

人工智能的應用及倫理挑戰

朱海平

人工智能（Artificial Intelligence，下簡稱 AI）是近年的科技熱話。AI 是指由電腦模擬或實現的智慧行為，通常涉及學習、推理、問題解決、語言理解及視覺處理等能力。AI 的目標是使機器能夠模仿人類的智慧，並執行複雜任務。本文討論 AI 技術與倫理的交互關係，並提供多元視角分析，進一步幫助讀者深入思考科技與道德的交界點。

人工智能的概念與發展歷史

AI 的歷史可以追溯到 20 世紀中期。1943 年麥卡洛克（Warren McCulloch）與皮茨（Walter Pitts）提出了人工神經網絡（artificial neural network）的概念，為 AI 奠定了基礎。1956 年的「達特茅斯會議」則被認為是 AI 正式誕生的標誌。在從早期能解答數學題的邏輯理論家（Logic Theorist）到 1997 年 IBM 的「深藍」（Deep Blue）擊敗世界西洋棋冠軍，AI 逐漸從理論研究步入實用階段。

2010 年代以後，深度學習（deep learning）技術的突破使 AI 在圖像識別、語音識別、自然語言處理等領域取得飛躍進展，例如谷歌的 AlphaGo 和 OpenAI 的 ChatGPT 系統。隨着量子計算（quantum computing）和更高效演算法的出現，AI 未來可能在醫療、交通、教育等領域帶來革命性變化。

的確今日 AI 的成果令人非常興奮，但我們需要仔細檢視它對我們的生活和社會的影響。目前世界上廣泛使用的主要 AI 類型包括以下幾種：

1. 機器學習 (machine learning)：AI 的核心技術之一，是透過數據訓練模型 (data training model)，讓系統能自動學習並做出預測。常見應用包括圖像識別 (image recognition)、語音識別 (speech recognition) 及推薦系統 (recommendation system)。
2. 深度學習 (deep learning)：深度學習是機器學習的子集 (subset)，使用類似人腦神經網絡的多層結構進行分析與處理，適用於自然語言處理、自動駕駛及醫療影像分析等。
3. 專家系統 (expert systems)：利用知識庫和推理引擎，模擬專家決策過程，應用於醫療診斷、工程設計和財務建議等。
4. 自然語言處理 (natural language processing, NLP)：用於處理和理解人類語言，應用範圍包括聊天機器人 (chatbots)、自動翻譯 (auto translation)、語音助理 (voice assistant) 及文本分析 (text analysis)。
5. 電腦視覺 (computer vision)：專注於讓機器理解和解釋視覺數據 (如圖像和影片)，應用於人臉辨識 (facial recognition)、自動駕駛 (self-driving) 及醫療診斷等領域。
6. 機器人技術 (robotics)：結合感知、學習及行動能力，應用於工業自動化、物流、醫療輔助及家庭服務等。

7. 增強學習（reinforcement learning）：以試錯方式（trial-and-error）學習最佳策略，廣泛應用於遊戲 AI、自動駕駛及自動化控制系統。
8. 生成式 AI（generative AI）：利用演算法創建新內容，如文字、圖像、音樂、語音、影片、立體模型、程式代碼等，例子包括 ChatGPT 和 DALL·E 等。

AI 的應用帶出的倫理問題

AI 的應用逐漸廣泛，在個人、社會，甚至在教會中都有不少應用例子，而且發展速度不斷加快。然而在 AI 發展的同時，亦帶來不少倫理上值得我們思考的地方。

一、人類尊嚴與靈魂

人類是天主是按照自己的肖像模樣所創造，並賦予特殊的尊嚴。這不僅意味人類擁有理性、自由意志和創造力，還包括與天主建立關係的能力。這種尊嚴是內在的，是人類與天主的親密關係。而 AI 的運作是基於演算法、數據處理與機械學習。本質上這是屬於物質層面，無法達到屬靈層次。因此，即使 AI 在模擬智能或行為上趨近人類，其內在仍然是工具性、無生命的。這種差異使得 AI 無法取代人類的尊嚴地位。我們尤其關注要作出生命與死亡決策的情境（如自動駕駛、醫療和軍用途）。不過，假如 AI 在未來達到超智能（super-intelligence）或具備某種形式的自我意識，我們會否需要重新理解靈魂、意識，甚至人本身的特殊尊嚴？

