



The Holy See

教宗方濟各

第 57 屆世界和平日文告

2024 年 1 月 1 日

人工智能與和平

新年，是上主賜給我們每一個人的恩寵時刻，在這新的一年開始，我願向全體天主子民，向不同的民族、國家及政府首長、不同宗教和民間社團的領袖，以及我們這個時代的所有男士、女士致意，獻上我對和平的熱切祝願。

1. 科技的進步作為走向和平的途徑 聖經裡證實天主賜下聖神給人類，使他們有「智慧、技能和知識，能做各種工程」（出卅 五 31）。人類的聰明才智是造物主賜予尊嚴的表現，因祂按自己的肖像和模樣創造了我們（參閱：創一 26），且讓我們能有意識並自由地回應祂的愛。科技則以特別的方式，彰顯了人類智慧基本的相關特質——人類創造潛力的傑出成果。

梵二大公會議文獻《論教會在現代世界》牧職憲章裡重申此一真理，憲章提到「人類經常勞動並動用腦筋，以改善自己的生活。」¹ 當人類藉助技術，努力使大地成為適合全人類的住所時，² 他們就履行了天主的計畫，同時與祂的旨意合作，使這創造臻於完美，並促成各民族之間的和平。科學技術的進步，已使人類社會更加有序，並更具有兄弟情誼般的共融與自由，導致了人類的進步和世界的變革。

由於科學與技術上令人矚目的成就，過去使人類生活飽受煎熬且造成極大痛苦的疾病，如今得以醫治，我們理當為此高興與感激。科技的進步，使得人類能控制現實的程度前所未有，於是我們可以有大量的選擇，包括那些對我們的生存和家園會造成威脅的選擇。³

新的資訊科技，特別是在數位領域的驚人進步，給了我們令人興奮的機會，也帶來重大的風險，對於追求人民之間的正義與和諧產生了嚴重的影響，因此必須提出許多迫切的問題：這些新的數位科技會帶來那些中、長期的後果？對個人的生活和社會、國際的穩定及和平，會有什麼影響？

2. 人工智能的未來：在承諾與風險之間 近數十年來，資訊科技和數位技術的進步，已使全球社會及各種動態產生深遠的改變。新的數位工具甚至正在改變傳播、公共行政、教育、消費、人與人之間的互動，以及我們日常生活中無數其他層面的樣貌。

此外，從遍布網際網路的數位足跡中，科技利用各種演算法來提取數據，使它們能在我們不知情的情況下，控制我們的心理和相關習慣，以達到它們的商業和政治目的，因而限制了我們有意識的選擇自由。在承載過量的網路空間中，他們可以經常在使用者察覺不到的情況下，根據使用者的選擇標準來建立數據流。

我們必須記住，科學研究和技術創新不是脫離實體的，也不是「中立」⁴的，總會受到文化的影響。作為全體人類的活動，它們所發展的方向反映出在任一特定年齡的個人、社會和文化價值觀之下所作的選擇。人類以什麼方式對待周遭世界，一如所產生的結果，是有道德的層面，是與人所作的決定密切相關，意即設計實驗計畫，其結果是針對一些特定的目標。

用在人工智能的形式也是一樣。至今為止，在科學和技術領域中，對於人工智能，還沒有一個單一的定義。這個詞彙本身如今已成為日常生活的用語，包含各種科學、理論和技術，其目的是使機器的功能可以複製或模仿人類的認知能力。我們用複數談論「不同形式的智能」，更有助於強調這些系統與人類之間不可逾越的差距，不論那智能是多麼令人驚嘆和強大，最終也只是「零碎」的，意思是說，也只能模仿或重現人類智慧的某些功能。同樣，多種的智能形式也指出以下的事實，那就是這些設備之間存有極大的差異，應該始終被視為是一種「社會技術系統」，因為任何人工智能設備的影響——不論其基礎技術如何——都不僅取決於技術設計，也取決於其擁有者及開發者的目標和利益，以及應用的場景。

為此，人工智能應該被理解為恍如星系的不同現實。我們不能事先假定它的發展會有益於人類的未來及人民之間的和平。只有在我們展現自己有能力負責，並尊重「包容、透明、安全、公平、隱私和可靠」⁵等人性基本價值的情況下，才有可能得到這正面的結果。

