出國報告（出國類別：訪問）

國家科學及技術委員會
荷蘭 World Summit AI 論壇參訪行程

服務機關：國家科學及技術委員會
姓名職稱：王凱平科長
派赴國家：荷蘭(阿姆斯特丹)
出國期間：114 年 10 月 4 日至 114 年 10 月 11 日
報告日期：114 年 11 月 3 日

摘　要

本次出國目的為參與 2025 年阿姆斯特丹世界人工智慧週（World AI Week）與人工智慧世界高峰會（World Summit AI），其為全球規模最大的國際性人工智慧活動，為探索人工智慧未來發展以及建構人工智慧治理的關鍵議題。本次會議主題「回到未來：時機已至」具有三重意涵：1.歷史借鑑：從人工智慧發展歷程中汲取經驗與教訓；2.現實挑戰因應：解決當前人工智慧發展面臨之技術、倫理與監管難題；3.未來軌跡塑造：規劃人工智慧對人類社會長期發展之積極影響。

我國在人工智慧的推動上，行政院已於 2023 年核定臺灣 AI 行動方案 2.0，從產業端出發，透過深耕 AI 技術與發展 AI 產業及產業應用 AI，帶動我國整體產業轉型升級。此外，我國政府亦推動「AI 新十大建設方案」進一步推動百工百業導入 AI 技術，驅動產業數位應用落地，並創造產業新市場。因此，借鏡國際上對於人工智慧發展及人工智慧治理法制發展趨勢，研析我國人工智慧發展策略及治理法制之調適方向，為本次出訪重點。

我國人工智慧基本法草案，已於 113 年 7 月 15 日預告，經過多番討論，在 114 年 8 月 28 日通過行政院審議，草案揭示永續性、人類自主性、隱私保護及資料治理、安全性、透明性及可解釋性、公平性、可問責性等七大基本原則，以及創新合作及人才培育、風險管理及應用負責、權益保障及資料利用、法規調適及業務檢視等四大推動重點，作為引導我國各機關研發與應用人工智慧之原則。

本次參與荷蘭阿姆斯特丹「世界人工智慧週」與「世界人工智慧高峰會」對後續科學發展、技術研究與應用政策上具參考價值，可做為臺灣後續施政推動之借鏡。

目　錄

表目錄

圖目錄

壹、目的

人工智慧技術近年發展快速，被世界普遍認為可為整體產業與社會活動帶來廣泛之經濟及社會效益，並為我國企業及國家發展提供關鍵之競爭優勢。在氣候變遷、環境、醫療、金融、交通、內政、農業、公共服務等對民眾具廣泛影響力之領域中，更亟需積極採用人工智慧技術以推動數位轉型與永續發展。

人工智慧技術雖帶來經濟及社會效益，同時也可能對個人或社會帶來風險或影響。鑑於人工智慧技術創新之速度及可能面臨之挑戰，全球主要國家皆致力在不妨礙技術發展下，尋求建立人工智慧之治理方針與原則。

2025 年荷蘭阿姆斯特丹「世界人工智慧週」與「世界人工智慧高峰會」不僅為全球人工智慧領域最重要之年度盛會，更為推動人工智慧技術發展、促進國際合作、建立治理框架之關鍵平台，對於塑造人工智慧未來發展方向具有深遠影響。

我國現行 AI 發展政策以多元布局架構，透過跨部會協力共同推動，實現「人工智慧島」的國家願景。為因應人工智慧技術創新之速度和可能面臨的挑戰，行政院 2025 年 8 月 28 日院會審查通過「人工智慧基本法草案」，期建立我國 AI 發展的基本方針。後續人工智慧基本法要落實到 AI 應用的各行各業，完善 AI 運作環境，必須由各部會依基本法規定檢討及調整所主管之法規及業務，由各部會在技術、應用、治理、人才培育、國際合作、基礎環境整備等面向，共同推動我國 AI 發展。因此，期透過本次荷蘭阿姆斯特丹「世界人工智慧週」與「世界人工智慧高峰會」的活動，獲取國際發展經驗與掌握國際最新趨勢。

貳、行程與議程

一、行程

本次行程自 114 年 10 月 4 日（六）起至 10 月 11 日（六），共計 8 日，行程如下：

國別	時間	行程
臺灣 → 荷蘭	10/4-5 (六) (日)	• 臺灣臺北→荷蘭阿姆斯特丹 臺灣時間10/4　　出發(臺灣桃園機場) 荷蘭時間10/5　　抵達(荷蘭史基浦機場) 搭乘中華航空
荷蘭	10/6 (一)	• 世界人工智慧週（World AI Week） • ALLAI 負責任 AI 研討會
荷蘭	10/7 (二)	• 世界人工智慧週（World AI Week） • AI & Partners AI 治理研討會
荷蘭	10/8 (三)	• 世界人工智慧週（World AI Week） • 人工智慧世界高峰會（World Summit AI）
荷蘭	10/9 (四)	• 世界人工智慧週（World AI Week） • 人工智慧世界高峰會（World Summit AI）
荷蘭 → 臺灣	10/10-11 (五) (六)	• 荷蘭阿姆斯特丹→臺灣臺北 荷蘭時間 10/10　　出發(荷蘭史基浦機場) 臺灣時間 10/11　　抵達(臺灣桃園機場) 搭乘中華航空

表 1 訪團行程表

二、會議議程

(一) 世界人工智慧週（World AI Week）

World AI Week（世界人工智慧週）是一場全球人工智慧領域最具代表性的國際盛事之一，於 2025 年 10 月 6 日至 9 日在荷蘭阿姆斯特丹舉辦，為全球規模最大之人工智慧活動。

