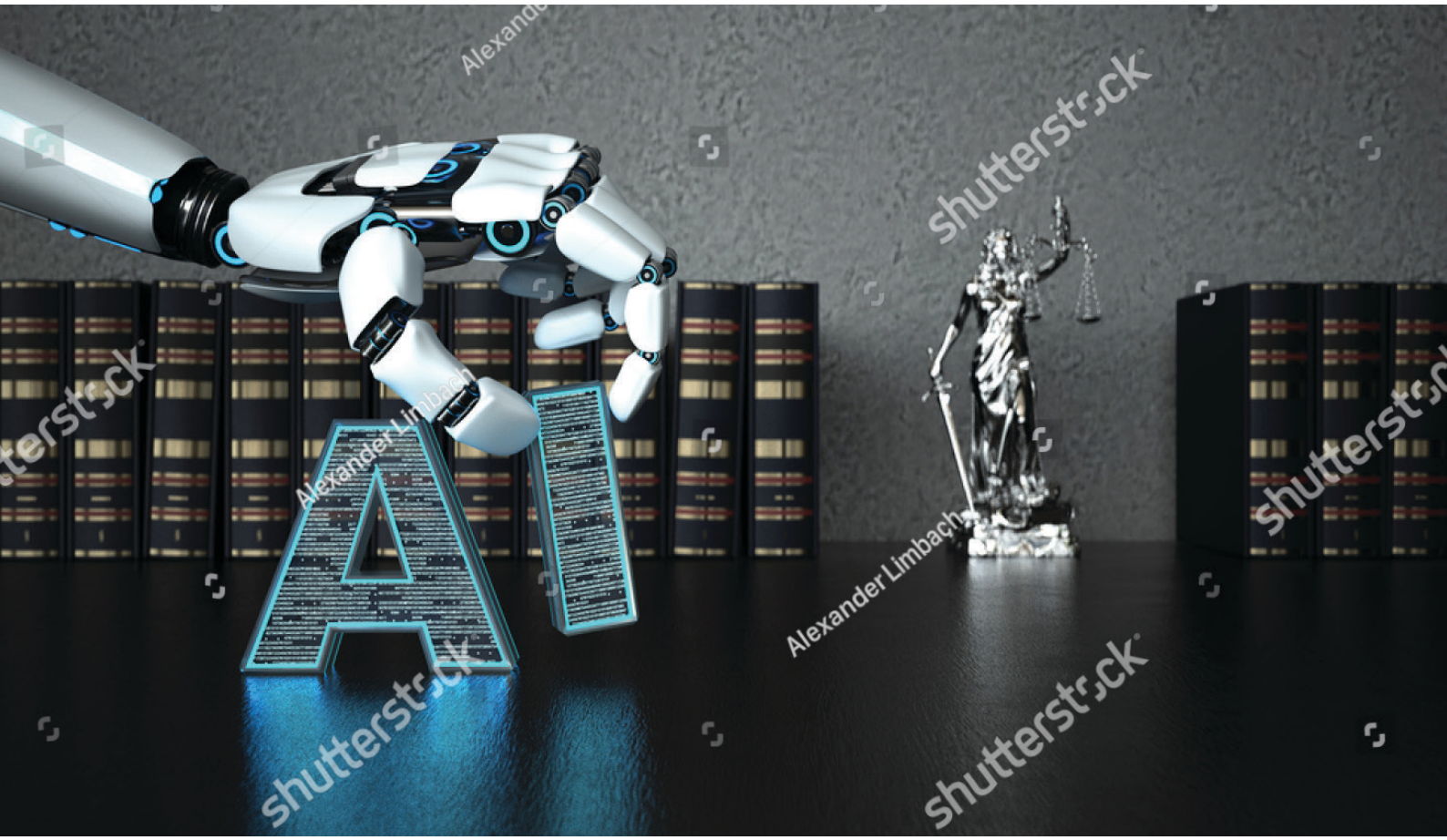
Building the Rule of Law in a Digital World

By Fred Kan, Senior Partner, Fred Kan & Co.