同樣，我們也不能只是假定那些設計演算法和數位科技的人，願意承諾對他們的行為負起責任且是合乎道德的。我們必須加強，或有必要時去建立機構，來負責審查這一領域中所出現的道德問題，並保護那些採用人工智能或受其影響者的權益。⁶

因此，科技的巨大擴張，必須隨之有適當的培育，使科技對未來的發展負責。只要人類一旦屈服於自私自利、追求利潤和渴望權力等，自由與和平的共存就受到威脅。為了個人及團體的整體發展，我們有責任拓寬視野，把科學技術研究導向追求和平及共同利益層面。⁷

每個人與生俱來的尊嚴，以及把我們凝聚成一個人類大家庭的兄弟情誼，必須成為新技術發展的根基，並在這些技術被應用之前，作為無庸置疑的評估標準，這樣，數位的進步才能與正義並行，且有助於和平大業。技術發展若不能提升全人類的生活品質，反而加重不平等及衝突，則永遠不能視之為真正的進步。⁸

人工智能會越來越重要。它造成的問題既是技術上的，也是人類學、教育、社會和政治上的。例如

，它保證我們免於辛苦的工作，提高製造效率、更便捷的交通運輸、更方便的市場，且大大革新了收集、整理和確認數據的過程。然而，我們必須小心留意如今的快速轉變，並要謹慎加以管控，好使基本人權得以保護，同時強調推動人類整體發展的制度及遵守法律。人工智能應有助於發展我們最佳的潛力和最崇高的志向，而不是與之相抗衡。

3. 未來的技術：能自行學習的機器 人工智能有多種形式，而基於「機器學習」(machine learning) 所發展的人工智能，雖然還在開創階段，但已經給社會的結構帶來可觀的改變，對文化、社會行為及建立和平等也產生了深遠的影響。

像這種機器學習或「深度學習」(deep learning) 的發展，引發了超越技術和工程領域的問題，事關深刻了解生命的意義、知識的建構及心靈獲得真理的能力。

例如，某些儀器雖有能力產生句法和語意一致的文本，但不能保證它的可靠性。據說它們是在「幻想」，也就是說，會創造出一些乍看之下似乎可信，其實是毫無根據或有不實偏見的陳述。當人工智能用於傳播假新聞，引起人們對媒體的日漸不信任，就會帶來嚴重的問題。隱私、資料所有權及智慧財產權等，也是此一技術會引起風險的領域。我們還可以加上這些技術的濫用所帶來的其他負面後果，諸如：歧視、干預選舉、對社會的監控增加、濫用網際網路排斥他人，以及與社會越來越脫節的個人主義等問題，日益嚴重。這種種因素都會助長衝突，阻礙和平。

4. 在技術官僚範例中對限度的意識 我們的世界太廣大、變化多端且複雜，永遠無法充分了解和完整分類，即使借助最先進的演算法，人類的思想也永遠無法將世界裡豐富的東西一一探討。這樣的演算法無法保證對未來的預測，只能提供統計上的近似值。不是所有事都能預測，也不是一切都能計算；「現實終究大於思想」，⁹ 不論我們的計算能力有多驚人，總是會有一些無法計量的空間，讓我們對任何量化的嘗試，力有未逮。

此外，人工智能所分析的龐大數據，本身不能保證客觀公正。用演算法推斷資訊時，總是存在著失真的風險，複製了原來環境中的不公正與偏見。演算法越快速、越複雜，就越難了解為何會產生某一特定的結果。

「智能」機器會以越來越高的效率，來執行指派給它們的任務，但它們操作的目的和意義，仍然會由人類自己擁有的宇宙價值觀來決定或操控。其風險就是，在某些決定背後的標準，會變得不清晰，對那些決定的責任被隱匿，製造商則能規避他們為社區利益而行動的義務。在某種意義上，這是技術官僚制度所青睞的，將經濟與技術連在一起，並特別重視效率，而忽視與直接利益無關的任何事物。¹⁰

這應該促使我們省思在當前技術官僚和效率導向的心態上往往被忽略的一些事，這對個人及社會發展極為重要，那就是「對限度的意識」。按定義來說，人必有一死。但人意圖藉著科技來克服每一種限度，不能抑制地渴望控制一切，我們就可能失去對自己的掌控；為了追求絕對的自由，我們可能陷入「技術獨裁」的漩渦。承認並接納我們身為受造物，即承認我們是有限度的，這是我們得到，或更好說是欣然接受「成就」這份恩賜所不可或缺的條件。在技術官僚的範例脈絡中，被普羅米修斯般的心態激發，也就是在自以為是的意識型態下，不平等的現象可能會急遽成長，知識和財富