World AI Week 由全球知名組織 Inspired Minds 主辦，該機構同時也是 World Summit AI 與 Intelligent Health Summit 的策劃者。這是一場為期五天、橫跨商業、科技與學術領域的盛會，涵蓋超過 50 場分場活動，吸引來自全球的企業高層、技術專家、創業者與學者逾 15,000 名參與者共同參與，範圍涵蓋科技巨擘至新創企業、投資者、協會及公共機構。

圖 1 World AI Week 活動會場

(二) ALLAI 負責任 AI 研討會

ALLAI 是一個致力於推動和培育負責任的人工智慧的獨

立組織，由歐盟人工智慧高級專家小組的三位荷蘭成員 Catelijne Muller、Virginia Dignum 和 Aimee van Wynsberghe 發起。

本次 ALLAI Responsible AI Conference 於 2025 年 10 月 6 日在阿姆斯特丹科學園區舉辦， 是一場專注於負責任人工智慧（Responsible AI）的重要會議，研討會目的是為了使人工智慧以負責任且永續的方式融入社會，包容性是理解人工智慧挑戰、建立對負責任的人工智慧應用的信任的關鍵。而人工智慧需要一種持續、系統化的方法，從各個角度審視這項技術，集結多學科、政策制定者、各領域的學者、社會夥伴、企業和非政府組織持續合作討論。

圖 2 ALLAI Responsible AI Conference

(三) AI & Partners AI 治理研討會

AI & Partners 是一家專注於 AI 治理和合規的領導者與專家團體，本場會議 AI Governance Conference 於 2025 年 10 月 7 日在阿姆斯特丹市區舉辦，深度研討歐盟人工智慧法案，該法案係全球首部人工智慧領域之綜合性法律框架，對各

國之企業、開發者及監管機關均具重大意涵。本次活動提供政策洞察與實務指導之獨特結合。人工智慧監管、倫理規範及法規遵循領域之專家學者將深入解析該立法之核心要素、對不同利害關係人之意涵，以及如何為即將到來之執法作準備。

圖 3 AI & Partners AI 治理研討會

(四) 人工智慧世界高峰會（World Summit AI）

2025 年世界人工智慧高峰會於 10 月 8 日、9 日兩天，在 Taets 藝術暨活動園區舉行，與世界人工智慧週同期舉辦，為其主要論壇活動。

World Summit AI 廣受認可為全球領導性人工智慧研討會。本次主題「重返未來：是時候了」旨在強調人工智慧對科技與社會的深刻影響，以及各國和組織在制定 AI 策略方面的加速和緊迫性。它呼籲在 AI 快速發展的同時，進行倫理反思和果斷行動，活動匯聚逾 7,500 名參與者、200 名演講者，以及來自 160 餘國之代表，設有 17 個討論主題，涵蓋 AI

的最新發展、風險、應用、倫理與政策，並吸引超過 200 位專家講者登台發言。

圖 4 World Summit AI 會場入口

參、研討議題一、促進 AI 應用

促進 AI 應用，並讓政府、企業及學研機構加速導入 AI，為本次活動熱門主題。根據活動主辦方調查，歐洲 78%的組織在其業務中至少中使用一個 AI 功能。相較於一年前統計數據為 55%，且其中 71%採用生成式 AI 模型，整體有大幅提升的趨勢。

本議題主要討論加速 AI 導入之方式，主要需要結合政策支持、企業準備度、人才、技能及基礎設施上的投資，以及負責任的 AI 治理，才能促使 AI 導入各行業，增進創新及效率的競爭優勢。

圖 5 World AI Week：加速 AI 導入專場

一、歐盟「AI 大陸行動計畫」

2024 年 9 月，歐盟執委會收到前義大利總理 Mario Draghi 的報告「歐盟競爭力策略」，對於歐盟缺乏產業動態彈性、創新及投資量能，敲響歐盟對於整體競爭力的警鐘。本次高峰會一

重要部分，在探討歐盟與美國之間，創新能量的落差，特別是在高科技領域自 2013 年起歐盟的研發投資已落後美國。因此，2025 年 4 月，歐盟發布「AI 大陸行動計畫」（AI Continent Action Plan），該計畫由五大支柱組成，首要為基礎建設，預計打造 13 個歐洲超級電腦（EuroHPC supercomputers）給會員國使用；第二，則是強化 AI 開發者的資料可近用性，第三，是確保戰略性產業及公部門使用 AI；第四，培育 AI 專業人才；第五，簡化現有法規。

二、Google 導入 AI 案例

圖 6 World AI Week：Google 攤位

(一)　Princess Maxima Center 兒科腫瘤中心

荷蘭 Princess Maxima Center 兒科腫瘤中心的兒科腫瘤學家面臨的挑戰，是掌握兒科腫瘤醫學研究大量且不斷增長的數據，光 PubMed 一個資料庫就擁有 3,700 萬篇論文，遠遠超出臨床醫師的閱讀負荷量。為此，荷蘭兒科腫瘤中心與 Google 合作，開發 Capricorn。這是一款由 Google Gemini 驅動的 AI 工具，可以掃描 PubMed 中的論文，並將論文研究與實際去識別化的患者病情比對，從而更快、更全面的發現潛在的治療方案。為臨床醫生節省大量時間，以前需要兩三天才能完成的事情，AI 只需要 40 秒就能完成。

