AI法制的現在與未來

劉靜怡

國立台灣大學國家發展研究所特聘教授兼所長
中央研究院法律學研究所合聘研究員
中央研究院資創中心合聘研究員

May 15, 2025@成功大學法律學系

全球人工智慧合作夥伴聯盟（下的對抗態勢？）

- 從全球人工智慧合作夥伴聯盟（Global Partnership on Artificial Intelligence, GPAI）會員國的2022年底東京高峰會部長宣言談起 (https://www.gpai.ai/events/tokyo-2022/ministerial-declaration)
- Reaffirm our commitment to the OECD AI Principles, which are based on human-centred values, protecting dignity and well-being and promoting trustworthy, responsible and sustainable use of artificial intelligence.
- Affirm our commitment to protecting and promoting human-centred values and democracy that underpin an inclusive, development-oriented, sustainable and peaceful society.
- Oppose unlawful and irresponsible use of artificial intelligence and other technologies, which is not in line with our shared values.

OECD AI Principles

Values-based principles

Inclusive growth, sustainable development and well-being

Human rights and democratic values, including fairness and privacy

Transparency and explainability

Robustness, security and safety

Accountability

Recommendations for policy makers

Investing in AI research and development

Fostering an inclusive AI-enabling ecosystem

Shaping an enabling interoperable governance and policy environment for AI

Building human capacity and preparing for labour market transition

International co-operation for trustworthy AI

src: oecd.ai/en/ai-principles

GPAI現狀與未來

- **就AI相關事務成立國際平台的必要性**
  2018-19年之間，法國與加拿大兩國政府決定共同推動成立國際平台討論AI相關議題，其主張各國政府在鼓勵AI科技的研發與應用之際，應該就諸如資料隱私的保護和演算法的公開透明度等AI必然涉及的價值與道德議題，進行兼具專業與民主考量的分析與討論，避免上述價值的詮釋與發展方向產生偏差，對人類社會造成負面影響，因此倡議成立一個具有獨立特質的國際性專家平台，持續促進AI所涉的價值討論，藉此不但催促公權力在AI研發過程中應負起道德責任，也可避免科技遭到誤用，並強化公民對AI的信任

- **參與方式**
  有「會員國」、「觀察員」或「專家」等三種參與方式，惟其地位與權利義務存在差異（詳見GPAI Terms of References可下載全文）

GPAI現狀與未來

- **倡議過程**
  2018年加拿大主辦的G7高峰會上，兩國首度共同發表「AI聯合宣言」，呼籲成立常設性國際平台。接著，兩國又在2019年8月法國主辦的G7高峰會期間，正式提出成立GPAI的倡議，主張各國公私部門與學界等利害關係人，可以透過此一國際平台共同形塑AI的全球化策略

- **正式成立時間與依據**
  2020年6月G7科技部長會議，在美國主導下與各國提出共同聲明──*Joint Statement from Founding Members of the Global Partnership on Artificial Intelligence*

GPAI現狀與未來: GPAI 2.0

Dear GPAI Experts,

I hope my email finds you well.

On behalf of the OECD Secretariat, I would like to share with you the exciting news that the GPAI and the OECD's work programme on artificial intelligence are joining forces to work together to advance coordinated efforts to promote trustworthy AI.

This new model for GPAI – as an integrated partnership with the OECD under the GPAI brand – will maintain your expertise and contribution at its core. It will bring together the GPAI Multistakeholder Expert Group (MEG) and the OECD's Network of Experts on AI (ONE AI) as a broad expert community to collectively support, shape and inform an ambitious work programme on AI decided jointly by all GPAI and OECD member countries on equal footing. The Partnership will remain open to new countries, on the basis of commitment to the OECD Recommendation on AI, and to additional experts.

In particular, expert engagement will include both continuation of the work within the Working Groups dedicated to specific priorities and projects, co-ordinated by the OECD Secretariat, as well as the active participation of the Working Group Co-Chairs and the Expert Support Centres in the meetings of the governing bodies of the new model (the Council, Plenary, and Steering Group).

GPAI現狀與未來: GPAI 2.0

As we are transitioning to this new model, GPAI relies on your support and involvement. In this regard, we will hold dedicated meetings in September between each of the Working Groups and the OECD Secretariat team supporting the integrated partnership to exchange about ongoing projects, the preparations for the next programme of work, and the next steps to implement the new model. We will be in touch over the coming weeks to provide dates and logistical details for these meetings.

I and the team on CC remain at your disposal for any questions in the meantime, including with our new dedicated email address (gpai@oecd.org).