On 28 September 2018, at the conclusion of the 2nd Belt and Road Law Conference organized by The Law Society of Hong Kong, 34 law associations from 17 jurisdictions signed the following declaration on the formation of the Law Tech Alliance on promoting a shared ethical artificial intelligence (AI) framework between lawyers from the "Belt and Road" regions:

> *We, the undersigned organisations, share our support for innovation and technology development. We are also cognizant that there are legal and ethical challenges to the design, development and deployment of artificial intelligence around the world which may pose threats to the human race. There are many issues, such as whether A.I. should be used against human in warfare, in profiling and discrimination; whether communications and decision-making by and between machines should be fair and be examinable intelligently by human; effect of A.I. on rights and expectations to data protection and privacy; speed of replacement of traditional jobs and capacity building for new jobs, etc. They all need to be addressed. We therefore agree to collaborate to raise and critically examine relevant issues from a legal perspective, to heighten public awareness and to make submissions to our respective governments and the United Nations General Assembly.*

The 34 signatories of the declaration on the formation of the Law Tech Alliance included, in addition to The Law Society

of Hong Kong, 15 Chinese Mainland law associations, the Taiwan Bar Association, the Macau Lawyers Association, the Bar of Attorneys in Gdańsk, the Beirut Bar Association, the Czech Bar Association, the Federal Chamber of Lawyers of the Russian Federation, the French National Bar Council, the Indian National Bar Association, the Luxembourg Bar Association, the Milan Bar Association, the Mongolian Bar Association, the Myanmar Bar Association, the German Federal Bar, the Law Society of Brunei Darussalam, the Law Society of Singapore etc.

It is therefore most heartening to note that at the 41st session of the General Conference of the United Nations Educational, Scientific and Cultural Organisation (UNESCO) held in Paris on 24 November 2021, 193 Member States passed a resolution to adopt the "Recommendation on the Ethics of Artificial Intelligence" ("Recommendation"), the first ever global standard on the ethics of artificial intelligence adopted by the Member States of UNESCO at the General Conference. According to Audrey Azoulay, Director-General of UNESCO, this historical text defines the common values and principles which will guide the construction of the necessary legal infrastructure to ensure the healthy development of AI.

Perhaps it is interesting to note that China is a Member State of UNESCO, and the United States is not, as President Donald Trump took the United States out of UNESCO in 2017.

The preamble of the Recommendation canvasses the sweeping premise upon which the Recommendation has become pressingly important:

1. AI has profound and dynamic positive and negative impacts on societies, environment, ecosystems and human lives, including the human mind.

2. AI technologies can be of great service to humanity and all countries can benefit from them, but they also raise fundamental ethical concerns. AI technologies can deepen existing divides and inequalities in the world, within and between countries. Bringing benefits can also amplify tension around innovation, asymmetric access to knowledge and technologies. Justice, trust and fairness must be upheld so that no country and no one should be left behind.

3. A standard-setting instrument developed through a global approach, based on international law, focusing on human dignity and human rights, as well as gender equality, social and economic justice and development, physical and mental well-being, diversity, interconnectedness, inclusiveness, and environmental and ecosystem protection, can guide AI technologies in a responsible direction.

The Recommendation seeks to define AI systems as systems which have the capacity to process data and information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control.

The principal aim of the Recommendation is to provide a basis to make AI systems work for the good of humanity and to prevent harm. Its principal objective is to provide a universal framework of values, principles and actions to guide States in the formation of their legislation, policies or other instruments regarding AI, consistent with international law.

The values propounded hinge on the respect, protection and promotion of human rights and fundamental freedoms and human dignity. Environmental and ecosystem flourishing should be recognised, protected and promoted through the life cycle of AI systems as well as ensuring diversity and inclusiveness. Furthermore, AI actors should play a participative and enabling role to ensure peaceful and just societies, which is based on an interconnected future for the benefit of all.

