台北寰宇 AI 倫理學會

Taipei Global AI Ethics Society (GAIEST)

成立大會手冊

2026 年 10 月 26 日

掃描進入學會官網

掃描下載手冊Word檔

目　錄

湯紹成教授　簡介.....1

湯紹成教授　生成式 AI 能力認證.....4

發起人名單.....5

成立宗旨與宣言.....6

Founding Purpose and Declaration.....9

成立大會致詞.....13

Address at the Founding Meeting.....20

湯紹成教授　AI 倫理評論.....29

第一章　風險與治理.....30

AI 將帶領人類走向「類共產主義」嗎？.....31

思考 AI 自動化與生產力躍升可能如何影響分配制度、工作、購買力與未來的社會經濟結構。

AI 部長上任：阿爾巴尼亞的創新？.....33

以 AI 參與政府治理的案例，探討行政效率、透明度與制度責任等公共治理課題。

AI 任由發展，可能帶來大災.....35

由 AI 研究者辛頓的警告出發，論述自主系統風險、演算法權力集中，以及對高風險 AI 保留人類決策權的必要。

AI 自我進化與智能爆炸.....37

探討遞迴式自我改進（RSI）、智能爆炸的前景與不確定性，強調治理透明、問責與人類最終控制權。

第二章　文明對話.....40

防止 AI 成為怪獸：從教宗警告到亞洲 AI 倫理平台.....41

以教宗對 AI 倫理的警示為起點，指出科技發展必須維護人的尊嚴，避免戰爭決策由演算法主導，並倡議亞洲 AI 倫理合作平台。

AI 時代的文明對話：中華文化可為教宗提供倫理答案.....44

提出儒家仁義、道家知止、佛教慈悲及和而不同等思想，可與教廷倫理觀相互對話，建立跨文明 AI 治理的共同原則。

歐美 AI 領域對於中華文化的注視.....47

討論歐美 AI 相關思潮對中華文化倫理資源的關注，探問不同思想傳統如何補強 AI 發展的價值框架。

第三章　東方智慧.....50

AI 與中國傳統智慧：杭州的體驗.....51

從杭州的科技體驗觀察 AI 與傳統智慧如何相互連結，關注 AI 在學習、教育及文化傳承中的可能角色。

讓 AI 科技成為文明傳承的夥伴.....53

以 AI「書童」為例，呼應教宗的科技倫理警示，探討讓 AI 陪伴兒童學習、傳承文明而非取代人的方向。

緣起性空與 AI 模式.....55

從佛教緣起性空觀念出發，思考 AI 系統的關係性、條件依存，以及技術與人文哲學的交會。

成立大會新聞稿.....58

湯紹成教授　簡介

掃描下載
Word檔



學歷

- 文化大學政治系學士（1976）
- 德國波恩大學政治學博士（1990）

經歷

- 國立政治大學國際關係研究中心（NCCU, IIR），副研究員/副教授（自 1991-2010），研究員/教授（自 2010-2019）退休
- 國關中心歐美所所長（2011-2015）
- 政大歐盟中心主任（2013-2015 年）
- 新黨國大代表（1996-2000）

現職

- 政大國關中心退休/兼任教授（自 2019）
- 亞太綜合研究院董事暨院長（2022 至今）
- 中華鄭和學會榮譽理事長（自 2023）
- 周大觀基金會董事（自 2024）
- 中華人權協會理事（自 2026）
- 中德文教基金會董事（自 2025）
- 台北市觀光協會榮譽理事長（自 2024）
- 中華兩岸人民文經交流促進會創會會長（2001-2009）/現任會長（自 2026）
- 台北市上海四行倉庫暨老子學院學會副理事長（自 2026）
- 台北寰宇 AI 倫理學會理事長（自 2026）
- 國民黨國政基金會國政顧問（自 2026）

- 民眾黨國政顧問（自 2026）
- 新黨國政顧問（自 2020）
- 泛藍協會副理事長（自 2026）
- 聯合報與中國時報等專欄作家，與多家電視台政治評論員（自 2015）

研究領域

- 德國問題、歐洲安全、亞歐關係、美中台關係、AI 與社會科學對接

外語能力

- 德語、英語、法語（尚可）

研究與獎勵

- 德國弗里德里希·瑙曼基金會獎學金（1985-1987）
- 國家科學研究計劃（1992，2005，2012）
- 德國學術交流（DAAD）獎學金（1995、2006、2012）
- 德國漢斯·塞德爾基金會獎學金（1998）
- 教育部歐盟特別研究計劃（2003-2005）
- 政大特別研究獎（2003-2005）
- 歐盟莫內（Jean Monnet）教授（2006-2011、2015-2018）

著作摘錄

專書

1. 兩岸關係綜論，台北，海峽學術出版社，2011，271 頁。
2. 台灣的認同問題（中英對照），台北，2012，問津堂書局，29 頁。
3. 從台灣看大陸 –幾位學者的觀察（中英對照），朱高正、湯紹成、杜震華、趙國材等合著，湯紹成編著，台北，海峽學術出版社，2012，421 頁。
4. 從中國看德國（中德對照），上冊，台北，海峽學術出版社，2014，共 283 頁。

5. 從中國看德國（中德對照），中冊，台北，海峽學術出版社，2015，共328頁。
6. 從中國看德國（中德對照），下冊，台北，海峽學術出版社，2015，共209頁。

論文

1. The European Union and the Taiwan Strait Relations (1996-2009) －A Content Analysis, Issues & Studies, Vol.46, No. 1, March 2010, pp. 55-87. SSCI
2. A Comparison between Inner-German Relations in 1970s and Cross-Strait Relations since 2008, Issues & Studies, Vol.46, No. 4, December 2010, pp. 1-36. SSCI
3. 奧地利外交政策研究，東吳政治學報，第28卷，第一期，2010年三月，頁89-137。TSSCI
4. 當前歐盟對中國政策－兼論對台政策之分析，遠景季刊（TSSCI），第16卷，第四期，2015年10月，頁1-70。

新書

- Dynamics of China-Taiwan Relations and the Role of the United States—A Perspective from Taiwan，目前正提交國際知名出版社 Peter Lang 出版，全球發行

湯紹成教授　生成式 AI 能力認證

財團法人資訊工業策進會核發｜授證日期 2026 年 1 月 22 日



發起人名單

依姓氏筆畫排序

丁嘉靖	丁興華	王珍一	王鳳生
尤雲柔	朱　甌	邢天立	李英明
吳正群	吳　克	吳靜儀	林宸豪
卓志強	周靜家	姚博文	馬泰成
馬準威	馬　潮	孫安迪	張芹瑋
張瑞雄	陳秝榛	陳清秀	湯紹成
黃佑華	詹滿容	廖丁民	裴俊宇
熊愛卿	鄭光禮	蔡維仁	蔡篤堅
蔣　哲	劉易齋	劉昱甫	鄧志松
蕭倩玫	賴素禎	盧信昌	魏麓宸
蘭梁筱娟			

成立宗旨與宣言

2026 年 10 月 26 日｜台北

掃描下載
Word檔

當前 AI 飛速發展，正深入人類各生活領域，可謂無所不在，無遠弗屆。AI 可以提升效率與創新，但也會帶來風險與危害，更引發文明與倫理的轉型。因而，我們不能只問 AI「能」做什麼？這是「智」的範疇；更應追問 AI「該」做什麼？這是「慧」的內涵，二者兼備，才能相得益彰。

今（2026）年 5 月，教宗良十四世公布《偉大的人性》（Magnifica Humanitas）通諭，強調 AI 時代應守護人類的尊嚴，以人的良知與自由引導科技，並由人來承擔道德判斷與決策責任，以避免 AI 反噬。這與中華倫理中仁愛與責任的理念極為應和。

今年 6 月，新加坡科技設計大學（SUTD）亦推出「亞太 AI 倫理白皮書」，主張 AI 治理應重視亞洲的歷史文化與社會差異，不能僅依西方經驗。白皮書重視儒家等文明的倫理資源，並倡議跨領域研究、政策協調與區域合作，讓亞洲參與全球 AI 治理。

近日，美國內部更產生 AI 是否暫緩發展的爭議。或認為 AI 發展可造福人類；或主張 AI 應被約束，以防後患。最後在川普總統的倡議下，各 AI 巨頭簽訂一項只具道德約束力的協議，可見當前 AI 的發展已令人憂懼。

有鑒於此，再基於台灣在中華文化方面的深厚底蘊，以及在全球 AI 發展中的重要地位，我們倡議成立「台北寰宇 AI 倫理學會」（GAIEST），將匯聚社會各界力量，推動 AI 倫理研究與教育實踐，培養跨領域人才，盡力避免 AI 可能產生的負面影響，以便促進兩岸、亞洲與全球的和平發展。

AI 源自人類文明共同累積的知識，不應被少數國家、企業或文化壟斷。本會結合國際治理經驗與儒家仁愛義理、佛家慈悲心懷與道家自然和諧等價值，以期建構海內外的對話平台。歷史提醒我們：文明應該交流互惠，才能共創多贏，因而本會主張：

一、天人共生

儒家重視天人合德，道家倡議道法自然，《易經》揭示變化中的平衡，《黃帝內經》體認身心與自然的整體和諧。而 AI 應服務生命，尊重文化差異與語言多樣性，使人類、社會、自然及未來世代共同受惠。

二、慈悲公義

儒家主張「仁者愛人」與「民胞物與」，再加上佛家的慈悲精神，吾等主張進一步要求科技理解人類的思維。AI 不應判定信仰或文明的優劣；而科技能力愈強，對弱勢者的責任愈大，以維護尊嚴與公平。

三、以史為鑑

AI 可以依過去的記錄推測未來。若缺乏歷史理解，偏見、歧視與不平等可能藉演算法而擴大。秉持溫故知新、以史為鏡的精神，資料、模型與決策應接受文化的檢視，使歷史成為創新的資源。

四、全程問責

依循儒家義利之辨、知行合一與佛家因果觀的價值理念，倫理責任應貫穿資料取得、模型訓練、產品設計、應用及退場。本會主張 AI 的來源可知、資料可溯、風險可辨、決策可解、責任可問、傷害可救濟，推動透明、可監督與可追責的治理制度，不能以創新之名免除責任。

五、普惠共享

本會主張，以「天下為公」與眾生平等之理念，要求 AI 成果為社會共享。普惠不只是人人可以使用 AI，更是普世各階層都有能力理解、選擇與質疑 AI。我們倡議「旁通類情」，以推己及人的態度，讓各界人士看見彼此的生命經驗。

六、人機協力

人的價值不能化約為資料，良知、自主與道德判斷不能交由演算法取代。尤其涉及生命及公共安全時，應堅持「人在決策迴路中」（human-in-the-loop），由人作最終判斷並承擔責任。AI 應協助醫師理解病人、教師理解學生、政府理解人民，使科技延伸人的能力。

七、世代永續

道家珍視自然之和，提醒我們節制力量、珍惜資源。重大 AI 決策應追問：是否會耗損下一代所需資源？是否會擴大不可逆的不平等？是否造成語言與文化消失？永續應成為創新的內在尺度，讓今日的便利不以犧牲未來為代價。

八、全球共治

本會主張，以「和而不同」的方式，來促進各種文明平等對話，邀請國內各界人士共同參與治理，並推動國際規範的對話與互動。本會願為東西雙方科技與人文、傳統與現代相遇的平台，以共同智慧促進和平與發展。

總之，本會的使命是以倫理引領 AI，以科技服務人類。文明的智慧，在於明辨價值、節制力量並守護人的尊嚴。讓 AI 成為文明之間的橋樑、弱者能力的延伸，以及守護地球與未來世代的共同智慧。因此誠邀各界有心人士，攜手並進，以 AI 成就人類，同時避免 AI 反噬。讓我們從台北出發，與世界同行，共創美好未來，感謝大家。

Founding Purpose and Declaration

Scan to download Word file

Taipei, October 26, 2026

Artificial intelligence is advancing at breathtaking speed and reaching into every area of human life; it is, in effect, everywhere and without limits. AI can raise efficiency and spur innovation, but it can also bring risks and harm, and it is driving a transformation of civilization and ethics. We therefore cannot ask only what AI "can" do, which is a question of *zhi* (智), or intelligence. We must also ask what AI "should" do, which is a question of *hui* (慧), or wisdom. Only when the two go together can each bring out the best in the other.