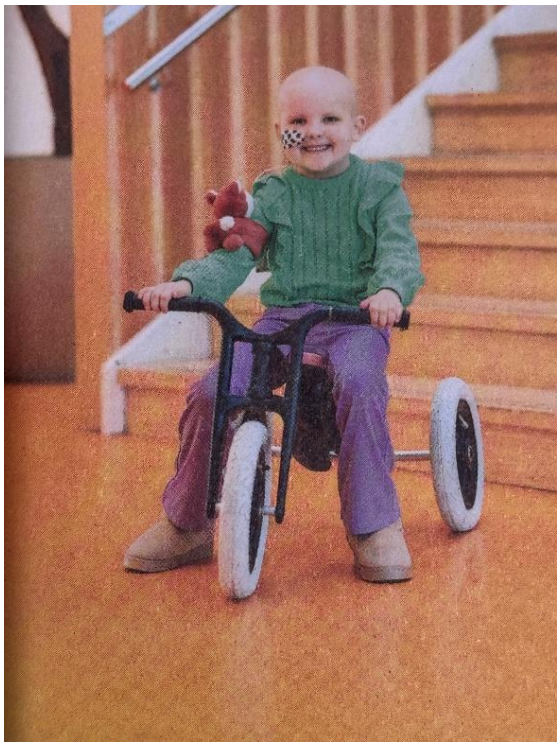


圖 7 World AI Week：Princess Maxima Center for pediatric oncology

(二) Cradle

荷蘭研究機構 Cradle 與它的 AI 平臺，主要從事藥品及食品領域的蛋白質工程。Cradle 執行長 Stef van Grieken 表示，透過 AI 平臺加速蛋白質設計工程，讓機構能夠更經濟有效地應對氣候變遷和疾病等全球挑戰。

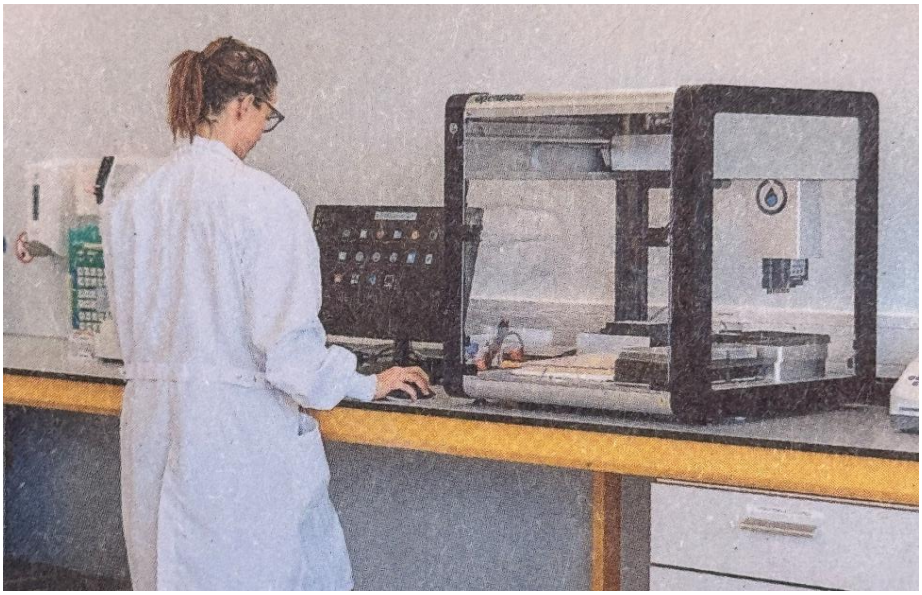


圖 8 World AI Week：Cradle Organization

(三) SUSE AI

SUSE AI 提供一個開放的基礎架構，讓使用者部署和運行生成式 AI 工作。透過 SUSE AI，可以無縫接軌各種 LLM 模型和元件。並且可在多款程式中觀測即時威脅。

圖 9 World AI Week：SUSE AI 攤位

肆、研討議題二、AI 監管規範與實務因應做法

2024 年 7 月 12 日，歐盟正式公告全球首部全面性的 AI 法律《人工智慧法案》（AI Act，以下簡稱 AIA），公告 20 日後法律分階段生效，該法案旨在規範歐盟境內的 AI 在符合可靠安全性與以人為本的前提下之開發與應用，並以風險分級為此法案之最大特色。

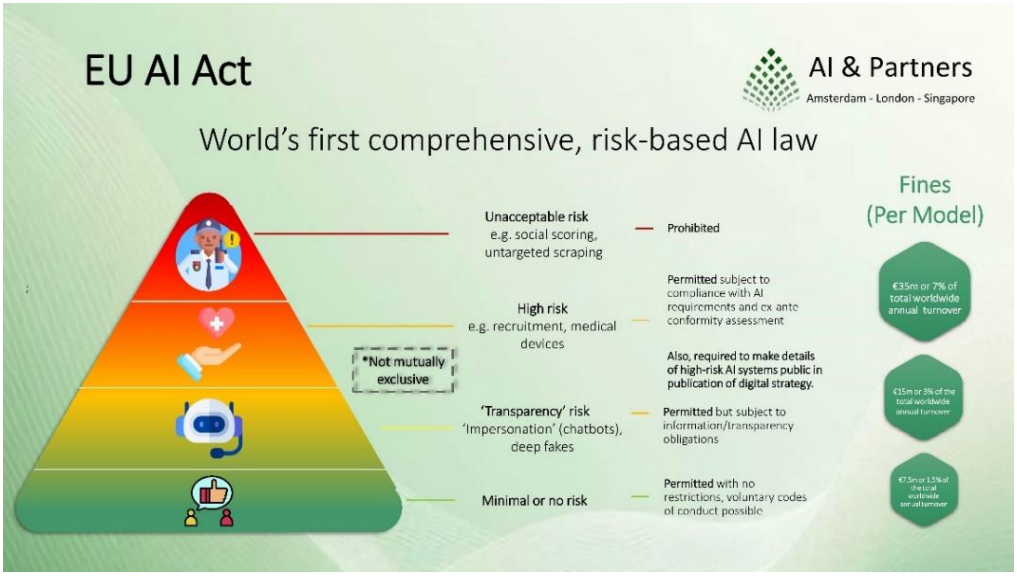


圖 10 歐盟 AI 法風險分級管理

AIA 訂定了更明確的法律規範細則，例如該法案要求 AI 開發公司在使用生物辨識、情緒辨識系統時通知使用者，並標記深偽技術（deep fake）與其他 AI 技術生成之內容等對細部 AI 種類使用進行規定。此外，保險與銀行公司也需提供評估報告，指出組織若使用 AI 工具，將對客戶權益造成哪些潛在侵害。