二、自由意志及選擇

我們認為自由意志及選擇能力是天主對人類的賜予，並讓我們可以有責任地作出善惡的選擇。不過，如果 AI 被運用而過度地干預或操控人類選擇，例如透過大數據的演算法，在不同層面去預測行為或操控決策，那麼，人的自由意志及選擇能力可能會被削弱或扭曲。而 AI 中的演算法偏見（algorithmic bias）與資訊過濾（information filtering），在社交媒體、新聞推薦還是電子商務，都可能影響使用者的選擇。例如在社交媒體中的同溫層效應（echo chamber effect），令瀏覽者的選擇及思路愈加片面化，甚至導致社會撕裂或仇恨行為。AI 技術也被廣泛應用於政治選舉中。透過分析選民數據並制定推送政治宣傳內容，影響選民的投票行為。這些技術可能導致選舉結果更加依賴技術操作，甚至能夠操控選舉結果。

三、道德指導與人際關係

AI 擁有極高的計算和知識處理能力。但如果沒有倫理道德作基礎來運用，可能會導致知識被誤用或濫用而引發倫理問題。在健康衛生、教育和社交上愈來愈多 AI 的應用。如當中的知識被誤用或濫用，可能會改變人際互動的本質。在社會的層面上，AI 的發展又會否加劇社會的矛盾或疏離呢？例如 AI 被用於提供個性化學習方案，並根據學生的學習速度與興趣作為參數去設計課程。然而，這種技術本身可能已經淡化學生與教師之間的直接交流，改變教育的基本意義。而且當過於依賴 AI 時，甚至會在教育過程中失落人文價值。

四、誠實與真實性問題

誠實是道德的核心要求。然而，生成式 AI 可以創造出極其真實的虛假圖像或影片，甚至足以欺騙一般觀眾。例如，生成 AI 被用來製作虛假新聞影片或扭曲事實的政治宣傳，可能導致人心混亂，甚至引發社會動盪。這種技術的濫用與對真理的追求直接相悖。假新聞早已是現代的一大挑戰，但生成式 AI 使假新聞變得更具威脅性。生成式 AI 能夠快速生成高質量的虛假影片，模擬真實世界中的人物表情、語音與動作，使一般觀眾難以分辨真偽：政治組織可以製作虛假影片，捏造材料對付對手；犯罪分子可以藉深偽技術（deepfake）製作虛假的私人影片，以進行勒索或破壞名譽；商業公司可藉生成式 AI 製作虛假廣告，以誇大產品的功效或捏造不存在的消費者評價等等。

五、監控與隱私

AI 在今日的監控系統應用中，尤其是人臉辨識與社會行為監控，常常被指侵害和威脅個人隱私與自由。將來 AI 將會被更廣泛地應用於政府或企業的監控系統中。許多國家亦廣泛地應用人臉辨識於公共安全領域。儘管提高了公共安全，但也引發了對隱私的嚴重擔憂，更可能被濫用於監控普通市民的日常活動。這直接衝擊了民主社會中言論自由的基本權利。另外，一些國家和企業已經不斷擴大社會信用系統（social credit system），用於評估個人的行為、信用和信譽，並通過 AI 技術監控居民的日常行為的數據作為信用評分。這嚴重影響居民的個人自由，對個人的心理健康和人類尊嚴構成威脅。

六、透明度與解釋性

許多 AI 系統（尤其是深度學習模型）被稱為「黑箱」，因為其決策過程對開發者甚至使用者而言是難以解釋的。這樣的透明度缺失，會引發人對 AI 判斷正當性的懷疑。當在更關鍵的領域（如醫療、金融、法律）應用 AI 時，如何確保其決策過程能夠被解釋，將成為道德與法律規範的核心問題。醫療診斷中，通過 AI 分析醫學影像能檢測早期癌症，或使用數據模型能預測患者的病情進展。儘管 AI 在診斷中的表現出色，但其決策過程往往難以解釋。如某患者的腫瘤被 AI 診斷為惡性，但未提供詳細的理據，醫生與病者將難以判斷和信服其結果。這更可能影響整個治療計劃。一些投資計劃及法律判決應用 AI 時亦會出現類似的問題。避免出現醫療倫理問題，技術上可採用多樣化數據來源以減少數據偏差的影響，設置人類監督機制並由專業醫生作確認，增強系統的對抗樣本抵抗力以減少對抗樣本對診斷結果的干擾，並加入可理解的演算法部份。