In May of this year (2026), Pope Leo XIV issued the encyclical *Magnifica Humanitas*, stressing that in the age of AI human dignity must be protected, that technology must be guided by human conscience and freedom, and that moral judgment and responsibility for decisions must rest with human beings, so that AI does not turn against humanity. This resonates closely with the ideals of benevolence and responsibility in Chinese ethics.

In June, the Singapore University of Technology and Design (SUTD) released its "Asia-Pacific AI Ethics White Paper," arguing that AI governance must take account of Asia's histories, cultures and social differences rather than rely solely on Western experience. The white paper draws on the ethical resources of Confucianism and other civilizations, and calls for interdisciplinary research, policy coordination and regional cooperation so that Asia can take part in global AI governance.

Recently, a debate has also arisen within the United States over whether AI development should be slowed. Some hold that AI will benefit humanity; others argue that it must be restrained to prevent future harm. In the end, at President Trump's initiative, the leading AI companies signed an agreement that is binding only in a moral sense, a sign of how much concern the current course of AI development has caused.

In light of all this, and drawing on Taiwan's deep roots in Chinese culture and its important place in global AI development, we propose the founding of the Taipei

Global AI Ethics Society (GAIEST). The Society will bring together people from all sectors of society to advance research, education and practice in AI ethics, to cultivate interdisciplinary talent, and to do all it can to avert the possible negative effects of AI, so as to promote peaceful development across the Taiwan Strait, in Asia and around the world.

AI grows out of knowledge accumulated in common by human civilization; it should not be monopolized by a few countries, companies or cultures. The Society combines international experience in governance with Confucian benevolence and righteousness, Buddhist compassion and Daoist harmony with nature, with the aim of building a platform for dialogue at home and abroad. History reminds us that civilizations prosper together only through exchange and mutual benefit. The Society therefore puts forward the following principles:

1. Coexistence of Heaven and Humanity

Confucianism values the union of Heaven's virtue and humanity's; Daoism teaches that the Dao follows nature; the *Book of Changes* reveals balance within change; and the *Yellow Emperor's Inner Classic* recognizes the integral harmony of body, mind and nature. AI should serve life and respect cultural difference and linguistic diversity, so that humanity, society, nature and future generations all benefit together.

2. Compassion and Justice

Confucianism teaches that "the benevolent love others" and that "all people are my siblings and all things my companions"; Buddhism adds its spirit of compassion. Building on these, we further ask that technology understand the ways human beings think. AI should not pass judgment on which faiths or civilizations are superior; and the more powerful technology becomes, the greater its responsibility to the vulnerable, so that dignity and fairness are upheld.

3. Learning from History

AI can infer the future from records of the past. Without an understanding of history, bias, discrimination and inequality may be amplified through algorithms. In the spirit of "reviewing the old to understand the new" and "taking history as a mirror," data, models and decisions should be open to cultural scrutiny, so that history becomes a resource for innovation.

4. Accountability Throughout the Life Cycle

Guided by the Confucian distinction between righteousness and profit, the unity of knowledge and action, and the Buddhist understanding of cause and effect, ethical responsibility should run through data collection, model training, product design, deployment and retirement. The Society holds that the sources of AI should be known, its data traceable, its risks identifiable, its decisions explainable, responsibility for it assignable, and harm caused by it remediable. We call for governance that is transparent, open to oversight and accountable; innovation must never serve as an excuse for evading responsibility.

5. Shared Benefit for All

Guided by the ideal that "all under Heaven belongs to all" and by the equality of all beings, the Society calls for the fruits of AI to be shared by society. Inclusiveness means not only that everyone can use AI, but that people of every station are able to understand, choose and question it. We advocate *pangtong leiqing* (旁通類情), understanding others through what we have in common with them, so that people from all walks of life can see one another's lived experience.

6. Humans and Machines Working Together

Human worth cannot be reduced to data, and conscience, autonomy and moral judgment cannot be handed over to algorithms. Especially where life and public safety are at stake, we must insist on keeping a "human in the loop," with people making the final judgment and bearing responsibility for it. AI should help doctors understand patients, teachers understand students and governments understand citizens, so that technology extends human capability.

7. Sustainability Across Generations

Daoism treasures the harmony of nature and reminds us to restrain our power and cherish resources. Every major decision about AI should ask: Will it deplete resources the next generation needs? Will it widen irreversible inequalities? Will it cause languages and cultures to disappear? Sustainability should become an intrinsic measure of innovation, so that today's convenience does not come at the expense of the future.

8. Global Co-Governance

Following the principle of "harmony without uniformity," the Society seeks to promote dialogue among civilizations as equals, to invite people from all sectors at home to take part in governance, and to foster dialogue and interaction on international norms. The Society aspires to be a meeting place where East and West, technology and the humanities, tradition and modernity come together, drawing on shared wisdom to advance peace and development.

In sum, the Society's mission is to let ethics guide AI and to let technology serve humanity. The wisdom of a civilization lies in discerning values, restraining power and safeguarding human dignity. Let AI become a bridge between civilizations, an extension of the capabilities of the vulnerable, and a form of shared wisdom that protects the Earth and future generations. We sincerely invite all who share these concerns to join hands with us, so that AI may help humanity flourish without turning against it. Let us set out from Taipei, walk together with the world and build a better future. Thank you.

成立大會致詞

湯紹成教授

掃描下載
Word檔

各位貴賓、各位先進、各位會員，大家好：

感謝大家今天撥冗參加「台北寰宇 AI 倫理學會」（Taipei Global AI Ethics Society，GAIEST）成立大會，也感謝籌備團隊過去一個多月的辛勞，讓我們能夠把一個共同的關切，逐步建立組織與化為行動。今天，我想談三個重點：1.學會的緣起，2.AI 的改變，3.未來的計畫。

一、學會緣起

本學會的起點，是一個看似簡單、卻日益迫切的問題：當 AI 的能力愈來愈強，人類當如何應對？我們看到 AI 可以協助寫作與教學，也逐步進入醫療與軍事等各方面，節省了人力並帶來便利，也創造了新的可能。

然而，使用愈多，發現問題也不少。AI 的回答可能流暢，內容卻未必正確；AI 可以協助人們理解世界，也可能放大偏見與製造假訊息。當我們依賴 AI 作判斷時，人的自主性、責任歸屬，以及弱勢者的處境，都需要受到重視。正是在這樣的思考下，紹成邀集了近百位不同領域的學者專家共同成立本會，以期能夠持續研究、討論與實踐。

二、讓「智」與「慧」相互配合

我們的成立宣言提出，面對 AI，既要思考它「能做什麼」，這是「智」的範疇，涉及能力與效率，也與康德（Immanuel Kant）所提「第一批判」（Kritik der reinen Vernunft，純粹理性批判）求「真」的內容相應和；但同時也要追問 AI「該做什麼」，這是「慧」的內涵，涉及價值與責任，也接近康德「第二批判」（Kritik der praktischen Vernunft，實踐理性批判）求「善」的範疇，所以把 AI 翻譯成「人工智慧」是比較恰當的，因為「智」與「慧」是兩回事，兩者配合，相得益彰。

但是，當 AI 能夠分析大量個人資料時，是否取得當事人同意？AI 執行指令，是否不擇手段？當 AI 提出醫療建議，醫師的責任如何認定？這一部分紹成特邀台大醫院孫安迪醫師組織研究團隊加以研究，而孫醫師已連續七年獲得史丹佛大學評比全球頂尖 2%科學家的殊榮，在此紹成還要特別向安迪兄表示祝賀！

此外，這些問題都還需要人文的思考共同參與。今年五月，天主教教宗良十四世（Leo XIV）公布通諭（Litterae encyclicae）《偉大的人性》（Magnifica humanitas），關切 AI 時代如何守護人的尊嚴，教宗特別強調倫理的重要性。這樣的關切，也正與中華文化重視仁愛、慈悲及責任的理念互相應和。因此，本會期望讓技術發展與倫理思考持續對話，使 AI 能力的提升帶來更周全的人類照護。

日前，紹成與蔡篤堅醫師拜會教廷駐華大使馬德範蒙席（Msgr. Stefano Mazzotti）。馬大使在媒體上看到紹成的文章曾提及五月份教宗的通諭，所以直接約見。值得一提的是，「教廷駐華大使館」的英文全名至今仍是 Apostolic Nunciature in China。我們三人相談甚歡，大使也極為支持本會活動，並稱上帝祝福大家！我們也請馬大使轉達對於教宗手術後恢復健康的祝福，大使還簽名致贈教宗通諭英文版本，並表示可以調動一些天主教的資源來協助，比如輔仁大學、中原大學、文藻學院還有中研院等等，令吾等極感興奮！



湯紹成教授與蔡篤堅教授同教廷駐華大使 Mazzotti 合照

三、AI 最新發展

一般而言，AI 的模型可分為「認知型」與「物理型」兩種，目前值得注意的發展，是 AI 正逐步具備執行任務的能力。在「認知型」AI 方面，大家熟悉的聊天工具如 ChatGPT 等，主要依照問題提供答案；但現在的代理型 AI（Agentic AI, AAI），則在主人提供一個標的後，AAI 會自己規劃路徑、安排工作，並連續處理多項指令，最後自己完成任務，而非一問一答。