著作權與智慧財產相關法規要求方面，AIA 規定所有 AI 模型，無論是提供給 AI 作為訓練學習的資料，或是 AI 的產出結果，都必須要符合歐盟的智慧財產法規。對於非常強力的超級 AI 模型，例如 GPT-4 或 Gemini 等頂級 AI 模型，開發該些模型的公司必須主動針對該些模型的安全可靠性與其他重要資訊提交報告。

根據 AIA 規定，違反法定義務的代價相當高昂，違反法案的公司，包含 AI 開發者、利用者與經銷商等相關企業，若違反 AIA 規定，根據違規態樣、行為持續時間、嚴重程度與公司規模等因素，罰款最高可達公司年度總營業額之 7%，實際裁罰金額視以上要件情況調整。此外，法案成立獨立的專家小組來針對相關 AI 執法運作治理，以最大化 AIA 之執行效率。

AIA 目前已完成第一、二、三階段生效（如下表），剩下高風險 AI 及部分一般性規範將於 2026 年 8 月生效。

生效階段	日期	內容/適用主體
第一階段	2025.2.2	禁止「不可接受風險」的 AI 系統、通用定義生效
第二階段	2025.2.4	發布「禁用 AI 實踐指引」
	2025.2.6	發布「AI 系統定義指導方針」
	2025.7.10	發布「通用型 AI 行為準則」
第三階段	2025.8.2	通用型 AI 模型治理、指定主管機關、部分罰則與報備義務生效
第四階段	2026.8.2	高風險 AI 及各項一般性規範全面生效

表 2 歐盟 AIA 生效階段

歐盟執委會為落實相關規範，已陸續發布「禁用 AI 實踐指引」、「AI 系統定義指導方針」、「通用型 AI 行為準則」，2025 年 9 月 26 日亦發布「AI 重大事件通報指引」草案。為因應 AIA 的生效，歐盟企業無不密切關注所謂「負責任 AI」的具體法遵要求，以即時規劃因應做法。本次會議則出現許多廠商，將 AIA 的法遵議題視為市場商機而推出相應之服務。