Kind regards,

Audrey PLONK

Deputy Director and Acting Head of Digital Economy Policy Division

Directorate for Science, Technology and Innovation

Tel: +33 (0)1 45 24 96 93 | Email: audrey.plonk@oecd.org

2, rue André-Pascal | 75775 Paris CEDEX 16 | France | www.oecd.org

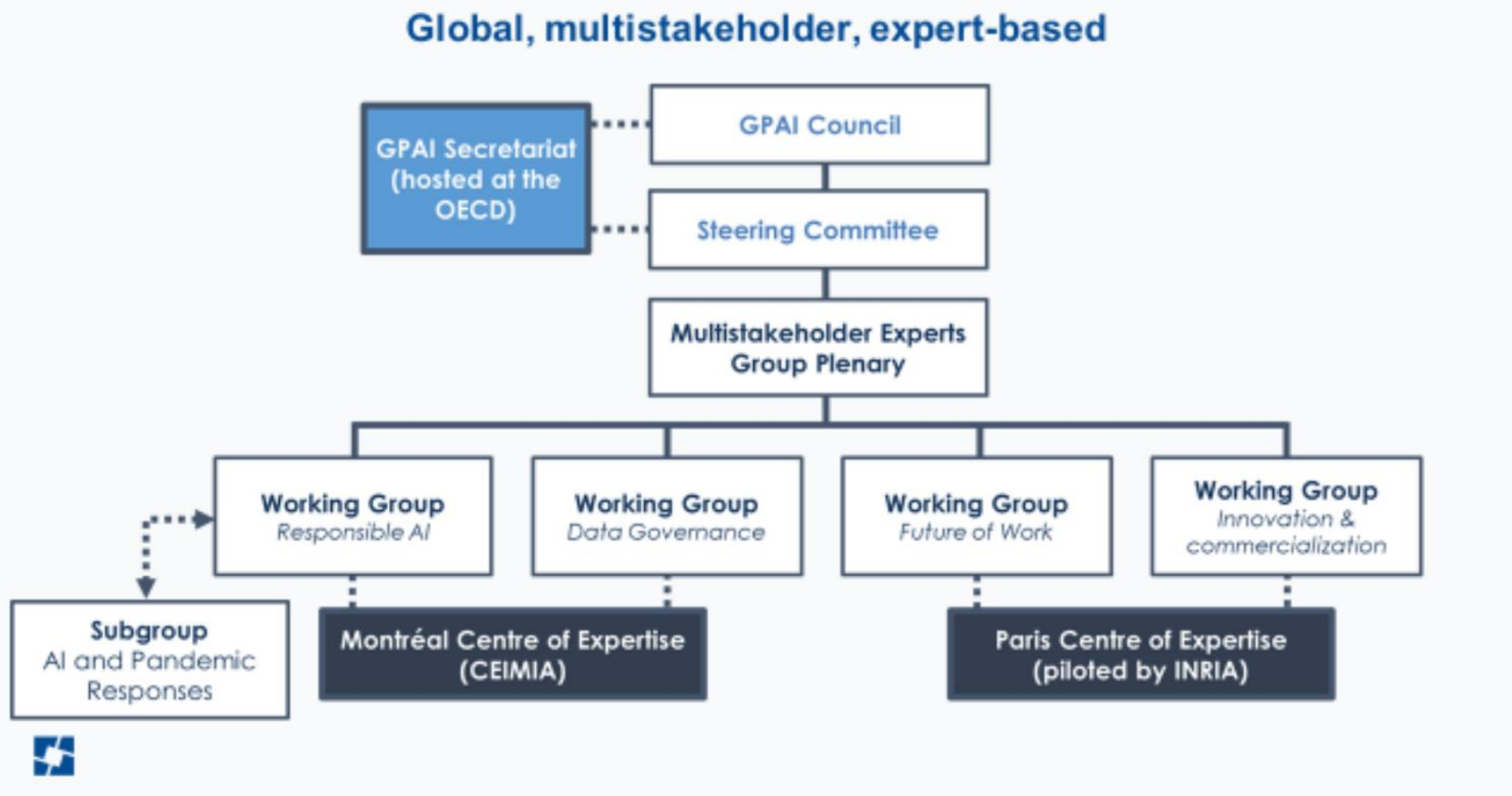
MEMBERS

44

Member Governments

GPAI Overview

GPAI STRUCTURE



 Src: gpai.ai/about/gpai-faq.pdf

SHARED PRINCIPLES

Section 1: Principles for responsible stewardship of trustworthy AI


Inclusive growth,
sustainable development and well-being


Human-centred values and fairness


Transparency and explainability


Robustness, security and safety


Accountability

Section 2: National policies and international co-operation for trustworthy AI


Investing in AI research and development


Fostering a digital ecosystem for AI


Shaping an enabling policy environment for AI


Building human capacity and preparing for labour market transformation


International co-operation for trustworthy AI

Based on the shared principles of the OECD Recommendation on AI

GPAI PRIORITIES

Four key priorities agreed for GPAI Work Plan 2024

GPAI | THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

Climate Change

Provide practical tools and methods to support the fight against climate change.

Global Health

Provide expertise on how to prepare the world for future pandemics.

Resilient Society

Foster and contribute to the responsible use of AI to mitigate the impact of world crises and protect vulnerable populations.

Human Rights

Study the impact of AI on human rights including gender equality and to consider AI's potential for social welfare.

GPAI Overview

ORGANIZATIONAL STRUCTURE

- 多方利害關係人專家小組大會（Multi-stakeholder Experts Group Plenary, MEG）：接受會員國推薦或自薦，共100-150位來自科學、產業、公民社會、工會組織、國際組織（如OECD、UNESCO）與各國政府等各界代表組成，按工作小組的工作成果提出建議。主席為指導委員會的當然成員。
- 工作小組（Working Group，WG）為工作核心

Co-Chairs of Data Governance Working Group

Dr Jeni Tennison

Founder and Executive Director

Connected by Data

UK

Dr Maja Bogataj Jančič

Founder and Head

Intellectual Property Institute

Slovenia

GPAI Overview

Member Nominated Experts

Christiane Wendehorst
EU

Kim McGrail
Canada

Paul Dalby
Australia

Ulises Cortés
Mexico

Yeong Zee Kin
Singapore

Mikael Jensen
Denmark

Alžběta Krausová
Czech Republic

Radim Polčák
Czech Republic

Robert Kroplewski
Poland

Ricardo Baeza-Yates
Spain

GPAI Overview

Self Nominated Experts

Andrea A. Jacobs
Antigua and Barbuda

Bertrand Monthubert
French

Kudakwashe Dandajena
Republic of Zimbabwe

Zümrüt Müftüoğlu
Turkey

Sarah Shoker
United States

Shameek Kundu
Singapore

Ching-Yi Liu
Taiwan

Jhalak Mrignayani Kakkar
India

Marc Rotenberg
United States

GPAI Overview

Observers

Christian Reimsbach-Kounatze
OECD

Alan Paic
OECD

Jaco Du Toit
UNESCO

GPAI | THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

The role of Government as a provider of Data for AI

Ching-Yi Liu
Taiwan

Co-Lead GPAI'S Government & Data

GPAI | THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

Project objectives

This project aims to support governments to make decisions about whether and how to share data they steward with AI developers.

The intended impact is to increase the availability of publicly held data for AI grounded in the principles of human rights, inclusion, diversity, innovation and economic growth by helping governments to prioritize their efforts and to reduce their concerns about the risks of sharing public data for AI by providing clear guidance, use cases and examples that demonstrate how it can be done safely and responsibly.

Collaborative Teams:
Research ICT Africa & D4D.net
Oxford Insights, UK
Taiwan AI CoE

四大工作小組

工作小組	聯絡處	合作夥伴	備註
負責任的AI Responsible AI (RAI)	蒙特婁 CEIMIA	The Future Society…..	原「AI與疫情應對」小組 (AI & Pandemic Response, AIPR) 2022年2月起併入RAI
資料治理 Data Governance (DG)		Open Data Institute, Alan Turing Institute, Aapti Institute, ICT Africa, Oxford Insights….	
勞動的未來 Future of Work (FoW)	巴黎 INRIA	OECD….	
創新與商業化 Innovation & Commercialization (I&C)		EU、OECD & UN轄下數機構….	

Src: www.gpai.ai/projects/

AI Connect Program
U.S. State Department/Atlantic Council

- The program consists of a series of virtual webinars and in-person workshops focused on the responsible development, design, and use of AI technologies. Between twelve virtual webinars spaced throughout 2022, the program also includes four in-person regional workshops: one each in the Indo-Pacific region, Africa, Eastern Europe, and South America. A select group of participants will be invited to join in-person workshops corresponding with their regional background with some opportunities for U.S. government funding for travel.
- Throughout the program, AI Connect participants will hear from globally renowned experts in their field; examine case studies and best practices from their own home regions; network with other AI Connect participants; and analyze the most pressing issues in AI policy and implementation.
- How is AI Connect working: https://www.atlanticcouncil.org/programs/geotech-center/ai-connect/

G7部長會議共同宣言

- https://g7digital-tech-2023.go.jp/en/, April 29-30, 2023, Takasaki, Gunma
- In this meeting, we discussed six political agendas:
  - Facilitation of Cross-Border Data Flows and Data Free Flow with Trust,
  - Secure and Resilient Digital Infrastructure,
  - Internet Governance,
  - Emerging and Disruptive Technologies in Innovating Society and Economy,
  - Responsible AI and Global AI Governance,
  - Digital Competition

  As a result of this meeting, "the G7 Digital and Tech Ministers' Declaration" was adopted.
- Points of the Ministerial Declaration are as follows.