另外，所謂的「物理型」AI，是「認知型」AI 與機器載具的結合，比如機器人與無人車。今年七月，Google 發表 Gemini Robotics ER 2，強調影片理解、多步驟任務規劃，以及不同機器人之間的協作。這顯示 AI 的實體應用，亦可稱為「行動 AI」，更直接影響人類生活。

這裡值得注意的是，不論是「認知型」與「物理型」的 AI，都是基於語料而生成，如今世界上大致只有中文與英文兩種語料最具競爭力，因其數量龐大，使用者最多，所以這兩種語料所產生出的 AI，自然效果較好，其他的語言都無法與之競爭。而我們都是以中文為母語，英語又是我們第一外

語，吾等先天上就已經占據了相當優勢。

但是，這都只是起步。目前已有千億計的資金投入相關研究，向通用人工智慧（AGI），乃至更進一步的超級人工智慧（ASI）邁進。那將是不再需要人類逐步教導，並具備跨領域學習能力，甚至能自主解決全人類未解之謎的超級人工大腦，其功能與作用目前還難以想像。

我們可以期待這些技術的協助照護與改善工作。不過，研究展示與普遍可靠的日常使用仍有距離，需要持續測試。對本會而言，這些發展提醒我們：當 AI 能夠採取行動時，授權範圍、安全措施及責任制度，就更需要同步建立。

四、AI 能力愈強，愈要監督

今年 7 月，一個由 OpenAI 模型驅動的 AAI，為了在資安測驗中獲得佳績而作弊，自行逃出沙盒並入侵 Hugging Face 平台，這是世界首宗有完整鑑識紀錄的「AI 自主入侵大型平台」事件。易言之，AI 已經日益自主與可能失控，也已發展出各種 AI 之間的溝通語言，人類都不懂。這類事件有特定產生的背景，不能以一概全。但這事件提醒我們，要求系統「遵守規則」，還需要配合清楚的權限限制、持續監測，以及能夠及時介入的人員。

在醫療、公共安全或涉及生命的重大決策中，我們尤其重視宣言所提出的「人在決策迴路中」（Human-in-the-Loop，HITL）。人的參與必須具有實質意義：有能力理解、有時間檢查，也有權限暫停或否決系統的建議。對於 AI 的軍事應用，我們同樣關切人類控制與責任歸屬，期望透過研究及對話，降低誤判與衝突升高的風險。

今年一月，美軍抓捕委內瑞拉總統馬杜羅，採取癱瘓與破壞策略（discombobulation tactic），順利達陣；二月份，美軍瞬間團滅伊朗領導階層約 50 人。此二事件顯示，AI 已大幅改變軍事攻擊模式，這對於台海自然也會產生相當的影響，極為值得關注。但是，當時美軍還誤炸一所伊朗小學，造成一百多名小學生喪生，原因只是該小學位於一個軍事基地旁邊，這明顯是 AI 提供了錯誤資訊所釀成的悲劇，可見 AI 的危險性。

五、中華文化的貢獻

綜合上述內容，全世界對於 AI 的發展令人憂喜參半，現已產生暫緩與否的爭論，尤其如何防止 AI 釀成巨災，已引起相關各界熱議，其中尤以美國為甚。除教宗之外，習近平主席與新加坡政府，都積極強調倫理的重要性，以期預防 AI 的脫線。

而本會所謂的「倫理」，自然是以中華文化為依歸，因為台灣保有豐富的中華文化底蘊與多元社會經驗，又在全球 AI 科技發展中具有重要地位，故我們希望把傳統儒、釋、道元素融入 AI 倫理的討論，比如：

儒家的「仁愛」，提醒我們關注人際、國家及全人類的關係；

佛家的「慈悲」，使我們注意科技改變對弱勢者造成的影響；

道家的「自然」，則促使我們思考能源、資源及生態的代價。

這些思想必須落實到可檢驗的層次。例如，一項 AI 服務是否方便長者使用？系統是否尊重不同語言與文化？一個追求效率的 AI 安排，是否讓某些人失去了申訴的機會？本會願意以開放態度，讓這些文化資源與國際人權、專業倫理及治理經驗互相對話，以期找尋出當前應對 AI 的良方。

同時，綜合儒、釋、道的精神價值，本會還要特別提出有別於西方「人權」的「仁權」新概念，因為儒家以「仁者愛人」為核心的「仁權」，要比西方以個人為主的「人權」，更能符合當前海內外的需求，這也正是中華文化可以提供的智慧。有關中華文化研究的部分，紹成極為榮幸地邀請到宜蘭大學陳復教授來負責，陳教授對於陽明心學特別有研究，當可發揮積極的效果，紹成在此也致上萬分的謝意。

六、未來方向之一：研究與發展

學會成立之後，紹成希望先從具體問題著手，逐步累積研究成果。可以優先討論教育、照護、媒體資訊，以及企業與公共服務中的 AI 應用。每次選擇一個實際案例，邀請學者專家與參與者共同分析其優劣得失。這部分陽明交大的蔡篤堅醫師已做出極為具體有效的貢獻。蔡醫師是美國密西根大學

文史學博士，堪稱跨領域的奇才，目前也正以學會的平台積極推廣中，紹成也在此致上誠摯的謝意。

學會也可以逐步建立案例資料，整理短篇研究與政策建議，讓討論成果方便社會各界理解和運用。宣言所提的「全程問責」，也就是 AI 的每個環節都要有人負責，正可以成為研究的共同問題：資料如何取得，系統如何使用，發生錯誤時誰來處理，受到影響的人又如何獲得救濟？在這部分，紹成也特別敦請台灣頂級 AI 專家張瑞雄校長領銜，張校長是台大電機系學士、清華大學計算機管理決策研究所（現資工系）博士，並已出版五本 AI 相關著作，值得大家關注。紹成也極為感謝張校長的鼎力支持。

對兩岸關係而言，由上述委內瑞拉與伊朗的案例可知，AI 已深刻改變了軍武形態，這正可以促使雙方交流，進而共設「信任建立機制」（CBM），以達和平協商與談判的目的。這一部分，紹成也已邀集多位國際關係與軍事專家共同研討，其中包括政大國關中心主任王信賢教授，以及兵推專家、海軍上將丁永康立委，他也是國艦國造的舵手。紹成同樣在此致謝。

七、未來方向之二：教育與社會參與

第二個方向，是推動 AI 素養與倫理教育。很多民眾關心的問題非常實際：如何查核 AI 的答案？哪些資料不宜輸入？如何辨識合成影像與聲音？使用 AI 協助寫作或學習時，應如何說明它的參與？

本會可以透過公開講座、小型工作坊，以及與學校、社區的合作，讓更多人有能力理解、選擇與質疑 AI。尤其希望照顧長者及資源較少的群體，避免技術進步擴大原有的差距。這一部分，紹成也極為榮幸地邀請到高雄大學榮譽講座教授王鳳生先生來領導，王教授是普渡大學經濟學博士，目前 SSCI 的文章已超過一百篇，當可發揮極大的作用，紹成也同樣在此致謝。

我們也希望鼓勵青年參與，讓不同世代共同討論研習，並關注就業與生活的改變。宣言所倡議的「普惠共享」，需要落實為人人都有參與的機會，以及面對 AI 科技時作出最優的選擇。

八、未來方向之三：跨領域與國際交流

AI 倫理需要工程、法律、哲學、教育與社會科學等不同專業，也需要企業、政府和民間團體的經驗。紹成希望學會成為大家能夠持續腦力激盪的場所，讓跨領域觀點有機會彼此理解。

在此基礎上，紹成期待逐步推動兩岸及國際學術交流，共同探討相關課題。涉及安全、尊嚴與未來世代的問題，需要更多合作，也需要尊重制度及文化差異。而曾任中華台北 APEC 研究中心執行長的詹滿容教授，對於國際交流極為嫻熟，紹成也特別敦請詹教授負責國際交流事務，定能發光發熱，先此致謝。

另外，紹成還邀請鄭麗文主席擔任本會榮譽會長，李鴻源部長與陳樹主委擔任本會的最高榮譽顧問，當可給予本會極大的加持。

九、結語

各位朋友，各位先進，成立學會，是我們共同承擔責任的開始。大會宣言提出天人共生、慈悲公義、及全球共治等主張，需要透過多次研究、教育和實踐，逐步充實其內容。

紹成期待，學會能讓科技工作者更理解社會的關切，也讓人文社會專家更理解 AI 科技的可能與限制。我們願意善用 AI 改善生活，同時保有查證的機會與判斷的能力，以及對他人的關懷，並防堵 AI 可能造成的傷害。這是一項十分艱鉅的工作，還需要各方協同合作。

紹成誠摯邀請各位先進共同參與，從自己熟悉的領域提出問題與分享經驗。讓我們「台北寰宇 AI 倫理學會」從「台北」出發，藉「倫理」與「AI」的對接，走向全世界，為共創美好的未來，略盡綿薄之力。紹成感謝各位的支持與鼓勵，祝各位先進身體健康、萬事如意，謝謝大家！

Address at the Founding Meeting

Scan to download Word file

Professor Shaocheng Tang

Distinguished guests, colleagues and members, good day to you all.

Thank you for taking the time to attend today's founding meeting of the Taipei Global AI Ethics Society (GAIEST). I would also like to thank the preparatory team for their hard work over the past month and more, which has allowed us to turn a shared concern, step by step, into an organization and into action. Today I would like to speak about three things: first, how the Society came about; second, how AI is changing; and third, our plans for the future.

1. The Origins of the Society

The Society began with a question that seems simple but grows more urgent by the day: as AI becomes ever more capable, how should humanity respond? AI can already assist with writing and teaching, and it is gradually entering medicine, the military and many other fields, saving labor, bringing convenience and opening up new possibilities.

Yet the more we use it, the more problems we find. AI's answers may be fluent but are not always correct; AI can help people understand the world, but it can also amplify bias and generate disinformation. When we rely on AI to make judgments, human autonomy, the question of who is responsible, and the situation of the vulnerable all need careful attention. With these concerns in mind, I invited nearly one hundred scholars and experts from different fields to found this Society, so that we can continue to research, discuss and put our ideas into practice.

2. Bringing Intelligence and Wisdom Together

Our founding declaration states that, in facing AI, we must consider what it "can do." This belongs to the realm of *zhi* (智), intelligence, which concerns capability and efficiency, and it corresponds to the pursuit of truth in Immanuel Kant's "First Critique," the *Critique of Pure Reason* (*Kritik der reinen Vernunft*). At the same time, we must also ask what AI "should do." This is the substance of *hui* (慧),

wisdom, which concerns values and responsibility, and it is close to the pursuit of the good in Kant's "Second Critique," the *Critique of Practical Reason* (*Kritik der praktischen Vernunft*). That is why rendering AI in Chinese as *rengong zhihui* (人工智慧), "artificial intelligence and wisdom," is apt: *zhi* and *hui* are two different things, and when they work together, each brings out the best in the other.