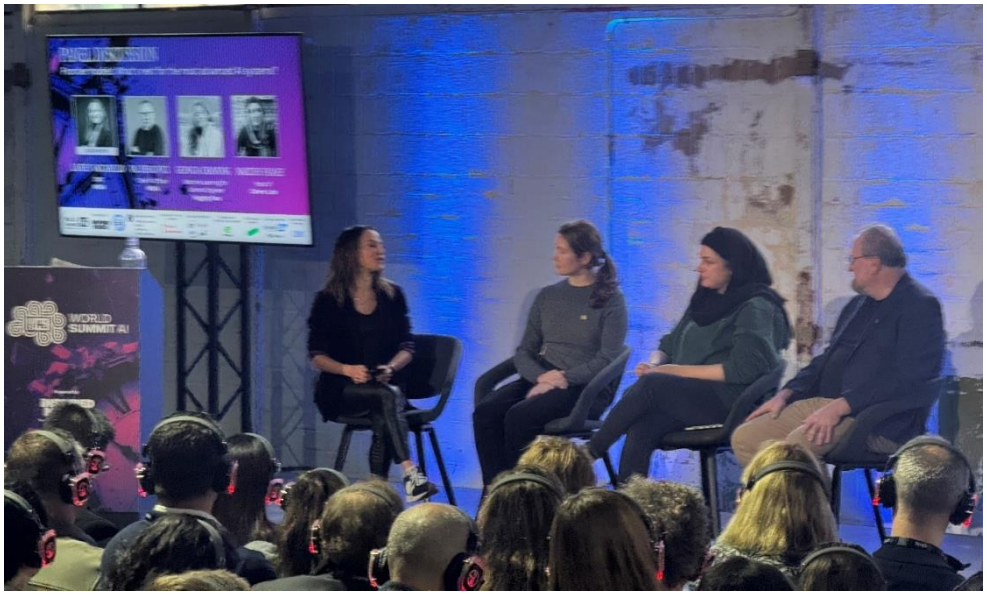


圖 11 World Summit AI：負責任 AI 專場

圖 12 歐盟 AI 法要求企業法遵

一、AI 透明度報告

（一） AI 掃描器

辨識軟體工具背後的大型語言模型、AI 模型、演算法及雲端運算平臺。管理開發者與使用者使用紀錄，以利後續透明度報告中揭露。

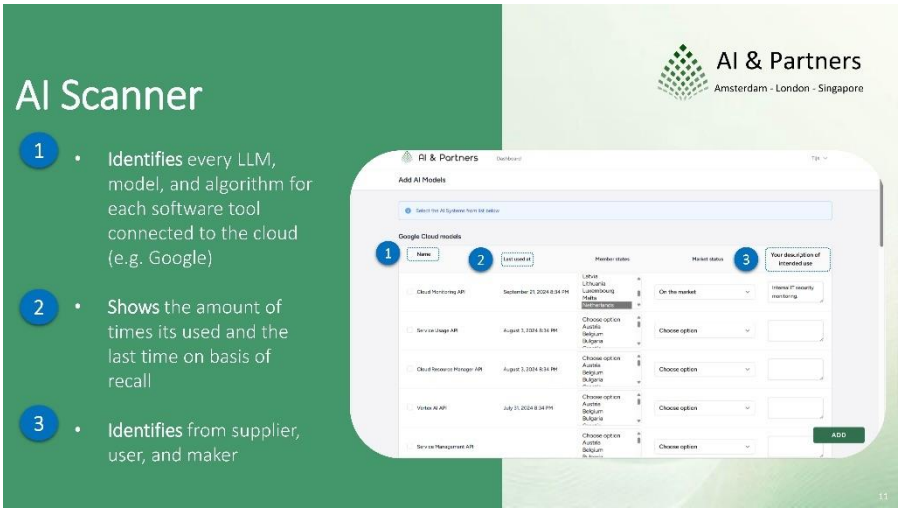


圖 13 AI Scanner

（二） 風險自我檢驗器

提供公司組織整體自我評估 AI 應用風險，幫助主管或主責單位辨識風險分級與對應之管理措施。

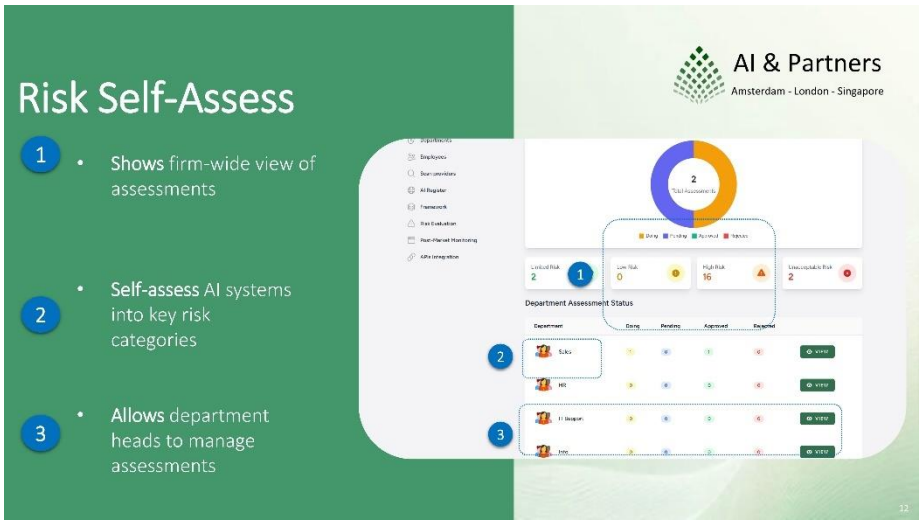


圖 14 Risk Self-Asses

（三） 高風險 AI 模型監控器

針對高風險 AI，即時監控模型異常狀態、未經授權之變更，以及提供視覺化即時狀態表供企業查看。

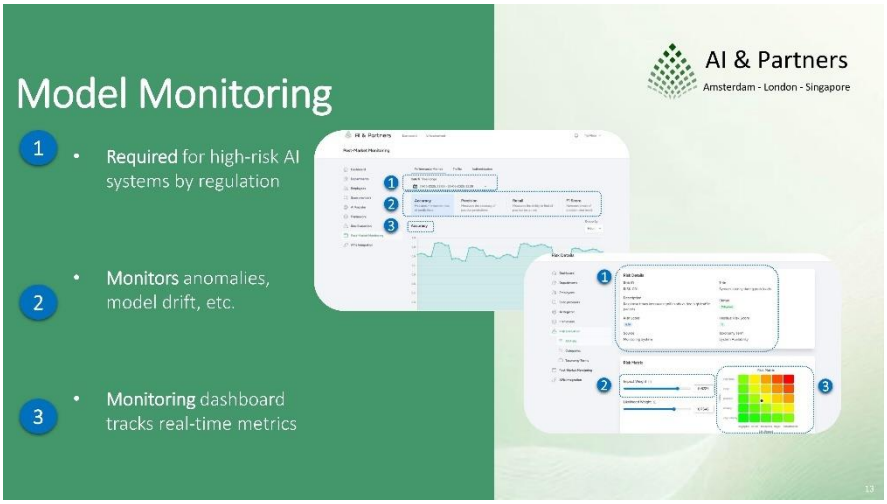
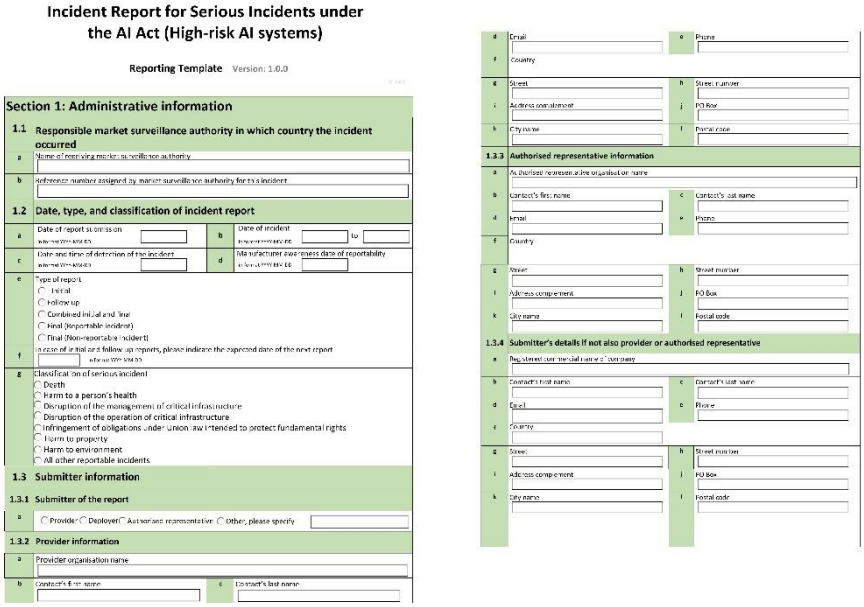