The Recommendation lays down the following fundamental principles:

1. There is to be proportionality in that AI system life cycle should not exceed what is necessary to achieve legitimate aims and objectives. In scenarios where decisions are understood to have an impact that is irreversible or difficult to reverse or may involve life and death decisions, final human determination should apply. In particular, AI systems should not be used for social scoring or mass surveillance purposes.

2. Safety and security risks should be avoided. Safe and secure AI will be enabled by the development of sustainable, privacy-protective data access frameworks that foster better training and validation of AI models utilizing quality data.

3. AI actors should promote social justice and safeguard fairness and non-discrimination of any kind in compliance with international law.

4. The advent of AI technologies can either benefit sustainability objectives or hinder them. Therefore, the continuous assessment of AI technologies on sustainability should be carried out against a set of constantly evolving goals across a range of dimensions, such as those currently identified in the Sustainable Development Goals of the United Nations.

5. There should be a right to privacy and data protection monitored by judicial systems.

6. It must always be possible to attribute ethical and legal responsibility for any stage of the life cycle of AI systems, as well as in cases of remedy related to AI systems, to physical persons or to existing legal entities. Life and death decisions should not be ceded to AI systems. Human oversight must include also public oversight.

7. The transparency and explainability of AI systems are essential preconditions to ensure the respect, protection and promotion of human rights, fundamental freedoms and ethical principles. Transparency is necessary for relevant national and international liability regimes to work effectively. A lack of transparency could also undermine the possibility of effectively challenging decisions based on outcomes produced by AI systems and thereby infringes the right to a fair trial and effective remedy, and limits the areas in which these systems can be legally used.

8. Appropriate oversight, impact assessment, audit and due diligence mechanisms, including whistle-blowers' protection, should be developed to ensure accountability for AI systems and their impact throughout their life cycle. Both technical and institutional designs should ensure auditability and traceability of the working of AI systems, particularly to address any conflicts with human rights norms and standards and threats to environmental and ecosystem well-being.

9. Public awareness and understanding of AI technologies and the value of data should be promoted through open and accessible education, civic engagement, digital skills and AI ethics training, media and information literacy and training.

10. Participation of different stakeholders, throughout the AI system life cycle, is necessary for inclusive approach to AI governance, enabling the benefits to be shared by all and to contribute to sustainable development.

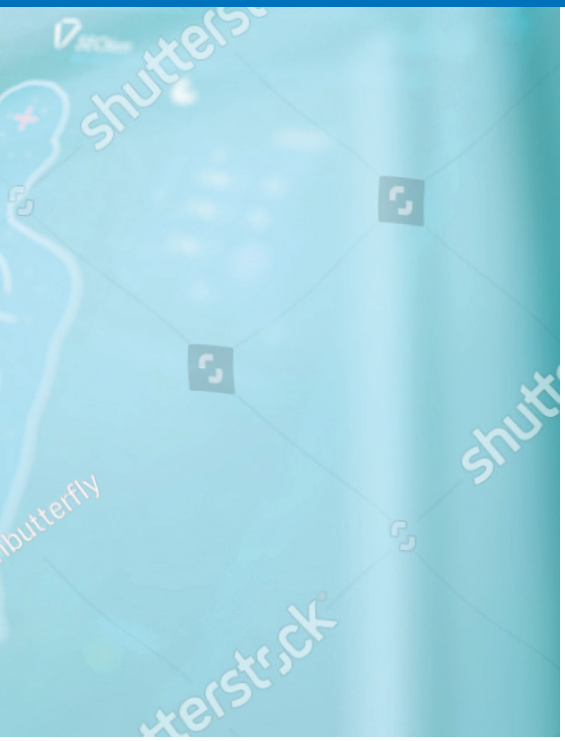
As to policy action enabling the effective implementation of the Recommendation, UNESCO will therefore: (1) develop a readiness assessment methodology to assist interested Member States in identifying their status at specific moments of their readiness trajectory along a continuum of dimensions; and (2) ensure support for interest Member States in terms of developing a UNESCO methodology for Ethical Impact Assessment (EIA) of AI technologies, sharing of best practices, assessment guidelines and other mechanisms and analytical work.