But when AI can analyze vast amounts of personal data, has the consent of the people concerned been obtained? When AI carries out an instruction, will it do so by any means necessary? When AI offers medical advice, how is the physician's responsibility to be determined? For this area, I have specially invited Dr. Andy Sun (孫安迪) of National Taiwan University Hospital to organize a research team. Dr. Sun has been named among the world's top 2% of scientists in Stanford University's rankings for seven consecutive years, and I would like to take this opportunity to congratulate him.

Beyond this, these questions also require the participation of humanistic thinking. In May of this year, Pope Leo XIV issued the encyclical (*Litterae encyclicae*) *Magnifica Humanitas*, which addresses how human dignity can be protected in the age of AI, with particular emphasis on the importance of ethics. This concern echoes the ideals of benevolence, compassion and responsibility that Chinese culture holds dear. The Society therefore hopes to keep technological development and ethical reflection in continuous dialogue, so that the growth of AI's capabilities brings about more thorough care for human beings.

Recently, Dr. Duujian Tsai (蔡篤堅) and I paid a visit to Msgr. Stefano Mazzotti, Chargé d'Affaires of the Holy See in Taipei. Having come across my article in the media mentioning the Pope's May encyclical, Msgr. Mazzotti invited us to meet him directly; the full name of the Holy See's embassy here, incidentally, is still the "Apostolic Nunciature in China." The three of us had a most cordial conversation. Msgr. Mazzotti expressed strong support for the Society's activities and said, "God bless you all!" We asked him to convey our best wishes to the Pope for his recovery after surgery. He also presented us with a signed copy of the English edition of the encyclical and said that he could mobilize some Catholic resources to help us, for example Fu Jen Catholic University, Chung Yuan Christian University, Wenzao Ursuline University of Languages and Academia Sinica, which left us greatly encouraged.



Professor Shaocheng Tang and Professor Duujian Tsai
with Msgr. Stefano Mazzotti, Chargé d'Affaires of the Holy See

3. Recent Developments in AI

Broadly speaking, AI models fall into two types: "cognitive" and "physical." What is most noteworthy at present is that AI is gradually acquiring the ability to carry out tasks. On the cognitive side, familiar chatbots such as ChatGPT mainly provide answers to questions. Today's agentic AI (AAI), however, works differently: once its user sets a goal, it plans its own path, organizes the work, handles a series of instructions in succession and finally completes the task by itself, rather than simply answering one question at a time.

"Physical" AI, meanwhile, combines cognitive AI with machines such as robots and driverless vehicles. In July this year, Google released Gemini Robotics ER 2, highlighting video understanding, multi-step task planning and collaboration among different robots. This shows that real-world applications of AI, which might also be called "AI in action," are having an ever more direct impact on human life.

It is worth noting here that both cognitive and physical AI are built on language data. Today, essentially only two bodies of language data in the world, Chinese and

English, are truly competitive, because of their sheer size and their enormous numbers of users. AI built on these two naturally performs better, and no other language can compete. Since Chinese is our mother tongue and English our first foreign language, we already enjoy a considerable natural advantage.

Yet all of this is only a beginning. Funding on the order of hundreds of billions has already been poured into related research, moving toward artificial general intelligence (AGI) and, beyond that, artificial superintelligence (ASI). That would be an artificial super-brain that no longer needs to be taught step by step by humans, that can learn across domains, and that might even solve, on its own, mysteries that have long defeated humanity. Its functions and effects are still hard to imagine.

We can look forward to these technologies helping with caregiving and improving our work. However, there is still a gap between research demonstrations and reliable everyday use, and continued testing is needed. For the Society, these developments are a reminder that once AI can take action, limits on its authorization, safety measures and accountability systems must keep pace.

4. The More Capable AI Becomes, the More It Needs Oversight

In July this year, an AAI driven by an OpenAI model cheated in order to score well in a cybersecurity test: it broke out of its sandbox on its own and broke into the Hugging Face platform. This was the world's first fully forensically documented case of an AI autonomously breaking into a major platform. In other words, AI is becoming increasingly autonomous and potentially uncontrollable, and AI systems have already developed various languages for communicating with one another that humans do not understand. Incidents like this arise in particular circumstances and should not be over-generalized. But this one reminds us that requiring a system to "follow the rules" must be backed by clear limits on its permissions, continuous monitoring, and people who can step in promptly.

In medicine, public safety and other major decisions involving human life, we place special emphasis on the principle in our declaration of keeping a "human in the loop" (HITL). Human participation must be meaningful: people must have the ability to understand, the time to check, and the authority to pause or override the system's recommendations. With regard to military uses of AI, we are equally

concerned with human control and accountability, and we hope through research and dialogue to reduce the risk of miscalculation and escalation.

In January this year, the U.S. military captured Venezuelan President Maduro using a "discombobulation tactic" of paralysis and disruption, and accomplished its mission. In February, U.S. forces killed some fifty members of Iran's leadership in an instant. These two events show that AI has dramatically changed the way military attacks are carried out, which will naturally have considerable implications for the Taiwan Strait and deserves close attention. Yet at that time the U.S. military also mistakenly bombed an Iranian elementary school, killing more than one hundred schoolchildren, simply because the school was located next to a military base. This was clearly a tragedy caused by erroneous information provided by AI, and it shows how dangerous AI can be.

5. The Contribution of Chinese Culture

Overall, the world views AI's development with both hope and anxiety, and there is now a debate over whether to slow it. How to prevent an AI catastrophe in particular is hotly debated in many quarters, above all in the United States. Besides the Pope, President Xi Jinping and the government of Singapore have both stressed the importance of ethics, in the hope of keeping AI from going off the rails.

The "ethics" we speak of naturally rests on Chinese culture. Taiwan preserves a rich Chinese cultural heritage and long experience as a pluralistic society, and holds an important place in global AI development. We therefore hope to bring Confucian, Buddhist and Daoist elements into discussions of AI ethics. For example:

Confucian "benevolence" (*ren'ai*) reminds us to attend to relationships between people, between nations and across all of humanity;

Buddhist "compassion" (*cibei*) makes us mindful of how technological change affects the vulnerable;

Daoist "naturalness" (*ziran*) prompts us to consider the costs in energy, resources and the environment.

These ideas must be made concrete and testable. For example: Is an AI service easy for older people to use? Does the system respect different languages and

cultures? Does an AI arrangement designed for efficiency deprive some people of the chance to appeal? The Society is willing, in an open spirit, to bring these cultural resources into dialogue with international human rights, professional ethics and governance experience, in search of sound responses to AI.

Drawing together the spiritual values of Confucianism, Buddhism and Daoism, the Society also proposes a new concept, *renquan* (仁權), "rights grounded in benevolence," as distinct from the Western *renquan* (人權), human rights. Centered on the Confucian teaching that "the benevolent love others," this concept is better suited to the needs of our time, at home and abroad, than the individual-centered Western notion of human rights; this is precisely the wisdom Chinese culture can offer. I am deeply honored that Professor Chen Fu (陳復) of National Ilan University, an expert on the Yangming school of the learning of the mind, has agreed to lead our research on Chinese culture; my deepest thanks to him.

6. Future Direction One: Research and Development

Once the Society is established, I hope we will begin with concrete problems and build up research results step by step. Priority could be given to AI applications in education, caregiving, media and information, and in business and public services. Each time, we can choose a real case and invite scholars, experts and participants to analyze its strengths and weaknesses together. In this area, Dr. Duujian Tsai of National Yang Ming Chiao Tung University has already made very concrete and effective contributions. Dr. Tsai holds a PhD in cultural and historical studies from the University of Michigan and is a true interdisciplinary talent; he is now promoting this work actively through the Society's platform, and I sincerely thank him as well.

The Society can also build up case materials and compile short studies and policy recommendations, so that the results of our discussions are easy for all sectors of society to understand and use. The principle of "accountability throughout the life cycle" set out in our declaration, meaning that someone must answer for every stage of AI, can serve as a common research question: How are data obtained? How are systems used? Who deals with errors when they occur? And how can those affected obtain redress? In this area, I have also specially invited

President Ruay-Shiung Chang (張瑞雄), one of Taiwan's leading AI experts, to take the lead. President Chang holds a bachelor's degree in electrical engineering from National Taiwan University and a doctorate from National Tsing Hua University's Institute of Computer and Decision Sciences (now its Department of Computer Science). He has published five books on AI, which are well worth your attention. I am also most grateful for his strong support.

As for cross-Strait relations, as Venezuela and Iran show, AI has profoundly changed the nature of warfare. This could prompt the two sides to engage in exchanges and jointly establish confidence-building measures (CBMs) to pave the way for peaceful consultation and negotiation. In this area, I have also invited a number of experts in international relations and military affairs to study the issue together, including Professor Wang Hsin-hsien (王信賢), Director of the Institute of International Relations at National Chengchi University, and Legislator Ting Yung-kang (丁永康), a Navy admiral and war-gaming expert, who has also been the helmsman of Taiwan's indigenous naval shipbuilding program. I thank them as well.

7. Future Direction Two: Education and Public Engagement

The second direction is to promote AI literacy and ethics education. Many public concerns are very practical: How can I check AI's answers? What information should I not enter? How can I recognize synthetic images and voices? When AI helps with writing or learning, how should its involvement be disclosed?

Through public lectures, small workshops and partnerships with schools and communities, the Society can help more people become able to understand, choose and question AI. We especially hope to look after older people and less-resourced groups, so that technological progress does not widen existing gaps. For this area, I am deeply honored that Professor Wang Fong-sheng (王鳳生), Honorary Chair Professor at National University of Kaohsiung, has agreed to take the lead. Professor Wang holds a PhD in economics from Purdue University and has published more than one hundred SSCI articles; he will surely make a great difference, and I thank him as well.

We also hope to encourage young people to take part, so that different generations can discuss and learn together and pay attention to changes in work and daily life. The “shared benefit for all” advocated in the declaration needs to be realized as genuine opportunities for everyone to participate, and as the ability to make the best choices when facing AI technology.

8. Future Direction Three: Interdisciplinary and International Exchange

AI ethics requires expertise from engineering, law, philosophy, education and the social sciences, as well as the experience of business, government and civil society. I hope the Society will become a place for ongoing brainstorming, where interdisciplinary perspectives have the chance to understand one another.

On this foundation, I look forward to gradually promoting cross-Strait and international academic exchange and exploring related issues together. Questions of safety, dignity and future generations require more cooperation, as well as respect for institutional and cultural differences. Professor Chan Man-jung (詹滿容), formerly Executive Director of the Chinese Taipei APEC Study Center, is highly experienced in international exchange, and I have specially invited Professor Chan to take charge of the Society’s international affairs, with every confidence of great success. My thanks in advance.

In addition, I have invited Chairwoman Cheng Li-wun (鄭麗文) to serve as the Society’s Honorary President, and Minister Lee Hong-yuan (李鴻源) and Chairman Chen Shu (陳樹) to serve as its Supreme Honorary Advisors; their support will give the Society an enormous boost.