共同宣言重點

1. **Facilitation of Cross-Border Data Flows and Data Free Flow with Trust**

   Endorsing the establishment of the Institutional and Arrangement for Partnership (IAP) and agreement on Annex on G7 Vision for Operationalising DFFT (data free flow trust) and its Priorities.

2. **Secure and Resilient Digital Infrastructure**

   Endorsing the G7 Vision of the future network in the Beyond 5G/6G era, which include openness and interoperability as the elements, and the G7 Action Plan for Building a Secure and Resilient Digital Infrastructure.

3. **Internet Governance**

   Endorsing the G7 Action Plan for an Open Free, Global, Interoperable, Reliable and Secure Internet.

共同宣言重點

4. **Emerging and Disruptive Technologies in Innovating Society and Economy**

   Ensuring interoperability of digital infrastructures, cooperation on trust and security across the global value-chain including enhancing transparency of software components, and utilising innovation friendly governance (agile governance)….. we acknowledge the need for agile, more distributed and multi-stakeholder governance and legal frameworks, designed for operationalizing the principles of the rule of law, due process, democracy, and respect for human rights while harnessing the opportunities of innovation.

共同宣言重點

5. **Responsible AI and Global AI Governance**

   Endorsing G7 Action Plan for promoting global interoperability between tools for trustworthy AI.

   Convening G7 discussions on generative AI.

6. **Digital Competition**

   Agreeing to share issues and challenges in promoting digital competition such as those common to G7 members in planning and implementing existing and new laws and regulatory tools and to convene a digital competition summit in the fall of 2023.

AI Summits的影響及其未來

- **AI Safety Summit（英國/韓國：人工智慧安全高峰會）**
  促進AI發展的安全、可靠與值得信賴，並且管理AI衍生的風險
- **UK AI Safety Summit (Bletchley Park), Nov. 2023**
  https://futureoflife.org/project/uk-ai-safety-summit
- **Seoul AI Safety Summit, May 2024**
  https://www.gov.uk/government/topical-events/ai-seoul-summit-2024
- 目前主要參與國家與關切議題：58 countries signed a joint declaration, the *Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet*, at AI Action Summit, Feb, 2025.

United Nations關於人工智慧的倡議

- UNESCO
  https://www.unesco.org/en/artificial-intelligence?hub=32618
- ITU: AI for Good; Inter-Agency Working Group on Artificial Intelligence
  https://www.itu.int/en/action/ai/Pages/default.aspx
- WIPO: AI & Intellectual Property Policy
  https://www.wipo.int/about-ip/en/frontier_technologies/ai_and_ip.html
- Internet Governance Forum (IGF): Multi-stakeholder Working Group of Experts (MWG), the Policy Network on Artificial Intelligence (PNAI) of IGF, United Nations
- UN High-Level AI Advisory Body on AI & Global Digital Compact
  https://www.un.org/digital-emerging-technologies/ai-advisory-body

國際人工智慧治理趨勢2024-2025

一、歐盟

歐盟以2024年人工智慧法（EU AI Act, EUAIA）為基礎：

- **法律面**：於2024年成立AI辦公室（European AI Office），同年5月經歐盟理事會批准EUAIA、7月生效，預計生效後兩年全面適用。為了確保EUAIA在歐盟境內的一致性、效力與統一適用，2025年2月發布禁用AI指引（Guidelines on prohibited artificial intelligence (AI) practices）；Trump Administration 致函要求延緩上述指引的時程；美國公司對 EU Code of Practice on General Purpose AI
- **政策面**：歐盟也公布了新的行動計畫 - AI大陸行動計畫（AI Continent Action Plan），該計畫涵蓋了五大領域：算力基礎設施、資料、技能、演算法的開發與採用、條文簡化，展現了在全球AI競賽維持領先地位的野心。

國際人工智慧治理趨勢2024-2025

二、美國

美國在川普總統二度上任後，已經相當幅度地著手改變既有的AI治理方向：

- 2025年川普上任後即廢除拜登政府的第14110號總統命令（executive order），隨後發布自己的第14179號命令，目的在於移除所謂美國人工智慧領導地位的障礙（removing barriers to american leadership in artificial intelligence），撤除對AI發展的障礙、創新、信賴、擴充可供訓練與使用之聯邦資料等
- 美國管理暨預算局 (Office of Management and Budget, OMB) 也於 2025 年 4 月完成修正並公告兩份新備忘錄：M-25-21及M-25-22取代原先為了14110號命令發布的兩號備忘錄。M-25-22的重點也從 M-24-10高度偏重於AI風險管理，轉向整體採購流程（procurement of AI）規範的建立
- 對於高度影響的使用案例（high-impact use cases）的規範方向

國際人工智慧治理趨勢2024-2025

三、英國

英國逐漸移轉管制重心，從Safety轉向Security

- Alan Turing Institute & National Physical Lab的AI Standards Hub (參見: 連結) 和 AI Safety Summit
- 2025年，工黨上台後，轉向更具約束力的法制。例如：宣示針對強人工智慧模型立法、將人工智慧安全研究所稱之「安全」，從safety更名為security，並預計將其轉化為法定機構，象徵英國新政府已從最初聚焦之偏見與歧視風險，轉向希望能維持人工智慧穩定性，著重於人工智慧的「國安風險」。
- 在基礎建設面向提出擴建國家人工智慧研究資源（AIRR）、推動人工智慧人才培育，並計劃成立國家資料圖書館（National Data Library, NDL），以釋出高影響力公共資料，且鼓勵私人部門共享資料並發展可信任人工智慧。

Alan Turing Institute & National Physical Lab的AI Standards Hub: Mar. 2025, https://aistandardshub.org/global-summit

國際人工智慧治理趨勢2024-2025

四、澳洲

- 研議以強制性防護措施管制高風險AI、其餘則採（產業）自願規範模式
- 2024年8月發布《自願性人工智慧安全標準》，提出供應鏈防護措施；
  接著於9月提出《強制性防護措施提案》，欲針對高風險人工智慧引入強制性的治理措施，相關單位包括澳洲工業、科學及資源部、澳洲法律委員會、澳洲人權委員會、澳洲資訊專員辦公室
- 關於「高風險應如何定義」、「應如何引入強制性措施」仍在研議中。