On the other hand, governments should adopt a regulatory framework that sets out a procedure, particularly for public authorities, to carry out ethical impact assessments on AI systems to predict consequences, mitigate risks, avoid harmful consequences, facilitate citizen participation and address social challenges.

Member States should ensure the institution of a governance and stewardship regime. Harms caused through AI systems are to be investigated and redressed. Certification mechanism for AI systems and the mutual recognition of their certification are to be developed. When developing regulatory frameworks, Member States should take into account that ultimate responsibility and accountability must always lie with natural or legal persons and that AI systems should not be given legal personality themselves.

Member States should work to develop data governance strategies that ensure the continual evaluation of the quality of training data for AI systems. Member States should ensure that individuals retain rights over their personal data.

Member States and transnational corporations should prioritise AI ethics



by including discussions on them in international, intergovernmental and multi-stakeholder fora.

Member States and business enterprises should assess the direct and indirect environmental impact throughout the AI system life cycle, including but not limited to, its carbon footprint, energy consumption and the environmental impact of raw material extraction for supporting the manufacturing of AI technologies, and reduce the environmental impact of AI systems and data infrastructures.

Member States should ensure that the potential for digital technologies and AI to contribute to achieving gender equality is maximized. Ethical Impact Assessment should include a transversal gender perspective.

Member States should foster new research at the intersection between AI and intellectual property (IP), for example to determine whether or how to protect IP rights the works created by means of AI technologies.

Member States should work with international organisations, educational institutions and private and non-governmental entities to provide adequate AI literacy education to the public on all levels in all countries in order to empower people and reduce the digital divides and digital access inequalities resulting from the wide adoption of AI systems.

Member States should use AI systems to improve access to information and knowledge.

Member States should assess and address the impact of AI systems on labour markets and its implications for education requirements in all countries and with special emphasis on countries where the economy is labour-intensive.

Member States should endeavour to employ effective AI systems for improving health and protecting the right to life, including mitigating disease outbreaks. Although it is not so stated in the Recommendation, it must imply improving access to healthcare and preventing disease outbreaks.

Member States should develop guidelines for human-robot interactions and their impact on human-human relationships. Particular attention should be given to the use of robots in health care and the care for older persons and persons with disabilities, in education, and robots for use by children, toy robots, chatbots and companion robots for children and adults. Special attention should be paid to the possibility of using AI to manipulate and abuse human cognitive biases. Member States should implement policies to raise awareness about the anthropomorphization of AI technologies and technologies that recognize and mimic human emotions, particularly in the context of robot-human interaction and especially when children are involved.

As for monitoring and evaluation, the Recommendation proposes the establishment of national ethics commissions to go together with the development of a UNESCO Ethical Impact Assessment methodology.