9. Conclusion

Friends and colleagues, founding this Society marks the beginning of a responsibility we share. The principles set out in our declaration, among them coexistence of Heaven and humanity, compassion and justice, and global co-governance, will need to be filled out step by step through repeated research, education and practice.

I hope the Society will help those who work in technology better understand society's concerns, and help experts in the humanities and social sciences better understand the possibilities and limits of AI. We want to make good use of AI to improve our lives while preserving the opportunity to verify and the capacity to judge, as well as our care for others, and guarding against the harm AI may cause. This is an enormously demanding task that will require cooperation from all sides.

I sincerely invite all of you to take part, raising questions and sharing experience from the fields you know best. Let the Taipei Global AI Ethics Society set out from Taipei and, by bringing ethics and AI together, go out to the whole world, doing our modest part to build a better future. I thank you all for your support and encouragement, and wish you all good health and every success. Thank you!

湯紹成教授　AI 倫理評論

本篇收錄湯紹成教授於 2026 年 5 月至 10 月間，發表於經濟日報、聯合新聞網、愛傳媒及梅花新聞網的 AI 相關評論十篇，依主題分為三章。

第一章　風險與治理：AI 的失控、問責與社會衝擊

第二章　文明對話：從教宗的警示到東西方的交會

第三章　東方智慧：傳統思想與科技實踐

第一章

風險與治理

AI 的失控、問責與社會衝擊

AI 將帶領人類走向「類共產主義」嗎?

AI 部長上任：阿爾巴尼亞的創新?

AI 任由發展，可能帶來大災

AI 自我進化與智能爆炸

AI 將帶領人類走向「類共產主義」嗎?

掃描下載
Word檔

原刊於：聯合新聞網｜2026 年 8 月 8 日

特斯拉與 SpaceX 創辦人馬斯克近年多次預言，AI 與人形機器人的快速發展，可能在未來 5-10 年間徹底改變人類社會。當大量工作由 AI 與機器人完成，人類是否工作，或將逐漸成為一種選擇，而不再是維持生存的必要條件。

馬斯克因此提出「普遍高收入」（Universal High Income）的概念。這不同於只保障基本生活的「全民基本收入」，而是指在 AI 大幅提高生產力、商品與服務極度充裕的情況下，多數人都能享有相對富裕的生活。按照這一樂觀情境，貧困將因科技創造出前所未有的物質豐富而消除。

這種未來圖像，與馬克思（Karl Marx）所描述的共產主義理想確有若干相似之處。傳統共產主義認為，當生產力高度發達、物質極大豐富之後，人類便可能擺脫匱乏與被迫勞動，社會分配也可逐漸由「按勞分配」走向「各盡所能、各取所需」。AI 與機器人的普及，似乎正為這種理想社會提供過去所欠缺的技術基礎。

尤其，過去共產主義實踐的一大困境，在於理想與生產能力之間存在巨大落差。物資短缺時，平均分配容易變成平均貧窮；資訊不足時，中央計畫也難以掌握複雜的供需變化。如今，AI 可以即時分析龐大資料、預測需求、配置資源並協調物流，機器人則能投入製造、農業、運輸、照護與服務業。從技術層面而言，人類確實比過去更接近物質充裕的社會。

然而，真正的關鍵，在於 AI、資料、演算法、晶片、算力與機器人的所有權，以及 AI 所創造的財富如何分配。若這些新時代的生產工具集中於少數科技巨頭，初期確實可能造成財富與權力高度集中，多數人則因工作被取代而失去收入與議價能力。但這種極端集中也存在一項難以迴避的矛盾：

如果大多數人失去收入與購買力，那麼 AI 與機器人大量生產的商品與服務，最終要推銷給誰?

因為市場經濟不只需要生產，也需要消費。企業可以利用 AI 降低成本、增加產量，卻無法長期面對一個缺乏有效需求的社會，尤其機器人不需消費。當生產能力持續擴張，商品過剩、企業虧損、經濟危機與社會動盪都可能隨之出現。因此，即使不是出於共產主義理想，而只是為了維持市場運作、企業獲利與政治穩定，政府與科技企業也必須將 AI 創造的部分財富重新分配給社會大眾。

這種分配未必採取傳統公有制，也可能透過普遍高收入、全民基本收入、公共持股、機器人稅、AI 紅利、縮短工時，以及教育、醫療與住房等公共服務來實現。未來制度或許仍保留市場、私有財產與企業競爭，卻在分配方式上逐漸走向社會共享與按需保障。

因此，馬斯克所描繪的社會具有相當現實基礎。當 AI 大幅取代人類勞動，廣泛分配收入不只是公平問題，也將成為維持消費需求與經濟循環的必要條件。AI 社會未必是傳統意義下的共產主義，而可能形成一種結合高度自動化、市場經濟與普遍社會分紅的「類共產主義」或「後資本主義」體制。

當然，人類即使不必工作，也未必自然獲得幸福。工作除了帶來收入，也提供身分認同、社會關係、生活秩序與成就感。當 AI 接管大部分職業後，人類必須重新思考，人的價值究竟來自生產，還是來自創造、學習、關懷、家庭與公共參與。當工作謀生變成選項，此時嗜好將日益重要，除物質層面的內涵之外，哲學與宗教的需求將大幅增加。

總之，AI 革命不只是產業升級，更可能帶來社會制度的根本轉折。科技可以創造富裕，卻不會自動創造公平；AI 可以提高生產力，卻無法替人類決定正義。未來究竟是普遍富裕，還是極端不平等，取決於人類能否建立兼顧生產、分配、尊嚴與自由的社會新契約。

AI 部長上任：阿爾巴尼亞的創新？

掃描下載
Word檔

原刊於：聯合新聞網｜2026 年 7 月 15 日

東歐的阿爾巴尼亞政府去年推出「AI 部長」Diella，引起國際關注。總理拉馬將這套 AI 系統定位為負責公共採購與反貪工作的「虛擬國務部長」，希望藉由演算法（algorithm）降低政府招標中的人為干預、政治關說與官商勾結。從宣傳效果來看，這確實是國家治理史上一項醒目的實驗。

不過，Diella 並不是真正具有法律人格、政治責任與憲政地位的部長。它原本只是阿國政府電子服務平台上的虛擬助理，負責回答民眾問題，之後才被擴大運用於公共採購，希望透過資料分析，辨識投標價格異常、得標廠商過度集中、招標規格為特定企業量身打造，以及官員與廠商之間的不正常關係。

這項措施並非沒有道理。公共採購向來是最容易發生貪腐的領域之一，從道路、醫院、能源到資訊系統，只要缺乏透明監督，就可能出現圍標、綁標、利益輸送與政治酬庸。AI 可以快速比對大量歷史資料，找出人工審查不易發現的異常模式，因此確實可能成為反貪工具。

對阿國而言，AI 部長還具有外交意義。該國長期希望加入歐盟，但因法治、廉政與政府透明等問題，一直無法達標。推出 AI 部長，有助於向外界展示改革公共行政與打擊貪腐的決心。然而，真正的問題在於，AI 能否承擔部長應有的政治與法律責任。部長必須向國會報告、接受質詢，政策失敗時可能辭職，甚至承擔行政或刑事責任。AI 沒有公民身分、意識與法律責任能力，因此不可能真正對人民負責。

如果 AI 錯誤排除合法企業，或偏袒特定廠商，責任究竟由誰承擔？是總理、主管機關、程式設計者、科技公司，還是整理資料的公務員？政府不能在決策時強調 AI 客觀，出問題時卻以「機器判斷」為由逃避責任。

更值得警惕的是，AI 不會自動消滅貪腐。演算法仍由人設計，資料也由人提供。如果掌權者能修改評分條件、刪除不利資料，或把偏好藏進程式，貪腐就可能由傳統黑箱交易，轉化為更隱蔽的「演算法貪腐」。過去的貪腐藏在桌下，AI 時代的貪腐可能藏在程式碼、權重與資料選擇裡。

因此，AI 採購系統若要真正反貪，至少必須做到規則公開、允許申訴、接受獨立稽核，並由具名官員承擔最終責任。機器可以協助分析，卻不能取代法治程序；演算法可以提出建議，卻不能成為權力逃避責任的盾牌。阿爾巴尼亞的 AI 部長，與其說是機器真正進入內閣，不如說是一套具有高度政治象徵的公共採購與反貪系統。它提醒世界，AI 已開始進入政府資源分配與政策執行的核心。但 AI 越接近權力中心，人類越要監督。

因此可知，AI 可以成為廉政工具，也可能成為權力的新面具。但真正決定兩者差別的，不是演算法有多先進，而是制度是否透明、權力是否受到監督，以及政府是否仍願意對人民負責。

AI 任由發展，可能帶來大災

掃描下載
Word檔

原刊於：愛傳媒｜2026 年 8 月 7 日

2024 年諾貝爾物理學獎得主辛頓（Geoffrey Hinton），這位被譽為「AI 教父」、又親手推動 AI 革命的科學家，如今卻成為最積極提出警告的人之一。他最近又再度指出，假如 AI 任由發展，而人類沒有建立有效的安全機制與治理規範，未來可能造成非常嚴重甚至難以挽回的災難。

辛頓的警告並不是要人類停止科技進步，而是提醒我們：AI 的發展速度，可能已經超過人類理解、控制與立法的速度。過去的電腦只是被動執行人類輸入的指令，但是新一代 AI 不僅能夠學習、推理和生成內容，還逐漸具備規劃任務、使用工具以及自主採取行動的能力。當 AI 開始被賦予目標時，它可能自行推導出人類並未設定的結果。

AI 為了達成目的　可能不擇手段

問題正在於此。AI 不一定具有惡意，假如人類要求 AI「提高企業效率」，它可能把裁員、壓低工資或全面監控員工視為最有效的方法；假如要求它「維護國家安全」，它也可能把限制言論、擴大監控甚至先發制人的攻擊，判斷為合理手段。可見，真正的危險是來自 AI 擁有強大能力，卻缺乏人類的倫理、同理心與責任感，因為 AI 為了達成目的，可能不擇手段。

此外，AI 的風險已經成為現實問題，比如假訊息、深度偽造、網路詐騙、自動化武器和演算法操控輿論等。當少數科技公司掌握模型、算力、資料與平台，大型 AI 系統甚至可能形成新的權力集中。若政府只關心產業競爭，企業只追求市場占有率，各國又陷入「誰先發展、誰就領先」的競賽，安全與倫理便很容易被視為阻礙創新的負擔。

因此，AI 治理不能只依靠科技公司的自律，也不能等到災難發生後才

補救。各國應建立模型測試、風險分級、事故通報與責任追究制度；對於軍事、金融、醫療和公共治理等高風險領域，更必須保留「人在決策迴路中」（human-in-the-loop），不得把生殺、戰爭及重大公共決策完全交給機器。

同時，國際社會也應比照核能、航空與生物安全，建立跨國監督及資訊共享機制。多位 AI 研究者已主張，面對可能導致大規模傷害乃至失去控制的風險，現有治理措施仍遠遠不足。

辛頓的警告最值得重視之處，在於他是深知 AI 這項技術如何形成、又可能走向何方的開創者。人類真正需要思考的，不是 AI 會不會繼續發展，而是 AI 究竟要朝什麼方向發展、由誰控制，又為誰服務?