圖 15 Model Monitoring

二、AI 重大事件通報

依據 AIA 第 73 條，高風險 AI 提供者，必須遵循重大事件通報義務。此義務不但能提供市場監管機關，即早辨識通案性之重大影響事件，並能即時做出因應處置；此外，該規定讓高風險 AI 提供者有隨時追蹤、確保其 AI 系統安全性、穩定性的動力。因此，歐盟執委會發布「AI 重大事件通報指引」草案時，一併提供重大事件通報之範本表單如下表：



Incident Report for Serious Incidents under the AI Act (High-risk AI systems)

Reporting Template Version: 1.0.0

Section 1: Administrative information

1.1 Responsible market surveillance authority in which country the incident occurred

a Name of receiving market surveillance authority

b Reference number assigned by market surveillance authority for this incident

1.2 Date, type, and classification of incident report

a Date of report submission (dd/mm/yyyy hh:mm)

b Date of incident (dd/mm/yyyy hh:mm) to

c Date and time of detection of the incident (dd/mm/yyyy hh:mm)

d Manufacturer awareness date of reportability (dd/mm/yyyy hh:mm CET)

e Type of report
- ○ Initial
- ○ Follow up
- ○ Combined initial and final
- ○ Final (Reportable incident)
- ○ Final (Non-reportable incident)

f In case of initial and follow up reports, please indicate the expected date of the next report (dd/mm/yyyy hh:mm)

g Classification of serious incident
- ○ Death
- ○ Harm to a person's health
- ○ Disruption of the management of critical infrastructure
- ○ Disruption of the operation of critical infrastructure
- ○ Infringement of obligations under Union law intended to protect fundamental rights
- ○ Harm to property
- ○ Harm to environment
- ○ All other reportable incidents

1.3 Submitter information

1.3.1 Submitter of the report

a ○ Provider ○ Deployer ○ Authorised representative ○ Other, please specify

1.3.2 Provider information

a Provider organisation name

b Contact's first name

c Contact's last name

d Email

e Phone

f Country

g Street

h Street number

i Address complement

j PO Box

k City name

l Postal code

1.3.3 Authorised representative information

a Authorised representative organisation name

b Contact's first name

c Contact's last name

d Email

e Phone

f Country

g Street

h Street number

i Address complement

j PO Box

k City name

l Postal code

1.3.4 Submitter's details if not also provider or authorised representative

a Registered commercial name of company

b Contact's first name

c Contact's last name

d Email

e Phone

f Country

g Street

h Street number

i Address complement

j PO Box

k City name

l Postal code

表 3 AI 重大事件通報表

三、AI 應用著作權合規

歐盟對於著作權及生成式人工智慧（Generative AI）的法律規範體系，主要建立在 2019 年《數位單一市場著作權指令》（Directive 2019/790）、2024 年《人工智慧法》（AI Act）及《通用目的 AI 行為準則》（General-Purpose AI Code of Practice）之基礎上。

（一）《數位單一市場著作權指令》（Directive 2019/790）

根據《數位單一市場著作權指令》第 2 條第 2 款，文本與資料探勘（Text and data mining, TDM）被定義為「任何自動化的分析技術，旨在分析數位形式的文本和資料，以生成資訊，包括但不限於模式、趨勢和關聯」。這種技術對於生成式人工智慧的發展至關重要，因為它使得 AI 模型能夠從大量數據中學習和識別規律，進而生成新的內容。

指令設有兩種 TDM 例外規定，分別適用於不同目的和受益人：

1. 針對科學研究的狹義例外（第 3 條）：僅限於研究組織和文化遺產機構，僅限於科學研究目的允許合法取得的作品的重製和截取。此外，該條例禁止透過合約條款來排除此例外的適用（第 7 條）。
2. 附帶條件的廣義例外（第 4 條）：此例外不設限制，包括商業性 AI 開發者也能允許作品的重製和擷取。同樣必須是合法取得的內容。最關鍵的條件是，著作權人有權選擇「退出」（opt-out）。
3. 在廣義 TDM 例外下，著作權人可以適當的方式保留其 TDM 權利。例如對於公開可線上取得的內容，透過機器可讀取的方式來表示退出。為促成退出或授權訓練內容，業界正發展多種實踐方式，包括建立技術標準、電子曲目清單、使用條款/標準條款、退出曲目聲明、連結至退出儲存庫的技術解決方案、AI 公司提供的個別退出方案（如媒體管理器），以及包含退出選項的著作權基礎設施解決方案。

（二）《人工智慧法》（AI Act）及《通用目的 AI 行為準則》

（General-Purpose AI Code of Practice）

除了 DSM 指令外，歐盟的《人工智慧法案》（AI Act）也針對通用人工智慧（GPAI）模型提供者引入了具體的著作權相關義務。所有在歐盟市場上推出的通用人工智慧模型提供者，均須履行以下兩項核心義務：

1. 歐盟執委會 7 月發布之通用人工智慧實務守則（Code of Practice）旨在支持通用 AI 模型提供者，正確實施符合歐盟著作權法的政策。提供者可依該實務守則來證明其合規性。守則規定通用 AI 模型提供者，在爬取全球資訊網時，僅重製和擷取合法可取得的受著作權保護內容，識別並遵守權利保留（退出機制）。