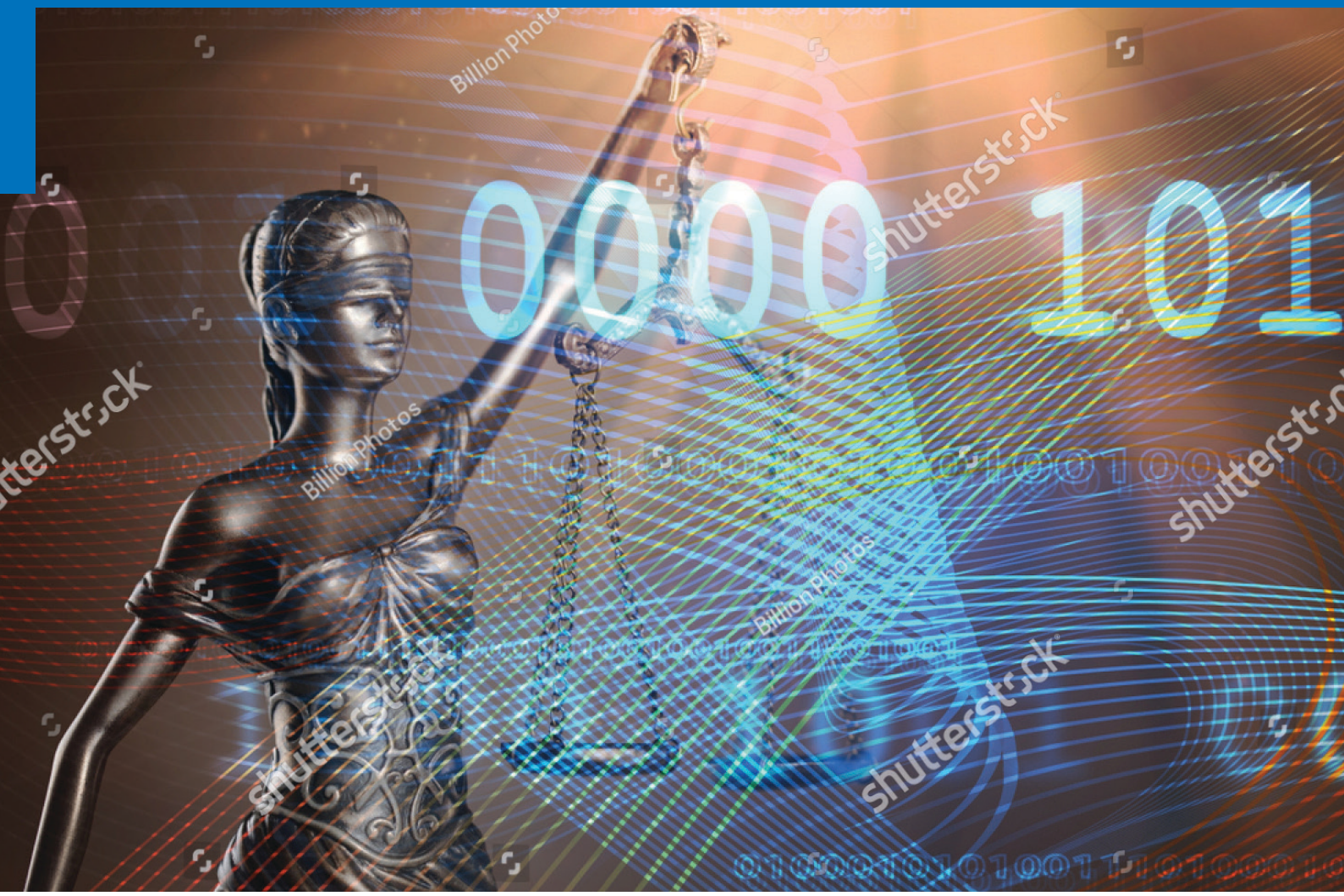
The Recommendation is only a set of recommendations. It will be up to the individual State to decide whether to act upon it, and if so, to what extent. But as remarked by Audrey Azoulay, Director-General of UNESCO, "[t]he world needs rules for artificial intelligence to benefit humanity. The Recommendation on the Ethics of AI is a major answer. It sets the first global normative framework while giving States the responsibility to apply it at their level. UNESCO will support its 193 Member States in its implementation and ask them to report regularly on their progress and practices."

The long, intensive and edifying process leading up to the birth of the Recommendation has indeed created a burgeoning interest and momentum for legislating based on an internationally accepted understanding of AI ethics. The publication of the Recommendation will no doubt greatly influence the negotiations afoot on the draft Artificial Intelligence Act presented by the European Commission on 21 April 2021. If passed by the European Parliament, it would be the world's first legally binding law on AI.

To legislate is the proper course to take. It must be so for the advancement of the rule of law, the core principle of which, according to Lord Bingham, is "that all persons and authorities within the state, whether public or private, should be bound by and entitled to the benefit of laws publicly made, taking effect (generally) in the future and publicly administered in the courts."