國際人工智慧治理趨勢2024-2025

五、新加坡

新加坡著重於AI標準形成與國際合作

- 依其發布之國家AI戰略2.0之步調持續發展。
- 2024年5月，由新加坡資訊通信媒體發展局（IMDA）指定數位信任中心（Digital Trust Centre）為新加坡人工智慧安全研究所。
- 2025年2月新加坡數位發展與資訊部部長發布了數項人工智慧政策，包括Global AI Assurance Pilot，以探討生成式人工智慧使用的最佳實踐標準。
- 與日本等國共同合作關於LLMs在非英語語境下的防護措施、以及發布探討LLMs安全性的報告AI Safety Red Teaming Challenge Evaluation Report，較著重於各項人工智慧標準形成的國際接軌與國際合作。

「我們」當前面臨的問題

六、台灣

- 還有加拿大、法國、日韓等國的AI治理方向的發展可供參考.....
- 是否務實地盤點出當前已經出現或者潛在的AI相關規範爭議？
- 是否真正瞭解國際規範趨勢的實質內容？
- 是否明確掌握了台灣現行法規與國際規範趨勢的落差？
- 是否釐清了台灣既有法制與現行規定，對於人工智慧社會的健全發展而言，有哪些不足與漏洞？
- 是否務實評估過應該如何面對這些國際規範趨勢？還是只想採取短線的技術性操作去因應長期的規範問題？
- 台灣的特殊國際處境，對於人工智慧社會發展的意涵為何？
- 現在可以做什麼？?

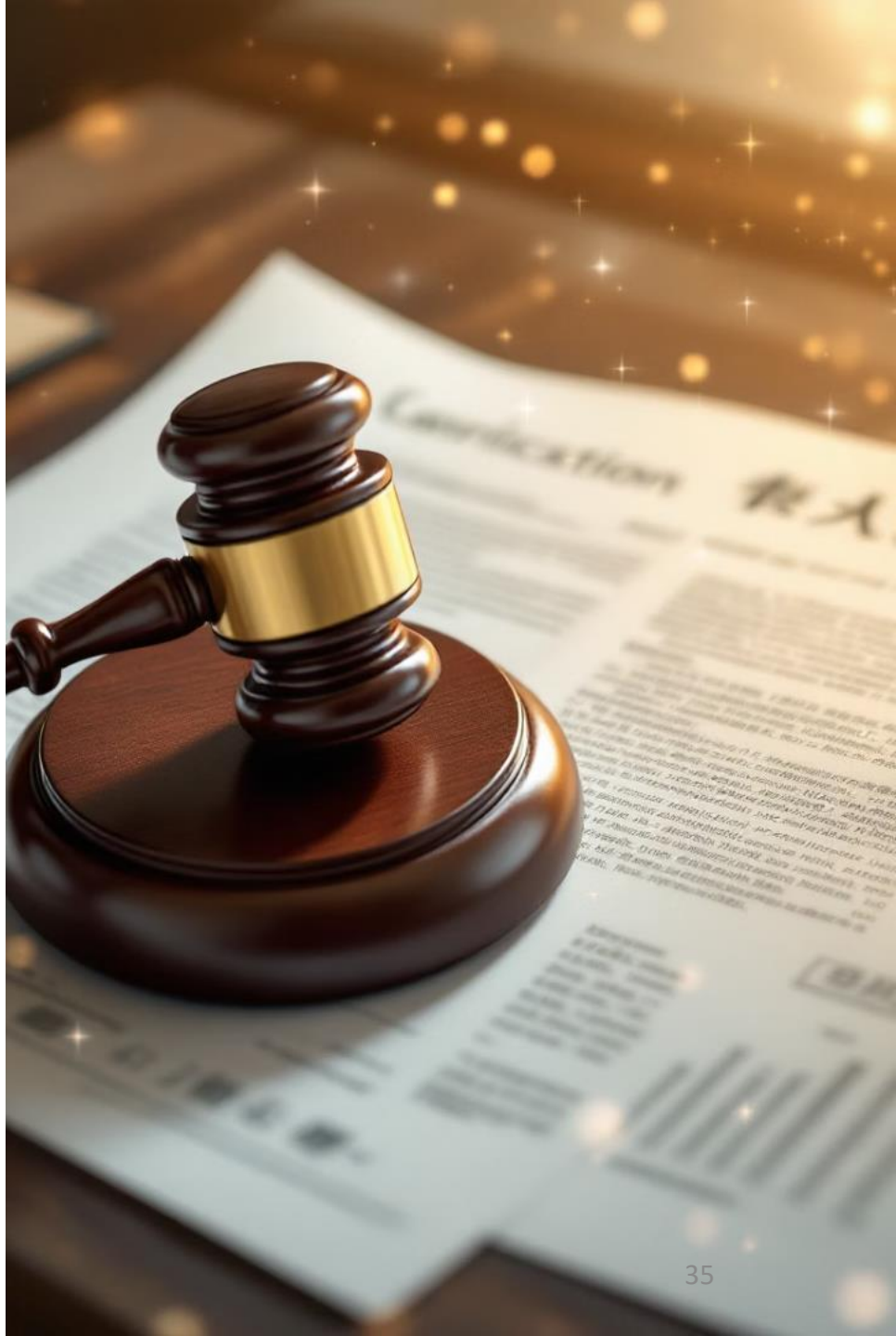
東亞人工智慧治理的發展

南韓、日本、台灣

目次

一、法律架構與目的比較
二、管制對象比較
三、倫理與人權保護的角色
四、個資保護vs資料利用
五、法規鬆綁的方式
六、管制的組織與架構

一、
法律架構與規範目的比較



法律架構與規範目的-南韓

- 2024年12月：《人工智慧發展和信任建立基本法》
  - 全球第二個建立人工智慧全面監理框架的國家。
  - 國會制定的特別法，於2026年1月施行。
- 立法目的
  - 第1條：「本法旨在規範促進人工智慧的健全發展與建立信任基礎所需的基本事項，以保護國民的權益與尊嚴，並致力於提升國民生活品質與強化國家競爭力。」
    - 對人工智慧進行適當規範來建立社會信任基礎
    - 提高國家人工智慧競爭力

法律架構與規範目的-日本

- 管制模式：
  - 2023年以前：主要依靠軟法（如倫理指引、產業指針）治理人工智慧。
  - 2024年後：新內閣明確主張「促進創新並因應風險」的人工智慧框架。
  - 2025年2月，內閣提出《促進人工智慧相關技術研究開發及利用法》。
- 立法目的：
  - 第1條：「鑑於人工智慧相關技術已成為我國經濟社會發展的基礎性技術，為推動人工智慧相關技術的研究開發與應用，規定基本理念，並就推動人工智慧相關技術研究與應用的基本計畫等施策的基本事項，以及設立人工智慧戰略加以規範。...」

法律架構與規範目的-台灣

- 尚未正式立法
  - 2023年：行政院及所屬機關(構)使用生成式AI參考指引
  - 2024年：國家科學及技術委員會於2024年7月提出《人工智慧基本法》草案。嗣後轉由數位發展部擔任主管機關。
  - 各黨立委草案版本？
- 立法目的：
  - 第1條：「為促進以人為本之人工智慧研發與應用，維護國民生命、身體、健康、安全及權利，提升國民生活福祉、維護國家文化價值及國家競爭力，增進社會國家之永續發展，特制定本

