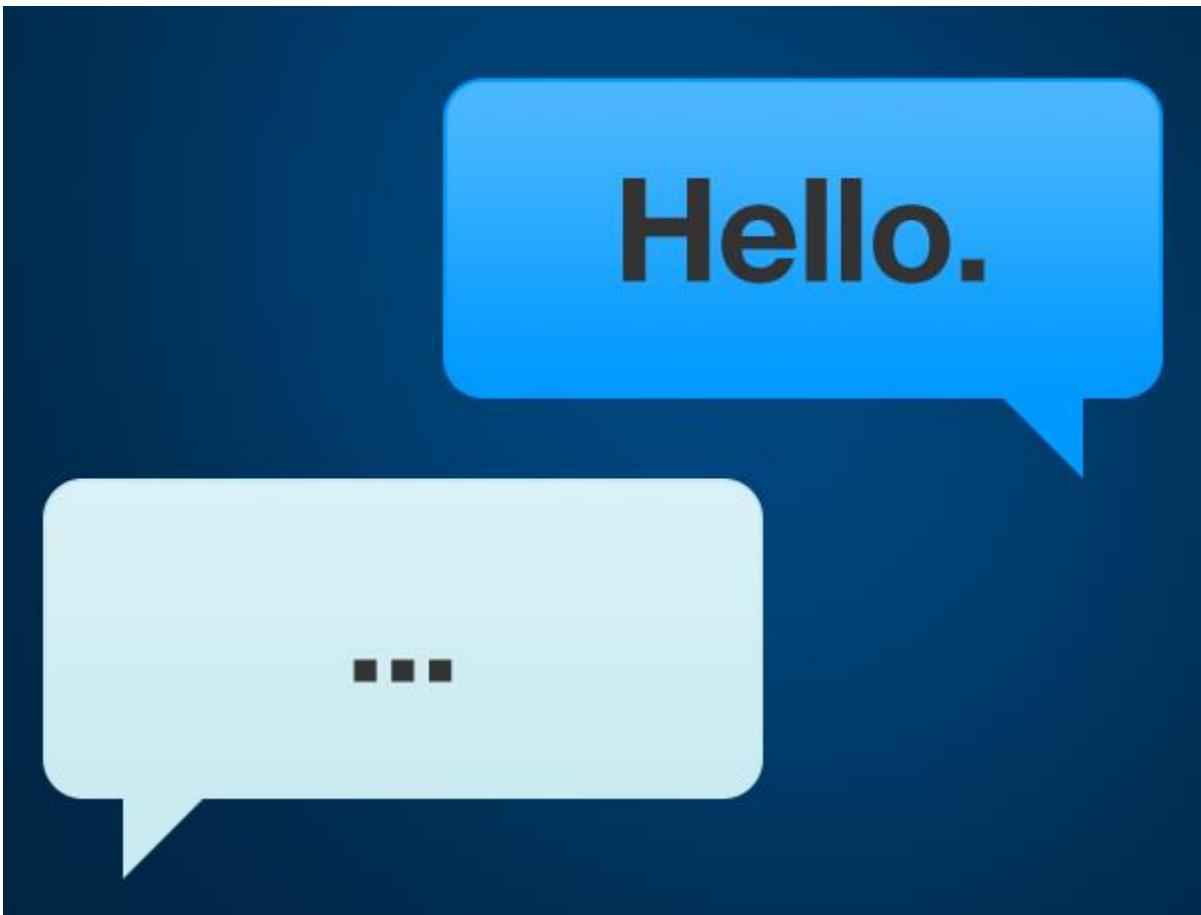


聊天機器人：音訊、人工智慧與機器學習



作者：[Paul Golata](#)，[Mouser Electronics](#)（貿澤電子）

我其中一個女兒在今年聖誕節找到一款叫 Hearing Things™ 的遊戲，我們全家玩得很開心，這款是很紅的「Whisper Challenge」遊戲的改版，玩法是：一人戴上耳機，耳機除了讓他聽不到，也會發出隨機的噪音。另一人面對著戴耳機的人，讀出卡片上的隨機句子，像是「鴿子喜歡依偎在一起」。然後，在沒有任何提示或線索下，戴耳機的玩家只能用讀唇語的方式來猜出句子。