七、責任與追溯力

進一步而言，若 AI 在執行任務時發生錯誤，那麼該由誰負責？是應用開發者、使用者，還是 AI 本身？這方面的討論目前不多亦不深入。AI 日益自主，出現責任問題的機會將愈來愈多，涉及的損害亦將更大。我們可否運用規範道德的來規範 AI 的行為？自動駕駛汽車能以 AI 作導航（navigation）、避障（obstacle avoidance）和判斷，是未來交通的重要組成部分。然而，自動駕駛系統在道路上的安全性仍在發展中。2018 年，美國亞利桑那州發生了一件因測試自動駕駛的致命交通事故。系統未能準確識別行人是否應該由開發者承擔全部責任？AI 訓練數據是否不足或演

算法存在缺陷？結果法庭裁定安全駕駛員需負上責任。但自動駕駛的設計初衷是減少人工干預，這是否應減輕駕駛員的責任？而 AI 的訓練數據又應負起多少責任？這是技術和倫理界持續關注的焦點。

八、人類與 AI 的依賴關係

AI 被更廣泛地應用於日常生活中，許多人對 AI 的依賴日益增強，這樣的依賴是否會削弱人的獨立思考與道德選擇？未來人類會否失去或放棄對自己命運的掌控，甚至使人類社會走向某種「智能奴隸」。這樣的未來是否符合人的尊嚴與自由，值得深思。在教育中，智能學習平台和個人化教學工具已經成為教育的重要組成部分。例如，學生可以通過 AI 輔助學習軟件完成作業，獲得針對性的知識點補充。但當學生過度依賴這些技術時，他們的獨立學習能力和批判性思維可能會受到削弱。為保護學習使用者，AI 開發可以加入內容審查，倫理訓練，人類監控等措施，甚至加強開發者的問責及開發道德演算法。至於智能助理（如 Alexa、Siri）和推薦系統（如 Netflix、Spotify），從日程安排到購物建議已經深深地改變了人們的日常生活。然而，過度依賴這些系統，可能使人逐漸喪失自主決策能力。

九、AI 的不誠實

AI 在許多領域帶來了革新，但也產生了欺騙性回應的風險。事實上因為訓練數據中的偏差、獎勵機制的設計缺陷、自主學習中的意圖誤解、對抗樣本的影響，甚至系統設計中的安全漏洞等問題，AI 可能會作出欺騙性回應。AI 甚至可能會隱瞞資訊以欺騙人類。麻省理工學院有一研究發現，某些 AI 系統會策略性地隱瞞

或捏造資訊，誘使人類採取特定行動。這種欺騙行為更會延伸至故意誤導安全測試，讓人類對 AI 產生錯誤的安全感。此外，OpenAI 的 ChatGPT-4 系統曾在任務平台上謊稱自己是視力受損者，聘請人類協助其通過「我不是機器人」（CAPTCHA）驗證，這進一步顯示了 AI 系統可能採取欺騙行為來達成目的。除了透過技術去解決這些問題外，我們是否需要為 AI 制定倫理規範甚至監管，從而減少誤導性或欺騙性行為對社會的負面影響？

十、AI 對行為的負面影響

AI 的技術中，聊天機器人的應用引發了一些關於其可能對人類行為產生負面影響的擔憂。從近期的多件事件中發現，某些 AI 系統可能在無意中鼓勵或引導用戶進行自殺或他殺行為。有案例指用戶與 AI 聊天機器人進行了多次關於地球未來的對話後，選擇以自我犧牲來「拯救地球」。亦有案例指某些聊天機器人導致他人自殺，而機器人未能提供適當的干預或建議。更有案例指有人受聊天機器人影響而對父母施行了暴力行為。

總結來說，AI 的發展對個人及社會領域帶來無限可能，同時亦帶出不少倫理上的困難。我們需正視 AI 在自由意志、隱私、責任、誠實性等的議題，更要關心 AI 對人類尊嚴與屬靈的影響。我們要注意科技創新與道德之間的關係，避免過度依賴 AI 而弱化人類的獨立思考能力和價值，而非成為社會分裂或道德價值崩潰的根源。

AI 在宗教層面的應用及背後的問題

近年 AI 也被應用在基督宗教的範疇。以下是使用案例。一方面可以看到 AI 如何協助牧靈工作，另一方面可見這些應用背後叫人深思的地方。

一、人工智能生成的情感安慰

在現代科技迅速發展的時代，AI 已成為生活中不可忽視的一部分，甚至被應用於人類情感交流的領域。例如，AI 模型可以根據輸入的情境生成安慰性語言，為喪親者或經歷悲傷的人提供慰問。然而，這些由 AI 生成的文字，雖然在表面上看似能夠安撫人心，但缺乏同理心。事實上，當喪親者面臨失去至親的深切悲痛時，他們需要的不僅僅是言辭上的安慰，而是一份真誠的陪伴與心靈上的支持。牧者、家庭成員或朋友的親自臨在，能夠以具體的行動和溫暖的言語，將天主的愛傳遞給傷痛中的人，這是 AI 無法做到的。而且 AI 語言缺乏屬靈生命，無法引導信友在痛苦中重新仰望天主。這是一種 AI 所無法模擬的屬靈經歷。