比較分析

- 三國的人工智慧法律架構各異：
  - 南韓已率先由國會通過專法實施全面監管。
  - 日本傾向既有法律及軟性指引、僅制定促進創新為主的法律。
  - 台灣則在草擬階段，預計制定一部融合國際經驗的原則性基本法。
- 重點略有不同：
  - 南韓明確提出建立信任與保障公民權益。
  - 日本強調社會經濟發展與國際協調。
  - 台灣則主張人權保障、社會福祉與產業永續並重。

二、管制對象比較

人工智慧定義-南韓

- 明確區分不同類型人工智慧
  - 韓國人工智慧基本法明確定義並區分「人工智慧」、「人工智慧系統」、「人工智慧技術」、「高影響人工智慧（high-impact AI）」及「生成式人工智慧」（第2條）。
    - 「人工智慧」指以電子方式實現學習、推理、感知、判斷及語言理解等人類的智慧能力。（第1款）
    - 「人工智慧系統」指為實現特定目標，具有各種程度的自主性與適應性，能對現實及虛擬環境產生影響，進行預測、推薦、決策等結果推導的人工智慧基礎系統。（第2款）
    - 「人工智慧技術」指為實現人工智慧所需的硬體、軟體技術或相關應用技術。（第3款）
    - 「高影響人工智慧」指對人的生命、身體安全及基本權利產生重大影響或有招來風險之疑慮，並應用於下列各目領域之一的人工智慧系統…（第4款）

人工智慧定義-日本

- 人工智慧促進法：
  - 第2條第1項：「以人工方式實現認知、推理、判斷過程的技術，以及產出此類結果的資訊處理系統」。
    - 廣泛涵蓋智慧計算技術，但未區分具體的類型。
- 避免對人工智慧一刀切的新規制，而優先運用既有領域法規，因此沒有建立類似歐盟的風險分類制度。僅在基本原則中提及對可能產生負面後果的人工智慧應注重「透明適切」以防範風險。

人工智慧定義-台灣

- 人工智慧基本法草案第2條：「人工智慧係指以機器為基礎之系統，該系統具自主運行能力，透過輸入或感測，經由機器學習與演算法，可為明確或隱含之目標實現預測、內容、建議或決策等影響實體或虛擬環境之產出。」
- 草案立法理由中明確區分不同「風險」的人工智慧。
  - 要求主管機關研參國際標準建立人工智慧風險分類框架：例如參考歐盟人工智慧法案的四級風險分類（包含對禁止行為的界定）。
  - 授權數位發展部等單位比照國際制定相應的分類準則 。
- 然而，截至草案階段，「高風險人工智慧」與「生成式人工智慧」等定義在法律中可能尚未細化列舉，而是留待未來制定子法或指引時明確。

比較分析

- 歐盟法的「風險分級」規範模式（歐盟人工智慧法第3章以下），影響了南韓及台灣治理模式。但仍有不同之處。
  - 南韓法制除了明確詳盡區分不同人工智慧外，著重「影響（impact）」，而非僅是「風險」。
  - 台灣行政院既有草案則趨向歐盟模式，強調建立分類管理，但細節有待配套規定完善。
- 日本迄今採取技術中立的寬泛定義，沒有正式風險分級，在法律上不區分高風險類別。

差別化的人工智慧管制-南韓

- 高影響人工智慧
  - 透明要求：高風險人工智慧的提供者須履行告知義務，在應用高風險人工智慧時向使用者事先告知其為人工智慧運作（第31條）。
  - 相關營運者在使用前進行影響評估，確認系統符合安全可靠性要求，並採取必要措施保障安全與可信任（第33、34、35條）。
  - 若業者違反告知、高風險事前審查等義務，主管機關（科學及資訊通訊部，Ministry of Science and ICT）可下令改正，並對未改正者處以罰款等懲處（最高可罰3,000萬韓元）。

差別化的人工智慧管制-南韓

- 對於生成式人工智慧的受知情權（第31條）
  - 一般的生成式人工智慧（第1項）
    - 事前受知情權：事前告知生成式人工智慧使用的使用。
    - 事後受知情權：標示義務。
  - 逼真內容的生成式人工智慧：嚴格的標示義務（第3項）
    - 「人工智慧事業者在提供使用人工智慧系統生成的、難以與真實區分的虛擬聲音、影像或影片等結果時，應以使用者能夠清楚識別的方式告知或標示該結果是由人工智慧系統生成的事實。在此情形下，若該結果屬於藝術或創意表現物，或構成其一部分之情形，得採取不妨礙展示或享用等的方式進行告知或標示。」

差別化的人工智慧管制-日本

- 高風險人工智慧：並未處理
- 生成式人工智慧：產業自治及倡議
  - 2023年10月，日本內閣發布了行政機關使用生成式人工智慧指南。
  - 產業界方面，經產省鼓勵企業遵循人工智慧治理指引，當中包括避免生成式人工智慧濫用、防止人工智慧輸出侵害版權或隱私等。
  - 日本執政黨自民黨在2024年也呼籲建立相應規範。

差別化的人工智慧管制-台灣

- 高風險人工智慧：授權管制
  - 要求數位發展部參照歐盟等國做法建立人工智慧風險分類框架，並建立風險評估機制（第9條、第10條）。
- 生成式人工智慧：尚未獨立管制
  - 草案條文未提及「生成式人工智慧」
  - 立法理由中引用了新加坡2023年提出的「生成式人工智慧治理架構草案」及英國2023年提出的「生成式人工智慧治理框架」作為參考。

比較分析

- 歐盟人工智慧法
  - 第三章就「高風險人工智慧」加以管制（第6條至第49條）
  - 第五章就「通用人工智慧模型」加以管制（第51條至第56條）。
- 受到歐盟法影響，南韓已經在法律中對高風險人工智慧和生成式人工智慧明定義務。
- 日本迄今在法律上對這兩類型AI皆無直接規範。
- 台灣則有意在基本法框架下授權主管機關管制高風險人工智慧。但並未處理生成式人工智慧。

三、
倫理與人權保護的角色

倫理與人權保護的角色-韓國

- 軟法功能及公私協力
- 政府有義務頒布「人工智慧倫理原則」（第27條第1項）
  - 確保人工智慧開發與應用等過程中的安全性及可信賴性。
  - 任何人均能自由且便利地使用人工智慧技術
  - 促進人類生活與繁榮，開發與應用人工智慧的相關事項。
- 民間「得」設立人工智慧倫理自治委員會（第28條）：
  - 調查、確認研發應用是否符合倫理要求及具有安全性
  - 企業教育以及提供意見
  - 落實主管機關的管制需求。