We are now witnessing the birth of the metaverse. We are living in a digital world.

With the great interest shown by the 34 law associations signing the declaration on the formation of the Law Tech Alliance and the engaging debate at the 2nd Belt and Road Law Conference on the ethical boundaries of AI, it is incumbent upon The Law Society of Hong Kong to follow through with the aims and objectives of the declaration by drawing together again these 34 law associations and other like-minded organisations to establish a platform to deliberate on an international model law on AI and for building the rule of law in a digital world in a UN-centric rules-based world order. ■



在數字世界建立法治

作者：簡家驄律師行高級合夥人 簡家驄

2018 年 9 月 28 日，在香港律師會舉辦的第二屆「一帶一路」論壇結束時，來自 17 個法域的 34 個法律協會簽署了「成立法律科技聯盟的宣言」，同意成立「法律科技聯盟」，推動「一帶一路」律師所適用的人工智能道德框架。該宣言的內容如下：

> *我們，即下方簽約協會，共同支持創新和科技發展。我們同時意識到，在全球各地設計、發展和應用人工智能，正存在著法律和道德方面的挑戰，並可能對人類構成威脅。眾多問題包括：應否在戰爭、罪犯特徵分析和識別中對人類使用人工智能；機器之間的通訊以及由它們所作出的決策可否做到公平及由人類智慧審查；人工智能對資料保障和私隱的權利及期望所帶來的影響；傳統工作被取代的速度以及新工作的能力建設等。這些問題都需要被處理。因此，我們同意攜手合作，從法律角度提出並以批判性思維審視相關問題，提高公眾的意識，以及向我們各地的政府和聯合國大會提出意見。*

除香港律師會外，成立法律科技聯盟宣言的 34 個簽署方還包括 15 個中國內地的法律協會、台灣律師聯合會、澳門律師公會、格但斯克律師協會、貝魯特律師協會、捷克律師協會、俄羅斯聯邦律師協會、法國國家律師協會、印度國家律師協會、盧森堡律師協會、米蘭律師協會、蒙古律師會、緬甸律師協會、德國聯邦律師公會、汶萊達魯薩蘭律師協會、新加坡律師公會等。

在 2021 年 11 月 24 日在巴黎舉行的「聯合國教育、科學及文化組織（教科文組織）大會」第四十一屆會議

上，193 個會員國通過決議，採納「人工智能倫理問題建議書」（”《建議書》”），這是會員國在教科文組織大會上通過的首個人工智能倫理全球標準。教科文組織總幹事 Audrey Azoulay 表示，這一歷史性文本確定了有關人工智能共同的價值觀和原則，用以指導建設確保人工智能的健康發展所必需的法律框架。這個消息實在令人鼓舞。

或許值得注意的是，中國是教科文組織的會員國，而美國並不是，因為特朗普總統於 2017 年宣佈美國退出教科文組織。

《建議書》的序言指出了《建議書》廣闊的前提，闡述了為何《建議書》變得極為重要：

1. 人工智能（AI）從正負兩方面對社會、環境、生態系統和人類生活包括人類思想具有深刻而動態的影響。

2. 人工智能技術既可以對人類大受裨益並惠及所有國家，但也會引發根本性的倫理關切。人工智能技術會加深世界各地國家內部和國家之間現有的鴻溝和不平等。人工智能技術可以帶來重大惠益，但也會加劇圍繞創新產生的矛盾衝突、知識和技術獲取不對稱。必須維護正義、信任和公平，不讓任何國家和任何人掉隊。

3. 以國際法為依據、採用全球方法制定且注重人的尊嚴和人權以及性別平等、社會和經濟正義與發展、身心健康、多樣性、互聯性、包容性、環境和生態系統保護的準則性文書，可以引導人工智能技術向著負責任的方向發展。