2. **Monarch AI 授權資料庫**

為因應 AI 模型訓練能取得權利人合法授權，Monarch AI 推出授權資料庫，標榜透過其資料庫訓練 AI 能夠取得合法且高品質的訓練資料。

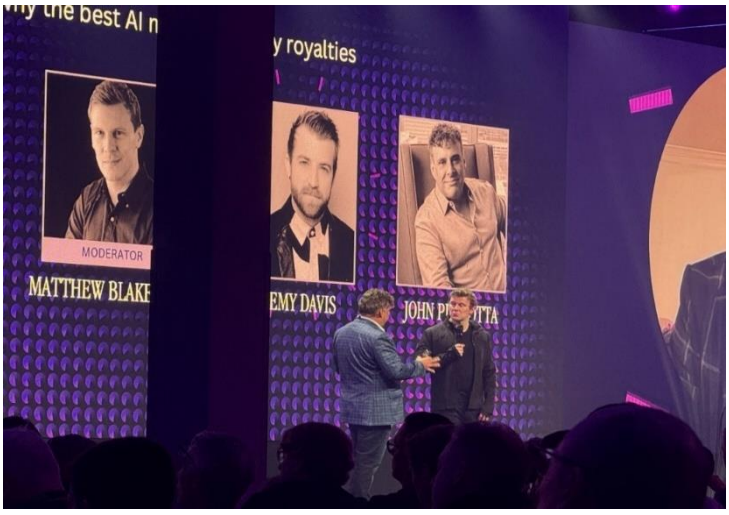


圖 16 World Summit AI ：AI 與著作權專場

四、AI 應用個資保護合規

鑑於 AIA 並未排除 GDPR 的適用，2024 年 12 月歐洲個資保護委員會（European Data Protection Board，下稱 EDPB）發布《第 28/2024 號關於人工智慧模型中處理個人資料保護意見》(Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models)，旨於尋求歐盟範圍內之監管協調如何將 AI 模型視為匿名、如何驗證資料控制者是否符合 GDPR 規定、開發之 AI 模型時使用非法處理之個人資料等議題。

2025 年 4 月，歐盟 EDPB 發布《人工智慧隱私風險與因應措施-大型語言模型》文件，提供大型語言系統(Large Language Models, LLM)之系統開發人員及使用者提供與管理相關技術相關隱私風險之指引及工具，說明相關關鍵技術之概念、風險管理流程、主要風險識別及評估措施，並提供案例，以協助開發者、部署者以及資料保護機關得識別、評估並降低 LLM 之隱私及資料保護風險。

（一） 2025 年 2 月法國資訊自由委員會發布與 AI 相關之個人資料利用指引

法國資訊自由委員會(Commission Nationale de l'Informatique et des Libertés, CNIL)於 2025 年 2 月 7 日發布兩項關於 AI 使用與個資保護之新建議，旨在促進 AI 領域之負責任創新並同時確保其能遵守歐盟《一般資料保護規則》(General Data Protection Regulation, GDPR)之規定。CNIL 指出，雖某些 AI 模型之訓練資料符合匿名化而不受 GDPR 之約束，但仍有其他大型 AI 模型內(如 LLM 等)可能包含個資，須遵守 GDPR 之規定。CNIL 建議機構落實隱私始於設計(privacy by design)，於

AI 系統開發階段即將隱私保護納入考量，並盡量以匿名化(anonymisation)或假名化(pseudonymisation)等替代措施補足現行 AI 系統無法完全讓當事人行使權利之限制。

（二） 法國 AI 公共服務專案的隱私始於設計

為了讓 AI 系統不受到 AIA 禁止生物辨識的限制，巴黎地鐵公司（RATP）開發 PRIV-IA 專案，透過對個人隱私影響最小的影像擷取辨識技術，偵測地鐵是否有乘客失足墜落軌道，連動列車自動煞停的緊急事故應變系統。在 AI 辨識影像時不辨識人臉，而以類似熱影像方式擷取人類動作如下圖，藉以達成匿名化資料的效果。

Comparaison d'images de la même scène issues d'une caméra classique (à gauche) et d'un capteur temps de vol (à droite) – Source : RATP

圖 17 巴黎地鐵公司 AI 公共服務專案

伍、研討議題三、負責任 AI 的法規最新趨勢

一、 義大利

義大利於 2025 年 9 月 17 日通過《人工智慧規範與政府授權》立法法案(Disposizioni e delega al Governo in materia di intelligenza artificiale，下稱 1146-B 法案)，為該國首次針對 AI 全面立法，亦為歐盟成員國內 AI 專法先驅。義大利將歐盟《人工智慧法》(AI Act，下稱 AIA) 框架轉化為國內法，並設立獨立窗口與歐盟對接 。為確保與歐盟溝通順暢，同時平衡國內 AI 分工治理平衡，預計採「雙主管機關制」，由隸屬於總理府 (Presidenza del Consiglio dei Ministri) 之數位局 (Agenzia per l’Italia Digitale，AgID) 及國家網路安全局(Agenzia per la Cybersicurezza Nazionale，ACN) 共同主導執行此法案。

（一） AgID 推動 AI 技術標準、互通性與資料治理，負責公共行政與數位服務之 AI 發展與監管，確保創新符合國家數位轉型戰略；ACN 保障 AI 系統之資安完整性與韌性，負責事故通報、應變機制與國安相關審查，確保高風險 AI 安全性。