倫理與人權保護的角色-日本

- 軟法及倡議
- 日本政府自2019年起即提出《人間中心人工智慧社會原則》，強調尊重人類尊嚴、資訊隱私、公平等價值。
- 2022年的《人工智慧戰略》中亦將尊嚴、多樣性與包容、公平正義列為指導原則 。
- 2024年4月經產省發布的《企業人工智慧治理指引》1.0版進一步細化了10項原則，包括：安全、公平、隱私保護、透明、問責、人權尊重、教育素養提升等。

倫理與人權保護的角色-台灣

- 政策性宣示
- 草案第3條提出了七大基本原則作為政府及機關推動人工智慧時的指導方針，包括：永續發展與福祉、人類自主、隱私保護與資料治理、資安與安全、透明與可解釋、公平與不歧視、問責。
  - 借鏡Hiroshima AI流程的行為準則、OECD AI建議、美國人工智慧權利憲章藍圖及歐盟可信賴人工智慧倫理指引等國際文件 。

比較分析

- 在倫理與人權保障上，南韓和台灣選擇了將主要原則納入法律框架來引導施政。日本則依賴政策性文件和指南，雖然沒有明文法律條款，但實際上政府和企業普遍認可這些原則。
- 不過，共同的特色（問題）是三國的人權或倫理條款都沒有具體效力。即便是宣示，也欠缺細緻的說明各該人權要求應該如何落實，應如何與法規鬆綁權衡。

四、
個人資訊隱私保護與資料利用的衝突

個資保護與資料利用的衝突-韓國

- 資訊隱私保護相關條文
  - 第13條privacy-by-design
  - 韓國人工智慧基本法沒有針對資訊隱私有特別立法，仍由個人資料保護法擔任管制腳院。
- 近期發展
  - 韓國個人資料保護委員會正在草擬個資法的修法，緩解人工智慧發展及應用的個人資料管制。

個資保護與資料利用的衝突-日本

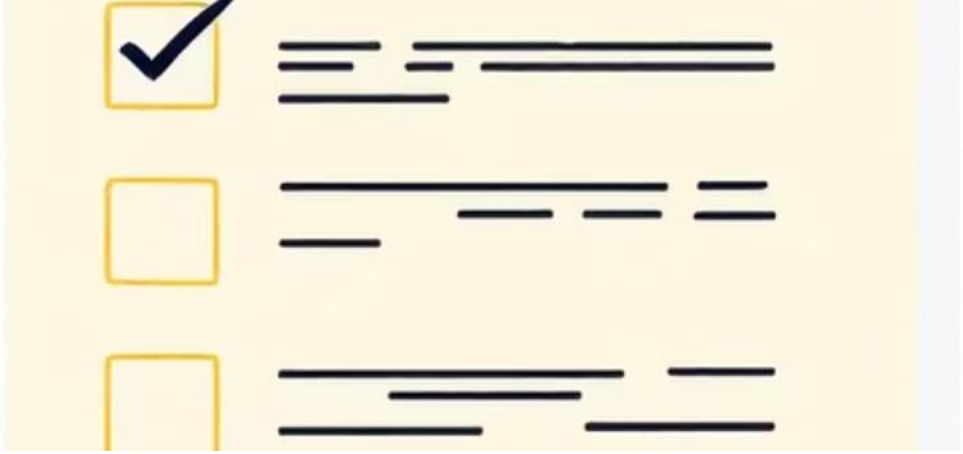
- 個人資料保護委員會考慮放寬某些情形下對敏感個資收集的事前同意要求，以便利人工智慧開發。
- 目前日本整體上而言仍依賴既有法規：例如，用人工智慧處理個人資料，仍須遵守日本個人資料保護法，取得同意或匿名化處理；人工智慧決策如果涉及消費者，企業須遵循消費者保護法上關於說明說明的義務等。
- 日本沒有針對人工智慧另外訂定新的資料治理規則，而是仍由個人資料保護的一般規範擔任管制角色。

個資保護與資料利用的衝突-台灣

- 人工智慧基本法草案並未豁免個人資料保護法的適用。
- 第14條明訂個人資料保護機關應協助各主管機關在人工智慧研發應用中防止不當的資料蒐集、處理或使用。
- 人工智慧時代的資料治理法制需求
- 中文訓練資料的需求
- 主權AI的需求與可能發展方向

比較分析

- 三個國家對於個人資料保護的情形，並未因爲人工智慧發展的利益就直接讓步。三國立法者均尚未針對個人資料保護在人工智慧時代，提出明顯不同的因應模式。
- 但這個現象未來如何發展，仍有待觀察，已經有跡象顯示各國主管機關打算透過修正既有規範，調整和人工智慧研發有關的個人資料應用規範。



五、
法規鬆綁的方式

法規鬆綁的方式-韓國

- 「監理沙盒」規定於第19條第2、3項
  - 「政府為支援人工智慧整合產品及服務的開發，必要時，得授予國家研究開發創新法下之國家人工智慧研發計畫優先權。」
  - 「關於第 2 項開發的人工智慧整合產品及服務， 政府應依資訊通訊振興及整合法第 37 條及第38條之2規定，給予臨時許可及豁免實驗測試規制得以順利開發。」

法規鬆綁的方式-日本

- 並無明確的法規鬆綁規定
  - 日本強調不用新的規制替代靈活運用既有法規。政府傾向讓人工智慧企業在現行法律框架內先行探索。
  - 日本內閣在2025年2月的中期報告中明確指出，除非自律措施不足，否則應盡量避免新增規制。