就跟電腦一樣，嘗試破解句子的玩家要尋找線索，雖然這些線索可能不是很明顯。玩家必須用眼睛「聽」，並解密對方的唇語。這就是好玩的地方（沒錯！），因為透過眼睛提示聽到的內容與真正的答案可能截然不同，這類遊戲利用人類天生無法將聽覺與視覺搭配在一起的限制。

那麼，電腦能表現得更好嗎？現在的聊天機器人已有長足的發展，這些應用程式採用了音訊技術、人工智慧 (AI) 和機器學習，這些技術結合後，產生出與人類在對話時類似的理解和回應能力。尤其，自然語言編程和神經網路兩個領域的技術發展經融合後，創造出動態的人機互動，對公司和終端使用者來說有相當大潛力。

技術融合

雖然人工智慧、機器學習和音訊這些技術到目前為止已有長足的發展，卻是直到最近才經過融合，創造出擬人般的人機互動。

人工智慧與自然語言編程

人工智慧 (AI) 正在幫助改變電腦運算的價值，透過可採取行動的見解，從自動化和擴充流程轉化為知識。經由解譯程式和集合流程，公司可透過人工智慧洞悉市場先機、產生新的價值、快速部署有力的決策。

從音訊來看，人工智慧可在許多方面幫助人類。這是因為許多我們所使用的人類語言和溝通要素都可編程到機器內，利用自然語言處理 (NLP) 演算法，經由電腦運算來處理複雜的模式辨識架構。目前設計人員正努力讓機器能以雙向的方式來使用自然人類語言，例如：

- 機器聽到人類語言，機器解讀人類語言
- 機器解讀人類語言，機器回應人類語言

編程雙向通訊的其中一項挑戰，就是確保演算法經過格式化和排列，能夠理解周遭環境，做出適當的回應。我們稱之為「框架問題」，人類確保電腦得到所需要的指示，以達成特定功能的程序。為了適當回應，機器必須經過編程，才能夠理解對話中明確語言和適切線索的意圖和態度。

人工智慧常用的編程語言包括 Python、Java、Lisp、Prolog 和 C++。Python 是一種很常用於人工智慧應用的電腦程式，其採用模組化架構，著重在特定功能的個別區域，採用相對簡單的原則組合來提供意義結構規則。其架構包含 NLTK、genism 和 Quepy，很適合用於 NLP 和文字處理。這些架構的個別定義如下：

- NLTK 是開放原始碼 Python 模組的集合，為自然語言工具組 (Natural Language Toolkit) 的縮寫，提供語言資料和詞彙資源，以供在經常部署的作業系統中用於開發 NLP 和文字分析。
- Genism 則為設計用於如無任何無法理解的複雜語意的情況下，可擷取文件中語意的模組。
- Quepy 的作用則是將自然語言問題轉換為使用資料庫查詢語言的查詢。

機器學習和神經網路

對人類而言，學習是我們先天的智慧、生理成熟和體驗中很自然的一部分。學習或許可理解為不斷的自我改善，透過回饋的調整，以得到更好的結果。研究人員經由研究人類大腦（具體來說，就是神經網路），構思出強大的機器學習概念，並將其轉化為人工智慧。神經網路由電子互連組成，這些互連能夠影響許多神經元的集合行為。換句話說，神經網路影響所及範圍更廣，而非只是單一的神經元。

再搭配上深度強化學習（或稱深度學習），透過回饋從正面或負面來強化目標導向的表現，這些最新的突破性技術引進許多極富創意的新策略，能用來處理大量的複雜問題，包括理解及回應自然人類語言和對話。像是 [Seeed Studio Google AIY Voice Kit](#) 和 [Seeed Studio ReSpeaker 2-Mics Pi HAT](#) 這類的產品是很有趣且方便的工具，讓我們開始跨入自然語言辨識工具的領域。

聊天機器人的興起

聊天機器人結合了音訊技術、人工智慧和機器學習，經由接收感測器的輸入，運用演算法決定可採取行動的見解，依照這些見解做出回應，然後再從後續的輸入中不斷學習。聊天機器人為單向、具前瞻思考的通訊工具，像是 Amazon、Apple、Facebook 和 Google 等公司皆使用它們來與客戶互動。聊天機器人會透過機器學習和預測性分析，持續加以改善；它們能夠感測、思考、決策，然後行動。聊天機器人應用程式與人工智慧、機器學習和深度學習互相重疊（[圖 1](#)）。