《建議書》把人工智能系統定義為有能力以類似於智能行為的方式處理類據和信息的系統，通常包括推理、學習、感知、預測、規劃或控制等方面。

《建議書》主要旨在提供基礎，讓人工智能系統可以造福人類，同時防止危害。《建議書》的目標是依照國際法，提供一個由價值觀、原則和行動構成的普遍框架，指導各國制定與人工智能有關的立法、政策或其他文書。

當中提出的價值觀包括尊重、保護和促進人權和基本自由以及人的尊嚴。應在人工智能系統的整個生命周期內確認、保護和促進環境和生態系統的蓬勃發展，以及確保多樣性和包容性。此外，人工智能行為者應為確保建設和平與公正的社會發揮參與和促進作用，這種社會的根基是惠及全民的相互關聯的未來。

《建議書》制定了以下基本原則：

1. 人工智能系統生命周期應具備相稱性，不得超出實現合法目的或目標所需的範圍。在所涉決定具有不可逆轉或難以逆轉的影響或涉及生死抉擇的情況下，應由人類作出最終決定。人工智能系統尤其不得用於社會評分或大規模監控目的。

2. 應避免安全和安保風險。開發可持續和保護隱私的數據獲取框架，促進利用優質數據更好地訓練和驗證人工智能模型，便可實現有安全和安保保障的人工智能。

3. 人工智能行為者應根據國際法，促進社會正義並保障一切形式的公平和非歧視。

4. 人工智能技術的出現可能有利於可持續性目標，但也可能阻礙這些目標的實現。因此，在就人工智能技術開展持續評估時，應考慮對一套涉及多方面的動態目標（例如目前在聯合國可持續發展目標中認定的目標）的可持續性的影響。

5. 隱私權和數據保護應受司法系統監察。

6. 始終有可能將人工智能系統生命周期的任何階段以及與人工智能系統有關的補救措施的倫理和法律責任歸屬於自然人或現有法人實體。生死攸關的決定不應讓給人工智能系統來作出。人類監督也指公共監督。

7. 人工智能系統的透明度和可解釋

性往往是確保人權、基本自由和倫理原則得到尊重、保護和促進的必要先決條件。透明度是相關國家和國際責任制度有效運作的必要因素。缺乏透明度還可能削弱對根據人工智能系統產生的結果所作決定提出有效質疑的可能性，進而可能侵犯獲得公平審判和有效補救的權利，並限制這些系統的合法使用領域。

8. 應建立適當的監督、影響評估、審計和盡職調查機制，包括保護舉報者，確保在人工智能系統的整個生命周期內對人工智能系統及其影響實施問責。技術和體制方面的設計都應確保人工智能系統可審計和可追溯，特別是要應對與人權規範和標準之間的衝突以及對環境和生態系統福祉的威脅。

9. 應通過開放且可獲取的教育、公民參與、數字技能和人工智能倫理問題培訓、媒體與信息素養及培訓，促進公眾對人工智能技術和數據價值的認識和理解。

10. 不同利益攸關方對人工智能系統整個生命周期的參與，是採取包容性辦法開展人工智能治理、使惠益能夠為所有人共享以及推動可持續發展的必要因素。

因此，教科文組織將采取如下政策行動，以促進有效落實《建議書》：(1) 制定準備狀態評估方法，協助有關會員國確定其準備進程各方面在特定時刻的所處狀態；及 (2) 確保支持有關會員國制定教科文組織人工智能技術倫理影響評估（EIA）方法，分享最佳做法、評估準則、其他機制和分析工作。

另一方面，各國政府採用監管框架，其中特別針對公共管理部門提出人工智能系統倫理影響評估程序，以預測後果、減少風險、避免有害後果、促進公民參與並應對社會挑戰。

會員國應建立治理和管理制度，確保調查並補救人工智能系統造成的損害，建立人工智能系統認證機制和此類認證的相互承認。會員國在制定監管框架時，應考慮到最終責任和問題必須總是落實到自然人或法人身上，人工智能系統本身不應被賦予法人資格。