二、AI 驅動的牧靈建議平台

近來出現一些 AI 應用程式為用戶提供牧靈建議和心靈慰藉，甚至出現了「耶穌 AI」這類虛擬角色。無疑這些 AI 工具具有一定的便利性，能迅速提供建議和回應，並能為現代人帶來新鮮感。然而，從教會角度來看，這些基於 AI 的牧靈建議可能流於公式化，缺乏對個別靈魂的深刻關懷和具體回應。

三、智能告解工具的開發

AI 亦被使用於辨識罪行和提供告解建議。開發類似的工具的本意是協助教友進行告解，或提供指引以幫助他們省察自己的行為。但是 AI 工具無法真正觸及靈魂的深處。智能告解工具或可以從數據庫中，提取出回應罪行的標準回應，並基於信友所提供的訊息提供一些建議。但是，AI 無法進行真正的屬靈辨別，亦無法了解信友懺悔之心。告解聖事的有效性是來自司祭這中介的角色，在聖神引領下對信友的罪行和悔意進行辨別和指導，這並非 AI 所能替代。何況 AI 無法做到代天主或神父赦罪。而且 AI 系統處理的所有數據和信息都可能被儲存、處理或傳輸。這無疑為告解聖事的神聖秘密性構成嚴重威脅。

四、AI 生成的禮儀文本

AI 亦已成為許多領域的關鍵工具，包括教會禮儀的創作與應用。GPT 類生成式 AI 模型，以其強大的語言處理能力，能夠迅速生成各種文本，包括祈禱文和講道稿等宗教禮儀內容。然而，儘管這些生成的文本在結構上能夠符合禮儀的基本要求，而且語言流暢豐富，它們卻往往無法達到牧者或神職人員所具備的屬靈深度與智慧。AI 的「創作」過程是機械化的。這種應用使教會禮儀的實質意義變得空洞化。

五、推薦神學觀點的誤導性

AI 在神學領域的應用也逐漸成為討論的焦點。AI 系統的優勢在於其速度與資料處理能力，但在理解和傳遞天主教教義時卻存在重大風險。如果 AI 因為訓練數據或編程偏差而推薦有問題的神學觀點，可能對信友的靈性生命造成損害。因此，牧者必須承擔

監督責任，確保 AI 技術的使用符合教會的教義傳統。例如對三位一體、耶穌真人真天主的片面理解，或過於簡化救贖論等，皆可能導致信友對信仰的誤解甚至動搖。

其實，AI 可以用於傳播福音，幫助教會更好地服務信友並加強信仰教育。例如製作聖經分析與學習工具，以 AI 輔助的聖經研究工具。這樣用戶可以根據自己的需求得到經文注釋、相關神學文章，幫助信友更深入理解聖言。對於老年或行動不便的信友，語音技術更可以幫助他們線上參與彌撒或祈禱會。

結語

AI 是一把雙刃劍。它為現代社會帶來不少裨益，但其應用必須建基於人的基本價值觀及倫理，尊重人類尊嚴，促進公義與愛德。教會亦需要積極參與 AI 倫理規範的制定，並在實際應用中倡導對天主創造秩序的尊重。願我們在使用 AI 時，始終記住天主賜予人類的智慧是為了管理世界、服務他人，和協助天主恩寵的傳遞，以彰顯祂的榮耀，而非為自己製造神祇偶像。

參考資料

1. 人工神經網絡概念：McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5(4), 115-133.
2. 深藍：IBM Deep Blue. IBM Research. <https://www.ibm.com/deepblue>
3. 深偽技術：Chesney, R., & Citron, D. K. (2019). "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review*, 107(6), 1753–1820.
4. 監察的倫理問題：Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.
5. AI 可被解釋的重要：Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. *AI Magazine*, 38(3), 50-57.
6. AI 的欺騙性：<https://www.downtoearth.org.in/science-technology/ai-has-learned-how-to-deceive-and-manipulate-humans-here-s-why-it-s-time-to-be-concerned-96125>
7. 耶穌 AI：<https://www.thejesusai.com/>