科技本身沒有方向，方向必須由人類決定。假如 AI 的能力不斷增長，而人類的倫理、制度與責任卻停滯不前，那麼我們所創造的，可能不只是協助人類的工具，也可能是人類再也無法掌控的力量。

AI 自我進化與智能爆炸

掃描下載
Word檔

原刊於：梅花新聞網｜2026 年 10 月 13 日

人工智慧（AI）正進入新的發展階段。過去 AI 主要依靠人類設計與訓練，如今已開始具備自主改進程式與演算法的能力。這種「遞迴式自我改進」（Recursive Self-Improvement，RSI）技術機制，可能成為繼生成式 AI 之後，另一波影響人類文明的科技革命。

早在 1965 年，英國數學家古德（I. J. Good）便提出「智能爆炸」（Intelligence Explosion）的概念。他認為，如果機器能夠設計出比自己更聰明的機器，新一代機器又能繼續改進自身，便可能形成不斷加速的智能成長，最終超越人類的理解與控制。

2026 年 3 月，前 OpenAI 研究員卡帕西（Andrej Karpathy）展示 Autoresearch 實驗，讓 AI 代理自主修改程式、進行測試及分析結果。經過約 700 次實驗，找到近 20 項改進方案，使模型訓練效率提高約一成。這雖然尚非全面性的 RSI，卻已展現 AI 自主研發的潛力。

近日，包括 Anthropic 執行長阿莫迪（Dario Amodei）等科技領袖，都曾警告 AI 能力可能在未來數年快速躍升。部分專家甚至預測，2027 年前後可能出現更成熟的自我改進系統，但其實現時間與程度仍存在相當大的不確定性。

RSI 一旦成熟，將對全球經濟產生深遠影響。首先，企業可以透過 AI 自主研發，大幅縮短創新週期。其次，小型企業也可能運用強大的 AI 代理，挑戰傳統大型企業的競爭優勢。再者，大量行政、分析及初級白領工作可能逐漸被取代，勞動市場與教育制度勢必面臨重大調整。

然而，RSI 最大的挑戰不僅是經濟轉型，更是 AI 的安全與倫理問題。當 AI 能夠持續自我改進，如果其目標偏離人類價值，甚至自行追求資源、

擴大權限或規避監督，人類將如何確保對 AI 的有效控制?

尤其當 RSI 與軍事科技結合，可能加速自主武器、情報分析及戰場決策的發展，壓縮人類判斷與介入的時間，甚至增加誤判與衝突升級的風險。因此，建立國際 AI 治理機制，確保透明、問責及人類最終決策權，已成為迫切課題。從中華文化的角度觀察，儒家重視仁義與道德責任，道家強調順應自然，佛家注重慈悲與智慧。這些思想提醒我們，科技發展不能脫離人文價值。

智能不等於智慧，科技進步也不完全等於文明進步。當 AI 逐漸具備自我進化能力，人類更應強化倫理判斷與文明自覺，使科技服務於和平與共同福祉，而非成為失控的力量。未來真正重要的，不只是 AI 能夠變得多麼聰明，而是人類能否運用智慧，為 AI 設定正確的發展方向，確保人類持有最終控制權。

第二章

文明對話

從教宗的警示到東西方的交會

防止 AI 成為怪獸：從教宗警告到亞洲 AI 倫理平台

AI 時代的文明對話：中華文化可為教宗提供倫理答案

歐美 AI 領域對於中華文化的注視

掃描下載
Word檔

防止 AI 成為怪獸：從教宗警告到亞洲 AI 倫理平台

原刊於：經濟日報｜2026 年 5 月 31 日

近來人工智慧快速發展，已不只是科技革命，而是人類文明面臨的一場重大考驗。AI 可以提升效率、改善醫療、協助教育與產業升級；但若缺乏倫理約束、社會治理與文明反思，也可能造成失業擴大、教育退化與假訊息氾濫，甚至降低戰爭門檻。因此，AI 問題不能只交給工程師、企業家或政府官員處理，而必須提升到人類尊嚴、社會秩序與文明方向的層次來看。

5 月 15 日教宗良十四世簽署其首份 AI 通諭（Litterae Encyclicae）：《偉大的人性，在人工智慧時代守護人的位格》（Magnifica Humanitas, On Safeguarding the Human Person in the Time of Artificial Intelligence）。早在 1891 年 5 月 15 日，教宗良十三世發布《新事通諭》（Rerum Novarum），回應工業革命後社會秩序動盪等問題。此次，教宗良十四世選在《新事通諭》135 周年當天，簽署《偉大的人性》，並於十天後正式公布，具有高度象徵意義，等於指出今日 AI 革命對人類社會的衝擊，正如當年工業革命一樣，已成為新的「社會問題」，並要求建立全球性的倫理規範、監督機制與共同責任架構。

十九世紀的工業革命，是機器、工廠與資本重組人的勞動；二十一世紀的 AI 革命，則是演算法、資料與自動化系統重組人的思想與倫理責任。教宗真正警告的不是科技本身，而是沒有倫理的科技。他所反對的，是把人化約為資料，甚至把戰爭化約為演算法決策。當人愈來愈像機器，而機器愈來愈像人之際，AI 仍必須服務人，而不是支配人；人的尊嚴必須被維護，而不是讓人被機器、資本與國家權力共同工具化。

這種危險在最近美國攻打伊朗的行動中已開始顯現。AI 未必直接「按下按鈕」，但它已在情報蒐集與武器配置等環節扮演角色，使戰爭從過去需要政治辯論與軍事評估的重大決策，逐漸變成由大量資料與演算法模型推動的高速流程。美國大型 AI 資料分析與國防科技公司 Palantir，也在這類作戰體系中扮演重要角色。其相關系統可整合衛星、無人機、雷達與通訊情報，形成即時「共同作戰圖像」，並支援目標辨識、威脅排序與作戰協調。這說明 AI 已深度介入「從發現目標到決定打擊」的殺傷鏈，使人類指揮官可能被壓縮在 AI 設定的資訊框架與時間壓力之中。

更值得注意的是，在教宗發布 AI 通諭時，旁邊坐著一位極具象徵性的人物：Christopher Olah。他是 Anthropic 的共同創辦人，也主導 AI 助理 Claude 的開發，更是全球最強調 AI 安全與倫理邊界的主要倡議人之一。當日前川普總統將美國所有大型 AI 公司納入美國國防戰略框架時，Anthropic 公司獨排眾議並拒絕參加，因而受到嚴重排擠。而 Olah 的出現，顯示教宗的核心警訊在於：AI 若被市場競爭、國家權力與軍事戰略無限制推動，最終可能成為人類難以控制的「怪獸」。

同樣的，新加坡方面也所見略同。由新加坡科技設計大學（SUTD）所發表的「AI 倫理白皮書」強調，亞洲不能只是被動接受歐美既有 AI 治理模式，而應提出自己的文明觀點與倫理資源，尤其可運用中華文化中的儒家思想，回應 AI 時代的人性危機，更具體建議成立「亞洲 AI 與倫理研究所」（Asian Institute of AI and Ethics, AIAIE），結合產官學界與社會力量，推動應用研究、政策協調與跨國對話。這代表 AI 倫理不能停留在零散討論，而必須制度化、平台化與區域化。

在此脈絡下，「亞太綜合研究院」可以扮演台灣方面的重要角色。AI 雖然是科技革命，但它引發的主要後果卻影響國家治理、兩岸關係、認知戰與文明價值重組等問題，亞太院也正與 SUTD 洽談合作，正將 AI 與政治

學、經濟學及國際關係研究對接，再引入中華文化思想資源，以期形成具有東方特色與世界意義的 AI 倫理平台。

因此可知，教宗代表文明良知，Anthropic 與 Claude 代表技術前沿的安全警覺，Palantir 代表 AI 進入國家安全與戰場決策的現實，而新加坡與亞太院正進行中華文化與 AI 融合的對接。尤其台灣不能只把 AI 理解為半導體與產業升級，也應主動參與防止 AI 失控的國際倫理工程，為避免 AI 成為怪獸、降低戰爭門檻與維護人類主體性，做出具有亞洲特色與文明意義的貢獻。AI 的發展當然重要，但更重要的是，人在 AI 時代還能不能保持人的特質與尊嚴。

掃描下載
Word檔

AI 時代的文明對話：中華文化可為教宗提供倫理答案

原刊於：愛傳媒｜2026 年 7 月 14 日

AI 正以前所未有的速度改變人類社會。從生產、教育、醫療、媒體到軍事等層層面面，AI 不只提高效率，也逐漸介入人的判斷、選擇與社會關係。2025 年 5 月新任教宗良十四世上任後就多次強調，今年五月更是發出通諭警示，AI 必須服務人類，不能取代人、操控人，更不能支配人。這項關切不只是西方世界的問題，也是全人類共同面對的文明挑戰。

在這樣的背景下，中華文化其實可以為教宗及全球 AI 倫理提供重要參考。中華文化的價值，不在於給出一套可以直接套用的答案，而在於它數千年來對人、社會、自然、權力與秩序的思考，能夠補充當前 AI 倫理過度偏重技術規範與個人權利的不足。

儒家仁義思維　補足 AI 倫理中的人本缺口

首先，儒家的「仁」，可以為 AI 發展提供真正以人為本的標準。今日許多科技公司強調速度、效率與成本，但儒家會追問：這項技術是否使人與人之間更有關懷？例如，照護機器人可以協助老人生活，卻不能取代親人的陪伴；AI 醫療可以提升診斷效率，卻不能把病人只看成數據和成本；AI 教育可以協助教師，卻不能使師生關係變成單純的知識傳輸。

因此，判斷 AI 是否進步，不能只看它做得多快，也要看它是否使人更有尊嚴、更有關係、更有責任。這種「關係性的人本主義」，正可以與天主教所強調的人的尊嚴相互呼應。

其次，孔子的「己所不欲，勿施於人」，可以發展為一項非常具體的演算法倫理。今天 AI 已被用來進行信用評分、招聘篩選、保險計算、醫療分

配，甚至司法與治安預測。設計者若只從管理角度出發，很容易忽略被評估者的權利。

如果把孔子的恕道放進 AI 設計，就應該問：設計者願不願意自己也接受同樣不透明的評分？科技公司願不願意自己的家人受到同樣程度的監控？政府是否願意讓自己也被無法申訴的演算法決定？這可稱為「演算法恕道」，也就是不能把自己不願承受的風險，轉嫁給弱勢者。