會員國應努力制定數據治理戰略，確保持續評估人工智能系統訓練數據的質量。會員國應確保個人可以保留對其個人數據的權利。

會員國和跨國公司應優先考慮人工智能倫理，在相關國際、政府間和多利益攸關方論壇上討論與人工智能有關的倫理問題。

在人工智能系統的整個生命周期內，會員國和工商企業應評估對環境產生的直接和間接影響，包括但不限於其碳足跡、能源消耗以及為支持人工智能技術制造而開採原材料對環境造成的影響，並應減少人工智能系統和數據基礎設施造成的環境影響。

會員國應確保數字技術和人工智能充分發揮促進實現性別平等的潛能。其倫理影響評估應包含橫向性別平等視角。

會員國應促進在人工智能和知識產權的交匯點上開展新的研究，例如確定是否或如何對通過人工智能技術創作的作品給予知識產權保護。

會員國應與國際組織、教育機構、私營實體和非政府實體合作，在各個層面向所有國家的公眾提供充分的人工智能素養教育，以增強人們的權能，減少因廣泛採用人工智能系統而造成的數字鴻溝和數字獲取方面的不平等。

會員國應利用人工智能系統改善信息和知識的獲取。

會員國應評估並處理人工智能系統對所有國家勞動力市場的衝擊及其對教育要求的影響，同時特別關注經濟屬於勞動密集型的國家。

會員國應努力利用有效的人工智能系統來改善人類健康並保護生命權，包括減少疾病的爆發。儘管《建議書》並無明示，但這必意味著改善獲得衛生保健和預防疾病爆發。

會員國應制定關於人機互動及其對人際關係所產生影響的準則，並特別關注人類身心健康，尤其應關注應用於衛生保健以及老年人和殘障人士護理的機器人、應用於教育的機器人、兒童用機器人、現具機器人、聊天機器人以及兒童和成人陪伴機器人的使用問題。應特別注意到利用人工智能操控和濫用人類認知偏差的可能。會員國應實施政策，提高人們對於人工智能技術以及能夠識別和模仿人類情緒的技術擬人化的認識，特別是在人機互動的情況下和涉及到兒童時。

在監測和評估方面，《建議書》提議制定教科文組織倫理影響評估方法和建立國家人工智能倫理委員會。

《建議書》僅提出一系列建議。各個國家決定是否採取行動及採取何種程度的行動，但正如教科文組織總幹事 Audrey Azoulay 所說：「世界需要為人工智能制定規則以造福人類。《建議書》是一個重要的舉措，它制定了首個規範性全球框架，同時賦予各國在相應層面應用該框架的責任。教科文組織將支持其 193 個會員國落實《建議書》的內容，並要求它們定期報告相關進展和做法。」

《建議書》的編制過程漫長、密集而具啟發性，確實為基於國際公認的人工智能倫理理解的立法，創造了濃厚的興趣和動力。《建議書》的發佈無疑會對歐盟委員會於 2021 年 4 月 21 日提出的《人工智能法草案》的談判產生重大影響。如獲歐洲議會通過，該法案將是世界上首部具有法律約束力的人工智能法律。

立法是正確的做法，是推進法治所必須的，根據兵咸勳爵（Lord Bingham）的說法，法治的核心原則是「國家內所有人和權力機關，不論公營或私營，都應當受法律約束，並有權享有法律帶來的裨益。法律必須公開制定，（一般）只在將來生效，並在法院公開施行。」

我們正在見證元宇宙的誕生。我們生活在一個數字世界。

既然有 34 個法律協會表達了強烈關注，簽署成立法律科技聯盟的宣言，並在第二屆「一帶一路」論壇上就人工智能的倫理界限展開激烈辯論，香港律師會將義不容辭，貫徹宣言的宗旨和目標，再次召集這 34 個法律協會和其他志同道合的組織，建立平台討論人工智能的國際示範法和在以聯合國為中心、以規則為基礎的世界秩序的數字世界中建立法治。■