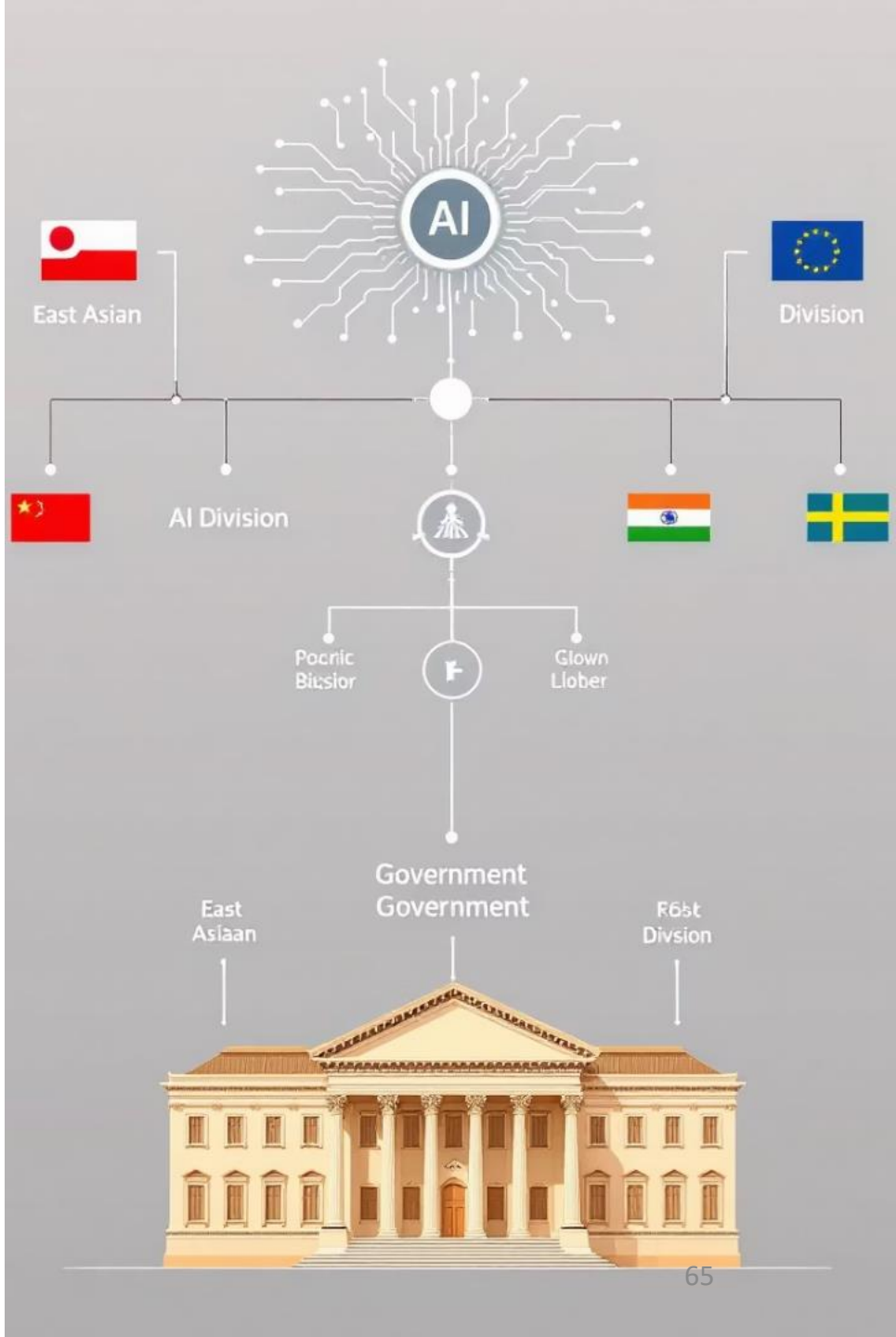
法規鬆綁的方式-台灣

- 草案明文強調「法規調適及業務檢視」（第5條），要求政府在推動人工智慧研發應用時，以不妨礙新技術和服務為原則來完善相關法規。
- 草案規定各主管機關應建置或強化人工智慧研發與應用服務的「創新實驗環境」（第6條）。
- 草案並建議在法施行三年內完成法規盤點，並可能推出跨領域的人工智慧沙盒計畫，降低新創人工智慧產品的合規門檻但附帶風險控管要求，以鼓勵試驗創新。

比較分析

- 南韓法制將沙盒法制化，但限於與「國家發展」有關的人工智慧研發，領域相當狹隘，而且取決於行政機關的判斷。台灣也意欲將將沙盒放入法律中，但是尚無明確的規範。
- EUAIA的第57條至63條，處理沙盒的規範。歐盟法基本上認為沙盒能夠促進「創新」，且能在行政管制上提供測試，減少行政負擔。並且給予個人資料「目的外利用」的條件以及誘因。
- 南韓法制僅限於涉及國家發展的人工智慧，誘因似乎有限，而且豁免的情形有限。

六、
管制的組織與架構


AI
East Asian
Division
AI Division
Government
Government
East
Asiaan

組織與架構-南韓

- 決策層面：
  - 總統直屬的「國家人工智慧委員會」（第7條），最多可有45名成員，包含政府高官及民間專家。負責審議國家人工智慧基本計畫統籌人工智慧治理方向。
- 執行層面：
  - 科學及資訊通訊科技部。該部設置「人工智慧政策中心」執行日常政策研擬並運作「人工智慧安全研究所」負責技術標準、風險評估方法等（第11條）。

組織與架構-日本

- 人工智慧戰略本部：最高協調機關，由首相擔任本部長，全體內閣大臣為成員。確保人工智慧議題上升到國家戰略高度，各部門協同推進。（第19條）
- 不另外成立專門的人工智慧管制機關，沿用各主管省廳按照既有法規監督人工智慧。
- 私部門的管制：依賴行業組織和自律團體，如軟體協會制定人工智慧倫理守則等。

組織與架構-台灣

- 主管機關：數位發展部（**或**國家科學國家科學及技術委員會）（目前草案方向尚未趨於清晰）
- 各目的事業主管機關依據基本法原則，在各自領域修訂配套規範。
- 管制之執行上，個資保護委員會、通訊傳播委員會等獨立機關在其權限範圍內協助參與。
- 在罰則與執法方面，台灣草案傾向不直接做為處罰依據，而透過既有法律來處理違規。

比較分析

- 韓國及台灣皆未建立新的管制機關，而是強化既有部門的統籌權。其他機關及地方政府在人工智慧推廣和管理上有配合義務。共同的挑戰則是：如何避免責任劃分不清所帶來的治理效率低下問題？
- 歐盟為了整合各會員國，在歐盟內部及會員國層級都建立了新的機關或單位，以達到協調及配合並落實執法。

主權人工智慧：

從數位主權的國際競爭分析我國發展途徑

2025/5/14

目次

前言：從數位主權到主權人工智慧

從黃仁勳到國家政策、從商業噱頭到政府宣示決定發展主權人工智慧，在此一發展脈絡下，我們實有必要釐清：

- 主權人工智慧的意涵為何？
- 為什麼各個國家會呼應特定商業公司的主張，發展主權人工智慧？
- 我國究竟是否應該發展主權人工智慧嗎？
- 我國目前發展主權人工智慧的現狀又是如何？
- 我國在發展主權人工智慧方面，到底有何等優勢與劣勢？

01

主權人工智慧的概念、目的及落實

主權人工智慧究竟所指為何？
為什麼各國政府積極發展主權人工智慧？
各國是透過哪些途徑發展主權人工智慧？

(一) 政策內涵：主權人工智慧作為科技國族主義的展現

- 主權AI為近年各國逐漸成形的「科技國族主義」發酵的結果
- 傳統科技國族主義接受全球化、重視透過補助強化本國企業競爭力；新興科技國族主義則基於國安等考量比較傾向排斥全球化，並藉由跨境干預抑制他國科技發展。
  - 在數位主權的實踐上，許多國家透過限制虛假資訊、言論內容與資料流動展現主權延伸至雲端空間的治理意圖。
  - 在資料治理方面，例如中國推行資料在地化，歐盟透過GDPR賦予個資規則域外效力，都展現出政府控制資料流向與使用之傾向。