集中在少數人手中，使民主社會及和平共存受到嚴重的危害。¹¹

5. 倫理上的急迫問題 未來，申請抵押貸款者的信用、申請工作者的適用性、罪犯再犯的可能性，或是接受政治庇護或社會福利等，都可由人工智能系統來決定。這些系統的引進，由於缺乏不同層級的調停，特別容易造成各種形式的偏見及歧視：系統性錯誤很容易倍增，不但在個別案例中會產生不公正，而且由於骨牌效應，也會產生真正的社會不平等。

有時人工智能的各種形式會受制於事先確定的、與激勵和勸阻相關的選項，或透過以資訊設計來操作的系統來規範人的選擇，這些似乎都能影響個人的決定。這些形式的掌控或社會規範，需要非常謹慎地去關注及加以監督，同時也意味著，製造商、部署者和政府當局都需要負起明確的法律責任。

對於自動化過程的依賴，例如：廣泛使用監視器，或採用社會信用體系，將個別人士分類，同樣會對社會結構產生深遠的惡果，因為它在人民之間造成了階級之分。這種人工分類的過程，也會導致權力衝突，因為它不僅關係到虛擬用戶，也關係到真實的人民。基於對人類尊嚴的基本尊重之要求，我們拒絕將人的獨特性視為等同於一組數據。演算法必不可用來決定我們如何了解人權，從而把人類的固有價值，如：同情、仁慈和寬恕置之一旁，而且也無法避免一個人有可能用它改變並擺脫自己的過去。

在這種情況下，我們也不能不考慮到新科技對工作場所的影響。過去專屬於人力勞動領域的那些工作，如今迅速被人工智能在工業上的應用所取代。在這方面，也存在著少數人獲得極大利益，而多數人卻付出貧困代價的風險。尊重勞工的尊嚴，以及就業對個人、家庭和社會經濟福祉的重要性，應該是國際社會最應優先關注的，因為這些科技形式已更深入地滲透我們的工作場所中。

6. 我們要化刀劍為犁頭嗎？在這些日子，當我們環顧這世界，會看到一個無法逃避的嚴重道德問題，與軍火產業有關。利用遙控系統指揮軍事行動的能力，會使人疏於留意系統所造成的破壞，以及對使用武器負責的心理負擔，以致對戰爭的巨大悲劇更加冷漠無感。對於正在興起的所謂「致命自動武器系統」，包括武器人工智能化的研究，會引起對道德倫理的嚴重挑戰。自動武器系統絕不可能負起倫理道德的責任，唯有人類有能力作道德判斷及合乎倫理的決定，這是一組複雜的演算法遠不能及的，畢竟那能力不能簡化為將機器程式化，不論那機器多有智慧，終究是一部機器。因此，人類亟需對武器系統進行充分、持續且有意義的監督。

我們也不能忽視尖端武器最後落在不當人士手中的可能性，例如：促成恐怖分子的攻擊或顛覆政府的合法體制。總之，這世界不需要有助於商業及武器交易的「新」技術，這只會導致不公義的發展或瘋狂的戰爭。若這樣做，不只是人的智慧，甚至人類心靈本身也有可能變得更「人工化」。最先進的科技，不應用來助長以暴力解決衝突，而應該為和平鋪路。

從更積極的角度來看，如果將人工智能用於促進整體人類的發展，就可以在農業、教育和文化上帶來重要的新發明，改善整個國家和人民的生活水準，並增進人類的手足情誼及社會的友愛。最後，我們如何以最小弟兄姊妹、易受傷害者以及最需要幫助者為念，來使用人工智能，才是衡量人性的真正標準。

一個真正人性化的觀點，以及對我們世界有更美好未來的渴望，無疑表示我們需要進行跨領域的交談，以實現演算法合乎倫理——即演算倫理——的發展，其價值觀會塑造出新科技的走向。¹²在研究工作的起始，即應考慮到倫理問題，並且持續到實驗、設計、生產、分發及營銷的各個階段。這就是設計倫理學的做法，也是教育機構和決策者必須扮演的重要角色。

7. 教育面臨的挑戰 一個尊重人類尊嚴並為其服務的技術發展，對我們的教育機構和文化世界有明顯的影響。數位科技大幅增加了傳播的機會，使我們能以新的方式彼此相遇。然而，我們仍必須不斷地省思，這科技會引導我們進入何種關係。我們的年輕人成長於科技普及的文化環境，這也不可避免地對我們教學、教育和培訓的方法帶來了挑戰。