圖1：聊天機器人與人工智慧、機器學習和深度學習互相重疊。

多數公司很聰明地利用這些聊天機器人來增加營收，提升客戶服務與客戶互動的程度。現在的客戶也非常懂得運用技術，期待公司能快速有效回應他們的需求。像 Siri、Alexa 和 Cortana 等這類的聊天機器人，目前已是家喻戶曉的名字。儘管設計人員還在努力，要讓聊天機器人能夠模擬、甚至超越人類的行為和表現，但現在設計人員的目標則是要讓聊天機器人能透過動態互動來提升自我的智能，變得更加人性化、更吸引人。目標是讓電腦在下列領域像人類一樣說話：

- 文字轉語音和語音轉文字
- 聲調分析和性格洞察
- 自然語言分類器和語言翻譯
- 虛擬人員和對話編程

像是 Google DeepMind 所部署的深度學習人工智慧，讓電腦聽起來不再像是電腦：電腦甚至能用合成的聲音像真人般對話而通過圖靈測試。意思是，其回應對人類來說是如此真實，人類甚至分不出是在跟電腦對話。深度學習是透過處理大量的文字對話，從人類的對話中去學習人類語言和通訊的各個面向，而達成這些結果。人工智慧能利用電腦運算來處理如此大量的文字對話，因此能建立一個夠大的經驗框架，從中找出結論，然後做出適當的人類語言回應。

跟對話用的人工智慧學習溝通的方式雷同，這些人工智慧也能用獨特的人聲和類似人類的感情來回應。除了將聲音儲存在記憶體內，人工神經網路還要處理數千小時的人聲錄音，推斷出特殊的細節，以模擬自然人類的聲調。如此一來，聊天機器人也能利用人工智慧來選擇符合場合的合適聲音和語氣。只需要一小段錄音資料，人工智慧就能從中複製聲音，而且這些錄音資料的說話模式還能跟聊天機器人的說話模式完全不同。

聊天機器人可利用對話人工智慧來提供個人化的服務，它們能用跟自然人類語言類似的對話，提出問題，從中接收高品質的資訊，甚至也能預期得到回饋，進而帶來更多的銷售機會，提高客戶滿意度。聊天機器人能設定成隨時提供最高水準的客戶服務。

聊天機器人也能跨媒介：無論客戶是透過電話、電子郵件或社交媒體應用程式對話，對它們來說都一樣。聊天機器人設定成能在各種情況下適當回應，它們是數位化的客戶應對選項，運用許多幫助人類商業環境轉型的技術和數位內容。聊天機器人適合用在特定的業務領域，包括銷售、行銷、客戶服務，以及其他類似的角色上。例如，某些公司早已開始運用聊天機器人的數位技術進行互動，它們同時也是公司與公司客戶互動的首選。

聊天機器人所面對最重要的一項問題，就是如何幫助公司來滿足特定的使用者需求。因為人的心思極其複雜，僅管對話內容可能極富動態且順暢，聊天機器人若能理解使用者的要求則再好不過。這牽涉到需要去理解人類語言中的細微差別和不同，以免解讀錯誤。

結合預測性分析，擁有優異效能的人工智慧聊天機器人能夠預測對話的走向，彷彿能讀出互動對象的心思。換句話說，聊天機器人能夠進化，著重在提出建議和預測，因此很適合不斷加以改良，讓它們採取特定的行動。打造智慧不斷提升的聊天機器人是一大挑戰，設計人員需要努力提升它們對情境式回應的關聯感知能力，與人類的互動和需求有正向的連結。

結論：討論出現

人工智慧、機器學習和音訊等技術演進經過融合，透過聊天機器人創造出類似人類般的人機互動。能夠辨識及解譯語音和聲調的聊天機器人變成了虛擬人員，提供基本的客戶服務和類似的互動，而且能夠解譯、回應，以及從語音輸入和眾多的枝微末節和線索中學習。

但是，為了適當回應，機器必須經過編程，方能理解與對話當中的意圖和態度有關的明確語言和適切線索。這便是自然語言編程的演進能協助推動人機互動和成功雙向溝通之處。神經網路也是另一項重要的發展，讓機器能從先前的互動中學習。

也許哪天我的電腦就能幫我寫出一篇技術文章，我期待那天的到來，希望到時電腦說出的第一句不是「我覺得這內容不成熟，簡直外行且愚蠢。我建議你寫...」。要是電腦現在就能讀我的唇語，或許其他人就不會笑我，就像我們在玩 Hearing Things™ 遊戲時一樣。