（二）政策目的：確保科技發展地位及建構妥適人工智慧

- 主權AI的第一個政策目的，是確保國家在AI技術領域取得或維持領先地位，回應科技國族主義中「科技即國力」的主張
- 第二個政策目的，是確保AI應用能切合本土脈絡，避免他國開發的AI系統錯置或誤解在地情境，例如台灣警方無法直接套用美國犯罪預測模型。
- 此外，主權AI亦可維護語言與文化脈絡之多樣性，避免他國主導生成式AI輸出錯誤資訊或價值偏誤，例如繁體中文使用情境遭簡化誤解等問題。

（三）政策現狀：各國發展主權人工智慧的不同情形

- 新興科技國族主義
  - 透過出口管制阻斷他國技術取得，例如美國對晶片、半導體、量子科技等人工智慧原料的出口，採取多層次的許可制。
- 傳統科技國族主義
  - 措施包括建置基礎建設、建置本國的訓練模型、調整技術移工的簽證政策。

建置國營基礎建設

法國斥資4千萬歐元升級超級電腦，歐盟在七個國家建置超級電腦，打造「人工智慧工廠」

訓練本土模型

日本、荷蘭、新加坡、西班牙、瑞典、台灣等國家投注資源建置本土模型

調整簽證政策

美國立法草案提出AI人才的簽證優惠，吸引專業人才進駐

02

對主權人工智慧政策的主要批評

初步整理重點

批評

- 批評一：科技國族主義可採？主權人工智慧應否發展？
- 批評二：政府主導人工智慧發展，將導致政府濫用其主導地位？
- 批評三：單一國家事實上無能力自行發展主權人工智慧？

批評一：科技國族主義可採？主權人工智慧應否發展？

- 對科技國族主義的批評多來自科技全球主義立場，認為此類政策將導致通用科技更加破碎、失去整合性，進而加劇國際科技落差，並降低創新可能性。
- 在缺乏單一價值標準的國際社群中，科技全球主義與國族主義各有其立場，難以一概而論。
- 是否採納主權AI並非價值選擇問題，而是應該回到政策如何實際落實、是否能發揮其目的的實用性判準。

批評二：政府主導人工智慧發展，將導致政府濫用其主導地位？

- 有批評指出，當主權AI由政府主導發展，可能會透過設計或控制，使AI產生特定意識形態或偏頗資訊，影響人民的資訊接受權利與民主體制的穩定。
- 生成式AI逐漸成為人們日常生活中的輔助決策工具，若其中內嵌政治意識型態或虛假內容，將嚴重危及資訊自由與公共討論空間。
- 若要避免此類濫權風險，需建立妥適的AI治理架構，從資料取得、處理到研發、應用階段皆納入規範，確保AI系統的可信賴性與公正性。

批評三：單一國家事實上無能力自行發展主權人工智慧？

- 存在現實不可行性，因AI開發所需的運算力、資料、專業人才與半導體原料，均非任何單一國家能獨立掌握。
- 即便具備資金興建高運算設施，仍可能受限於半導體製程、水準、原料取得等問題；更遑論人才整合亦曠日費時。
- 仍可透過「數位連帶」（digital solidarity）的模式，建立國與國之間的資源互補與合作網絡，以提升AI自主發展的可行性與民主治理的韌性。

03

對於我國主權人工智慧發展的政策建議

初步構想

（一）釐清主權人工智慧開發目的進而設計及開發本土人工智慧

- 觀察近年政策與「TAIDE」、「臺灣杉」建置情形，大致可得知我國主權人工智慧的政策，仍是朝向確保人工智慧在我國本土可以妥適使用的基本目的，至於是否進一步追求在國際間的人工智慧發展上佔據競逐上的戰略地位，則仍不明朗。
- 目前我國學研領域最接近主權AI的項目－TAIDE，在建置與訓練過程中，因訓練資料取得所遭遇的困境，以及與訓練資料相關的法律規範的建構問題，都值得進一步研究。

（二）從資料治理、人工智慧研發到應用所需的完善人工智慧治理規範

- **缺乏AI相關的專法**；現行如個人資料保護法、政府資訊公開法及著作權法等等，均顯現出**規範不足**的缺點。
- 針對資料治理，我國政府既有的資料（包括個資及非個資），應屬重要訓練資源，至少應有以下具體規範方向：
  - 已公開者可使用；未公開則應符合特殊條件始可利用；
  - 非公開資料為非個資，在不危害特定利益的情況下，容許利用；
  - 涉及個資，須符合個資再利用的憲法要求，即匿名處理與資料當事人自主控制權
  - 透過獨立管制機制，審查前開要件是否符合，並從事風險評估
- **《人工智慧基本法草案》仍停留於宣示層次，即使立法通過，仍須更多「作用法」或至少「指引」的配套**
- **可參考歐盟《人工智慧法》（EU AI Act）建立風險導向之規範架構**

（三）為人工智慧的發展建立穩固數位連帶

- 單一國家獨立發展人工智慧有事實上困難，各國政府應該確保穩固的「數位連帶」關係。
- 政府應釐清我國在全球AI供應鏈的策略地位－－**如何有效運用我國晶片產業的國際地位，以確保我國經濟安全的穩固：**
  - **具體政策：**例如，禁止所有本國晶片製造商的先進技術與製程出海晶片產業優勢；海外設施則僅能使用上一代的製程技術；
  - 確認缺乏何項資源，以挑選**合適合作夥伴；**
  - **AI人才短缺問題嚴重**，應同時推動本土培育與科技移工政策，引進國際人才，但亦應設限確保資訊安全與本土知識生產優先。

結論

在2025年4月9日，歐盟執委會正式公布了其醞釀多時的「人工智慧大陸行動計畫」（AI Continent Action Plan），其內容重點在於宣布將透過五大核心支柱（five key pillars)，亦即：AI基礎設施、高品質資料、產業AI化、AI人才和法規制度強化等五個面向的諸多具體措施，積極推進歐盟的AI實力，這個行動計畫除了呼應今年2月在法國巴黎AI行動高峰會（AI Action Summit）上，歐盟執委會主席Ursula von der Leyen所宣示的歐盟立場：全球的AI競賽才剛開始，真正的領先者尚未出爐，歐盟將盡其所能，確保歐盟能夠角逐全球AI領導者的地位

結論

這個歐盟方面的重大發展，無異於再次宣示了歐盟對於主權AI高度重視的基本立場，值得關注追蹤，而其即將針對以上五個面向採取的具體政策與措施，更是值得台灣比較甚至仿效。

台灣政府及業界對於主權人工智慧發展的重視，值得肯定。然而，綜觀台灣的科技法制現狀和政府政策，的確對主權AI的發展尚未釐清特有脈絡和應該補強之處，這將對台灣發展主權AI產生不適切結果。

政府若要積極協助我國人工智慧的發展，應更加審慎地規劃主權AI政策，並且根據妥善規劃後的政策目標做好配套措施，才能為我國發展「主權人工智慧」建立必要基礎和架構。

Thanks!

E-Mail: cytahr@gmail.com