使用各種形式的人工智能來從事教育，其首要目標應該是促進批判性思考。各個年齡層的使用者，特別是青年，對於在網上收集的數據，或由人工智能系統生成的內容，必須培養一種辨明的能力。學校、大學和科學社團都面臨挑戰，要幫助學生和專業人士理解科技的發展和應用上的社會層面及倫理道德。

訓練人們使用新的傳播工具，不但要考慮錯誤訊息和「假新聞」，也要考慮到下面令人困擾的情形復發，那就是「隱藏在新技術背後且不斷蔓延的由來已久的恐懼……。」¹³令人遺憾的是，我們發現自己要再次面對「建立圍牆文化的誘惑，豎立圍牆，以阻止我們與其他文化和與其他人民交流」¹⁴的惡果，以及對和平友好共存發展的影響。

8. 發展國際公約所面對的挑戰 從人工智能的全球化規模可以清楚地看到，除了主權國家有責任由內部規範它的使用外，國際組織在達成多邊協定及其應用和執行方面，也有決定性的角色要扮演。¹⁵關於這方面，我力促各國的全球社群能共同合作，以制訂一個有約束力的國際協定，來規範各種形式的人工智能的發展和使用。當然，規範的目的不應該只是防止有害的使用，也要鼓勵做最好的用途，要能激發新穎且有創意的做法，並鼓勵個人或團體的創新。¹⁶

總而言之，在尋求能對數位科技的開發者提供道德指導的規範模型過程中，必須要能指出人類的價值觀，才能駕馭社團制訂、採用並執行所需要的規範架構。為產生各種形式的人工智能的工作而制訂的道德指南，不能不思考人類存在意義、保障基本人權和追求正義與平等更深刻的議題。倫理和法律上的分辨，可成為一個可貴的機會，來共同省思科技在個人和公共生活中應扮演的角色，以及要如何應用才能有助於創造一個更公正、更人道的世界。因此，在談論人工智能的規範時，所有利害關係人的聲音都應納入考慮，包括那些在全球決策過程中經常被忽視的貧困者、無權無勢者及其他人士。

* * * * *

我希望上述的反思，能鼓勵大家努力，確使人工智能形式的進展最終將成為服務於人類友誼與和平的大業。那並非少數人的責任，而是整個人類大家庭的責任。因為在人類關係中，「和平」是承認並接納他人不可剝奪尊嚴的果實，也是在追求個人及民族的整體發展中，合作和承諾的果實。

值此新年伊始，我祈願各式各樣人工智能的快速發展，不致加劇普遍存在於當今世界的 不平等和不

公義情況，而是幫助這世界終止戰爭和衝突，減輕人類大家庭所遭受的多種苦難。願基督信徒、不同宗教的追隨者，以及所有善心人士，在和諧中共同努力，迎接數位革命帶來的各種機會，並面對所帶來的挑戰，讓我們的後代能有一個更精誠團結、更富於正義與和平的世界。

教宗方濟各

發自梵蒂岡

2023年 12月 8日

(臺灣明愛會 恭譯)

1 《論教會在現代世界》牧職憲章，33。

2 同上，57。

3 參閱：《願祢受讚頌》通諭，104，(2015 年 5 月 24 日)。

4 參閱：同上，114。

5 在密涅瓦對話 (Minerva Dialogues) 活動中向與會者致詞 (2023 年 3 月 27 日)。

6 參閱：同上。

7 參閱：〈致達佛斯世界經濟論壇會議主席的訊息〉(2018 年 1 月 12 日)。

8 參閱：《願祢受讚頌》通諭，194，(2015 年 5 月 24 日)；〈向「電子時代的公共利益研討會」與會者致詞〉，(2019 年 9 月 27 日)。

9 《福音的喜樂》勸諭，233，(2013 年 11 月 24 日)。

10 參閱：《願祢受讚頌》通諭，54，(2015 年 5 月 24 日)。

11 參閱：向宗座生命科學院大會與會者致詞，(2020 年 2 月 28 日)。

12 參閱：同上。

13 《眾位弟兄》通諭，27，（2020年10月3日）。

14 同上。

15 參閱：同上，170~175。

16 參閱：《願祢受讚頌》通諭，177，（2015年5月24日）。