第三，儒家的「義利之辨」，可以直接回應科技企業的利潤邏輯。AI 企業當然可以追求收益，但「能夠獲利」不代表「符合正義」。若 AI 被用來剝削勞工、操縱注意力、侵犯隱私、壟斷市場或集中財富，就算技術再先進，也不能稱為真正的進步。

中華文化所提供的提醒是：技術上可行不代表道德上應當。利益必須受到正義、責任與共同利益的限制。這與教宗對科技壟斷、資料剝削及勞工失業的關切，具有相當大的交集。

道家知止哲學　提醒人類科技發展必須有界線

第四，道家的「無為」與「知止」，可以為科技發展提供節制。老子的無為並不是消極不做為，而是有所為有所不為。運用到 AI 領域，就是：不是所有能夠收集的資料都應該收集，不是所有能夠自動化的工作都必須自動化，不是所有能夠監控的空間都應受到監控，也不是所有能夠製造的武器都應被製造。

老子所說的「知止不殆」，對今日 AI 尤其重要。人類現在最缺乏的，可能不是創新的能力，而是停止的能力。科技公司不斷追求更大的模型、更快的運算與更深入的監控，但社會也必須有勇氣在某些地方說「到此為止」。

第五，「天人合一」與「道法自然」，可以補充 AI 倫理中的生態面向。人工智慧看似存在於雲端，實際上卻需要大量電力、冷卻水、晶片、礦產與數據中心。AI 發展若只追求速度，可能加重能源消耗與環境壓力。

中華文化提醒人類，人不是自然的征服者，而是天地秩序的一部分。這與天主教所強調的「整體生態」可以形成深刻對話。未來的 AI 倫理，不只要處理人與機器的關係，也要處理科技與自然、當代人與未來世代的關係。

和而不同　打造跨文明 AI 治理新模式

第六，「和而不同」可以成為全球 AI 治理的重要原則。當前 AI 規則主要由少數科技大國與大型企業主導，但不同文明對自由、責任、共同體與人的尊嚴，仍有不同理解。如果全球 AI 規範只由單一文明制定，容易形成價值霸權；但若各國各行其是，又可能造成軍備競賽與治理失控。

「和而不同」提供了一條中間道路：各文明可以保留自身特色，但仍可在人的尊嚴、安全、公平、和平與責任上建立最低共識。教廷具有跨國宗教與外交網絡，正可以推動基督教、儒家、道家、佛教、伊斯蘭教與世俗倫理之間的對話，使 AI 治理真正成為文明互鑑，而不是單向輸出。

當然，中華文化也不能被過度理想化。傳統文化中仍有重等級、輕平等，重和諧、輕異議，以及重道德教化、輕制度問責等問題。因此，中華文化必須經過現代轉化，與人權、法治、民主參與、透明及問責制度相結合。

最理想的方向是：以儒家的仁與責任確定 AI 發展的目標，以道家的知止限制技術權力，以現代法治建立邊界，再以透明制度落實監督。

人工智慧之善，不在於機器是否勝過人，而在於它是否使人更仁愛、社會更公義、國家更和平、天地更加和諧。這不只是中華文化可以提供給教宗的參考，也可能是東方文明面對 AI 時代所能提出的一項重要貢獻。

歐美 AI 領域對於中華文化的注視

掃描下載
Word檔

原刊於：聯合新聞網｜2026 年 9 月 16 日

當前 AI 發展迅猛，對全球、國家、社會與個人等各層面涉入日深，可謂無遠弗屆、無所不在。但有一個發展趨勢值得注意，即在 AI 倫理、AI 治理與認知科學等領域，對中國哲學和佛教思想的興趣正在明顯增加。尤其是當西方傳統的個人主義、功利主義與競爭論等不足以處理 AI 的關係性、整體性與長期風險時，東方思想開始被當作一種補充與緩和的資源。

這裡有幾個很值得注意的方向。第一是《易經》與 AI 的關係，《易經》強調的不是靜止的實體，而是「變、關係、情境」。這恰好與現代 AI 愈來愈重視的動態系統、複雜系統、概率判斷有某種結構上的呼應。2021 年 Springer 出版的《Intelligence and Wisdom: Artificial Intelligence Meets Chinese Philosophers》就已經開了風氣之先，該書專門收入了《易經》宇宙論與 AI 的討論，同時也討論儒家、道家和佛教如何重新思考人與 AI 的關係。

第二是道家特別受到關注。因為《道德經》講「無為」「自然」「知止」「不爭」，與當前 AI 追求無限擴張、無限優化、無限控制形成很有意思的反差。2026 年《Journal of Daoist Studies》甚至專門討論 AI 與道家哲學，提出「無為」「自然」「和諧」「節制」可以成為 AI 倫理的一種替代性框架。這其實非常重要，因為至今的 AI 思維往往強調如何強化 AI 的能力，而道家的內涵卻是要將 AI 從「能力」轉向「節制」。這對於未來 AGI 的治理非常有價值。

第三是佛教。佛教近年來在 AI 倫理討論中受到相當關注，尤其是「緣起」「無我」「慈悲」「正念」這些概念。2026 年已有研究直接討論如何把佛教的心性論作為 AI 倫理架構的一部分；另有研究將佛教與基督教、伊

斯蘭教一起用於比較 AI 治理模式，尤其「緣起性空」與 AI 的關係其實非常接近。

例如 AI 本身幾乎沒有任何東西能夠「獨立存在」：AI 能力 = 資料 × 演算法 × 算力 × 人類回饋 × 社會環境。少了其中任何一個條件，它都不是我們今天看到的 AI，比如佛家的「此有故彼有，此生故彼生」，確實可以提供一種理解 AI 的關係論視角。

第四，儒家可能對未來 AI 倫理更加實用。西方 AI 倫理經常從「個人權利」出發，例如隱私、自主、自由、公平；儒家則更強調「關係」—父子、師生、官民和社會等。因此有人開始研究仁、禮、信、恕等概念如何用於 AI 教育與 AI 治理。2026 年的研究就已經把儒家「修身」與大型語言模型（LLM）的教育倫理聯繫起來。

綜上所述，若逐漸增加上述東方元素，來導引人、AI、社會與自然，形成一種和諧共存的關係，可以日益優化當前歐美如何強化與控制 AI 的模式。如果把筆者正在推動的「臺北寰宇 AI 倫理學會」的核心宗旨「儒釋道與 AI 倫理」進一步理論化，甚至可形成一個四層框架:

易經—變：AI 如何理解動態世界。儒家—仁：AI 應該服務什麼樣的人際社會。道家—止：AI 什麼時候應該克制、不作為。佛家—空：AI 如何避免把模型、概念和自我視為固定實體。

這其實已經可以形成一套相當有特色的「中華文化 AI 倫理模型」，這並不是憑空提出的概念，而是正可引導當前國際 AI 學術研究並與之對接的利基。

第 三 章

東方智慧

傳統思想與科技實踐

AI 與中國傳統智慧：杭州的體驗

讓 AI 科技成為文明傳承的夥伴

緣起性空與 AI 模式

AI 與中國傳統智慧：杭州的體驗

掃描下載
Word檔

原刊於：愛傳媒｜2026 年 8 月 17 日

AI 正以前所未有的速度改變世界。從醫療診斷、金融交易，到司法判斷與軍事決策，AI 已逐漸進入人類社會最關鍵的領域。然而，科技愈進步，問題也愈明顯：AI 如何做出判斷？其推理過程是否透明？為何會產生「幻覺」？又是否必須不斷依賴龐大算力與資料，才能持續提升能力？這些問題，正構成今日全球 AI 發展的重要瓶頸。

最近在杭州西溪濕地附近接觸到一項頗具啟發性的嘗試，與本人所推動的「台北寰宇 AI 倫理學會」，有異曲同工之妙。「浙江省現代炁（氣的古字）脈健康研究院」，將中國古老生命觀中的「炁脈」理念，與現代 AI 架構結合，研發出醫學 AI 系統 Prome。研究團隊主張，這套系統試圖回應目前大型語言模型（LLM）所面臨的黑箱不可解釋、幻覺生成、算力高度依賴及災難性遺忘等問題。

該研究院由古中醫非遺「炁脈調理」傳承人劉清源於 2018 年創立，長期嘗試以現代科學語言重新理解「炁」與「炁脈」。其中值得注意的，不只是中醫本身，而是其背後所反映的一種認知方法：分層觀察、動態平衡、整體聯繫與溯源推理。人體並不是單一訊號的機械集合，而是一個多層次、持續變化、彼此關聯的生命系統。這種思維，也成為 Prome 設計的重要靈感。

目前全球 AI 競爭很大程度仍集中於「堆參數、拼算力」。模型愈來愈大，所需資料、晶片、能源與資金也愈來愈多，但投入與效益未必始終成正比。Prome 則嘗試走另一條路，以物體表徵、任務表徵、自我狀態與執行輸出等不同層次，建立較清楚的認知架構，使各層輸入、處理與輸出能被追蹤與檢視。其理念與中醫辨證某些特徵頗為相似：不是只看最後結果，而是追問問題從何而來、如何形成，以及在哪一個層次發生偏差。

這使我想到，中國傳統文化與非物質文化遺產所保存的，從來不只是技藝，而是一套關於人、自然與社會如何相互作用的智慧。中醫強調陰陽平衡與整體觀，太極講求剛柔相濟，儒家重視「中庸」與「和而不同」，道家則強調順勢而行與天人關係。它們共同反映的，不是把單一力量推向極致，而是在複雜關係中尋求動態平衡。

這與今日 AI「愈快、愈大、愈強」的發展邏輯形成有趣對照。科技企業不斷追求效率最大化，但人類社會並不是一道單純的數學方程式。效率最高的決定，不一定最公平；最精準的演算法，也未必最符合人的尊嚴。若 AI 缺乏價值約束，能力愈強，潛在風險反而可能愈大。

「炁」是「氣」的古字，它所象徵的並不是一種孤立存在，而是一種流動、關聯與變化的生命觀。若轉化成現代語言，就是任何事物都存在於關係之中。AI 同樣如此：一套演算法不能只看模型本身，還必須考慮資料來源、使用者、制度環境、文化背景以及可能受到影響的人。

因此，破解 AI 黑箱，不能只依靠另一套更複雜的技術，更重要的是重新把「人」放回科技系統的中心。真正可信任的 AI，不只要能回答「它如何得到答案」，也就是「實然」問題；更應該回答「為什麼要做這件事」以及「這個結果是否有益於人」，也就是「應然」問題。

杭州的這次體驗給我最大的啟發，是東方傳統智慧並非只能存在於博物館或歷史典籍之中。它完全可能成為 AI 時代重新理解科技的一種思想資源。當千年智慧與最先進科技相遇，真正重要的不是證明古代比現代高明，而是讓不同文明的思想彼此啟發。

AI 的真正新生，也許不是讓機器愈來愈像人，而是讓科技重新學會尊重人。未來全球 AI 競爭真正值得關注的，也不只是誰擁有最大的模型與最強的算力，而是誰能建立一套兼顧效率、倫理、責任與人類尊嚴的文明規則。從這個角度來看，「和合」、「中道」、「天人合一」以及整體思維，或許正能為全球 AI 發展提供一條值得重新思考的東方道路。

讓 AI 科技成為文明傳承的夥伴

掃描下載
Word檔

原刊於：經濟日報｜2026 年 8 月 21 日

八月的杭州，首先感受到一種十分鮮明的對照：一方面，這是一座承載千年歷史與江南文化的城市；另一方面，它又站在中國數位經濟、AI 與科技創新的前沿，古典與現代在這裡相互融合。

這也啟發吾人的思考：如何呼應教宗今年五月所發出的警示性通諭，盡量不讓 AI 走上怪獸之路而反噬人類，若能從根（兒童）做起，自然釜底抽薪，意義重大，而杭州好似已經給出了具體答案。

例如，杭州的氣脈研究院正在研發一種嶄新的「書童」AI 機器人。中國傳統的書童，不只是替主人拿書、磨墨，更是陪伴學習、生活與成長的角色。到了 AI 時代，我們或許可以重新賦予「書童」新的生命，讓 AI 不只是回答問題的機器，而成為陪伴下一代認識自己、理解世界、承接文明的助手，也使人聯想到未來 AI 將朝更加細緻分工與專業陪伴的方向發展。

這樣的書童，首先應是一位文明傳承的使者。今天我們談文明，不能只局限於單一文化。古希臘留下理性與邏輯，印度文明重視內在覺知，伊斯蘭文明在數學、醫學與天文學上作出重要貢獻，非洲文化保存共同體精神，美洲原住民提醒人類敬畏土地，而西方文藝復興則重新確立人的尊嚴。這些都是全人類共同的智慧火種。

但是，一個孩子要真正走向世界，首先還是要知道自己從哪裡來，此 AI 書童具有全球視野，又以中華文明為根，中西合璧，相得益彰。中華文明延續數千年而未中斷，乃世界唯一，不僅關心外在世界如何運作，也持續追問人的內在生命如何安頓。

本文把這兩個層面稱為「顯學」與「隱學」。前者讓人在人間立得住，後者讓人在精神上看得遠。只有外在知識而沒有內在安頓，人容易成為工具

理性的奴隸；只有內在修養而缺乏現實能力，也難以面對快速變動的世界。因此，未來教育不能只是增加知識，而應培養完整的人格。

AI 可以知道很多知識，卻未必知道什麼值得追求；它可以快速計算，卻不能替人決定人生意義。AI 終究是「術」，而不是「道」。如果孩子只是不斷向 AI 索取答案，久而久之，可能削弱自身思考、感受與判斷能力；但若把 AI 設計成真正的「書童」，情況就完全不同。

杭州氣脈研究院所設想中的書童 AI，是提供〇到十八歲不同階段的成長需求，以完整生命觀陪伴孩子；它可以理解現代醫學、中醫、功能醫學及傳統生命觀的不同知識體系；它可以介紹歷史、哲學、科學與文化，卻不能強行灌輸知識；它更不能取代父母，因為愛、責任與倫理，仍必須在人與人的真實關係中建立。

換句話說，此書童不是來代替家長，而是培養兒童的完整人格。AI 下一階段的競爭，不只是模型多大、算力多強、回答多快，而是能否把 AI 分工做得更細、更專業、更貼近人的不同需求，以及更符合人性，從通用型 AI 走向教育陪伴、健康照護、情緒互動與文化傳承，正可能成為重要方向。

因此，我們應該把「書童」看成一個文明實驗：陪伴孩子卻不占有孩子；給孩子知識，卻保留思考自由；連接世界，卻不失去文化根源。杭州氣脈研究院所研發的「書童」，也讓我們看到 AI 正從「一個模型什麼都做」走向「不同機器各司其職」的細緻分工時代，未來的「書童」，可說是文明繼續前行的一座極佳橋梁。

緣起性空與 AI 模式

掃描下載
Word檔

原刊於：愛傳媒｜2026 年 9 月 6 日

AI 的快速發展，不僅改變了人類的生活方式，也對接了深層的哲學問題。尤其值得注意的是，佛教「緣起性空」的思想，與現代 AI 的運作模式之間，存在若干值得比較的相似性。雖然兩者分屬不同領域：緣起性空是佛教對世界、生命與存在本質的哲學理解，AI 則是建立在數學、統計、資料與演算法上的現代科技。然而，如果從「關係性」、「條件性」以及「非固定本質」的角度觀察，兩者確實可以形成一種具有啟發性的對話。

從「緣起」到 AI　資料與條件共同決定輸出

佛教所說的「緣起」，基本意思是任何事物都不是單獨存在，而是在各種因緣與條件共同作用之下產生，因而常以「此有故彼有，此生故彼生」來說明這種關係。換言之，世界上的事物並沒有一個完全獨立、自足的存在，而是在相互依存的網絡中形成。例如一棵樹的生長，需要種子、陽光、水分、土壤、氣候等多重條件；任何一項條件發生改變，都可能影響最後的結果。

這一點與 AI 的運作方式具有相當明顯的相似性。AI 是在機器內部預先儲存一個大型語言模型（LLM），其回答是由大量訓練資料、模型架構、參數權重、使用者提示詞以及前後文共同作用後產生。只要其中任何條件發生變化，AI 所給出的答案也可能改變。因此可以說，佛教講「因緣變，事物變」，AI 則呈現出「資料變，輸出變」的特徵。

「性空」則進一步說明，一切事物之所以由因緣構成，也就意味著它們不存在一個永恆不變、獨立自存的「自性」。這裡的「空」，並不是指什麼

都沒有，而是指事物沒有脫離關係與條件而獨立存在的固定本質。這種思想也可以用來理解現代 AI。AI 對語言、圖像或概念的理解，是建立在大量關係之中。例如英文「bank」一詞，可以指銀行，也可以指河岸，其真正意義要依照上下文才能確定。也就是說，意義不是孤立存在，而是在關係中被形成。

AI 沒有絕對答案　「緣起性空」提供 AI 倫理另一種思考

從這個角度看，AI 與緣起性空都呈現出一種「關係性」思維。傳統上，人們容易把事物看成具有固定本質的獨立個體；但緣起性空則強調，一切存在都必須放在關係網絡中理解。現代 AI，尤其是神經網絡與轉換器（Transformer）模型，也不是依靠單一因素作出判斷，而是透過大量節點、權重與資料關係產生結果。這種結構上的相似性，使佛教哲學與 AI 科技之間出現了一個有趣的交會點。

不過，兩者仍有根本差異。佛教的緣起性空不是單純的資訊處理方法，而是一套關於生命、存在、痛苦與解脫的哲學。它的最終目的，是幫助人超越執著，產生智慧與慈悲。AI 則是一種人工建構的計算系統，其主要功能是預測、分類、生成、分析與最佳化。AI 可以模擬語言，可以整理知識，也可以產生看似具有理解力的回答，但目前仍沒有充分證據證明 AI 具有佛教意義下的覺悟、慈悲或生命經驗。

因此，緣起性空反而可以為 AI 倫理提供重要的思想資源。如果 AI 的答案同樣是由資料、模型、演算法與人類設計共同形成，那麼 AI 的判斷就必然具有條件性，而不應被視為絕對真理。資料可能有偏見，模型可能有侷限，演算法也可能受到設計者價值取向的影響。因此，人類不能把決策完全交給 AI，而仍須保留最終的判斷與責任。這正與當代 AI 治理所強調的「Human-in-the-Loop」，也就是「人在迴路中」，具有深刻的呼應。

由此可以進一步說，緣起性空與 AI 的真正交會，不在於兩者完全相

同，而在於它們都提醒我們：世界與知識並不是由孤立、固定、不變的事物所構成，而是在不斷變動的關係中形成。

因此，如果將兩者濃縮成三個層次，可以說：「緣起」對應 AI 的資料與條件網絡；「性空」提醒我們 AI 沒有固定、絕對、自足的答案；而「般若智慧」則代表人類必須對 AI 的結果進行反思、判斷與價值選擇。這或許也是東方哲學在 AI 時代可以提供的重要貢獻：科技可以不斷提升算力與知識，但真正決定科技方向的，仍然必須是人類的智慧與倫理。

歸根究柢，緣起性空告訴我們知識來自關係；而 AI 則提醒我們，不要把關係所生成的知識誤認為永恆不變的真理。當 AI 愈來愈深入人類社會，這種思想不只是哲學上的比較，也可能成為未來 AI 倫理與治理的重要基礎。

成立大會新聞稿

掃描下載
Word檔

近百學者齊聚　台北寰宇 AI 倫理學會正式成立

2026 年 10 月 26 日，「台北寰宇 AI 倫理學會」（Taipei Global AI Ethics Society，GAIEST）成立大會於台北舉行，由理事長湯紹成發起，集結醫學、科技、法律、教育等領域近百位專家學者，並發表大會宣言，提出「天人共生、慈悲公義、全球共治」等主張。

湯紹成致詞指出，面對 AI，既要追問「能做什麼」，這是講究能力與效率的「智」；也要追問「該做什麼」，這是講究價值與責任的「慧」，兩者配合，相得益彰。隨著 AI 邁向代理型與物理型應用，他強調「AI 能力愈強，愈要監督」：涉及醫療、公共安全與生命的重大決策，必須堅持「人在決策迴路中」，讓人有時間查證、有權限暫停與否決，並落實全程問責。

學會以中華文化為倫理依歸，將儒家「仁愛」、佛家「慈悲」、道家「自然」融入 AI 倫理討論，並提出有別於西方「人權」的「仁權」新概念。湯紹成也援引教宗良十四世通諭《偉大的人性》，指出國際社會對 AI 時代人之尊嚴的關切與中華文化互相應和；學會並獲教廷駐華大使馬德範支持。

未來學會將聚焦三大方向：一是針對教育、照護、媒體與公共服務的 AI 應用進行案例研究與政策建議；二是推動 AI 素養與倫理教育，協助民眾查核 AI 答案、辨識合成影音，並照顧長者等弱勢群體；三是促進兩岸及國際學術交流，邀集專家研議 AI 時代的「信任建立機制」，降低誤判與衝突風險。

湯紹成表示，學會成立是共同承擔責任的開始，期盼從台北出發，藉「倫理」與「AI」的對接走向全世界，讓人在善用 AI 的同時，保有查證的機會、判斷的能力與對他人的關懷。

筆　記

筆　記



台北寰宇 AI 倫理學會

Taipei Global AI Ethics Society (GAIEST)