（二） 目前該法案已由參議院 (Senato della Repubblica) 審議並表決通過，2025 年 9 月 25 日已載於義大利《官方公報》(Gazzetta Ufficiale)，再經過 15 天緩衝期後，預計於 2025 年 10 月 10 日正式生效。

二、 歐盟

（一） 通用人工智慧行為準則

講者認為通用型人工智慧行為準則第三版草案在保護基本權利、準確定義系統性風險、實施一致的風險管理標準以及確保吹哨人保護和公眾透明度方面存在

嚴重不足。其對「安全衍生模型」和「可接受的系統性風險」的概念引入了潛在的監管漏洞和與人工智慧法案及預防原則不符的風險。起草過程未能有效整合多元利益攸關者的意見，特別是公民社會和基本權利專家的觀點，進一步削弱了 COP 的公信力和有效性。為確保歐盟市場上通用型人工智慧的可靠性，COP 亟需進行重大修訂，以更好地維護歐盟價值觀和基本權利。

依據歐盟 AI 法、GDPR 及各國專業標準，建立多層次合規框架，明確規範數據使用、模型開發、部署與監管責任。鼓勵制定行業自律準則，如醫療、金融、交通等關鍵領域，補充公共法規之不足。認為在多邊合作、法律制度與企業自律三者兼具下，才能確保 AI 發展既促進創新，又維護公共安全與基本人權。

圖 18 World Summit AI ：Karen Hao, Empire of AI

（二） 部署對抗性測試（adversarial testing）與紅隊演練

為避免偏差、歧視或受操控之資訊進入訓練流程，並使用數據溯源技術追蹤來源。歐盟主張部署對抗性測試（adversarial testing）與紅隊演練，主動發現模型弱點與攻擊路徑，並實施持續監控與異常偵測，透過行為分析及模型審計，及早預警性能退化或遭受攻擊。

（三） 吹哨者保護指令

雖有些國際草案文件於保護範圍與程序上趨於保守，卻仍必須確保吹哨人在揭露 AI 開發或部署中危及公眾安全、違反倫理或法規行為時，免受報復。為此，歐盟訂有《吹哨者保護指令》，將 AI 領域納入明確保護範疇。企業應設立匿名舉報渠道及外部聯絡窗口，並將 AI 安全與倫理列為內部合規考核指標。

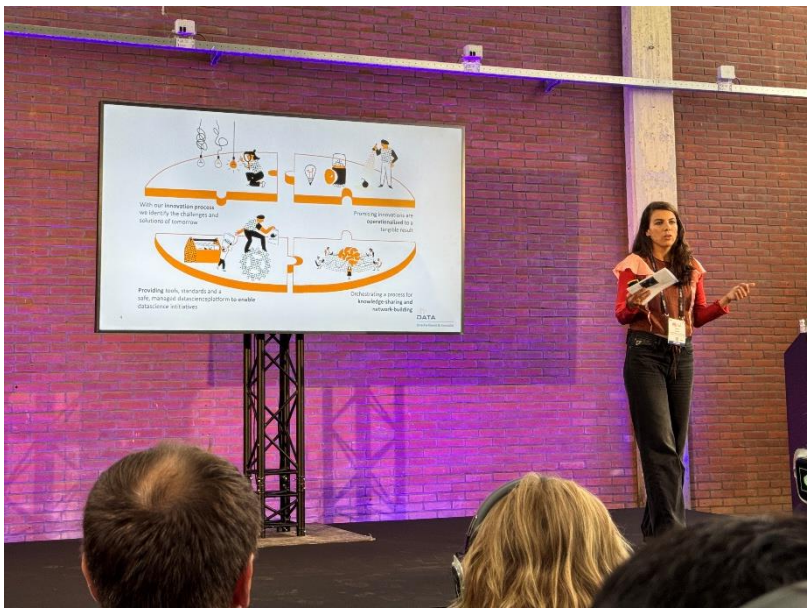


圖 19 World Summit AI：歐盟國防官員系統韌性說明

三、美國

（一）加州《前瞻 AI 模型透明度法案》

2025 年 9 月 29 日由加州州長簽署通過《前瞻 AI 模型透明度法案》，法案要求年營收在 5 億美元以上，或模型每秒浮點運算次數超過 10^{26} 次的大型 AI 模型，公開其安全框架，包含其安全協議，在部署、修改框架後 30 日內需提供透明度報告，如隱藏機敏資訊須載明理由。並且，如有重大安全事件，發生後 15 日內應通報檢察總長。

（二）反假冒法案

美國 2024 年 7 月 31 日首次提出《培育原創、促進藝術與維護娛樂安全法》；又稱《反假冒法案》，原文為 The Nurture Originals, Foster Art, and Keep Entertainment Safe (NO FAKES) Act。後於 2025 年 4 月 9 日提出新版本，目前刻正審議中。該草案將「數位擬真再現著作」定義為一種電腦系統生成、高度逼真的聲音或視覺呈現方式（例如 AI 生成擬真語音或臉部替換影像），即使該個人（如演員、歌手、配音員）從未創建該作品，但若該數位形式與某人聲音、形象高度相似，則該人可以主張權利，並及於生前與死後（70 年），包含可禁止未經授權之使用、散布等。未經授權使用該個人聲音或肖像來創造或散布 AI 生成之複製、再現版本，則該人或其代理人、繼承人可針對侵權行為提起民事訴訟，主張損害賠償或其他救濟。

此外，法案針對「線上服務（online service）」範疇做廣泛定義，讓法案適用於社群平臺(如 FB、IG)、影

音串流平臺(如 YouTube、Spotify)、App 線上商店等數位平臺。平臺在接獲其上架著作、商品等，已被標記為「未經授權數位擬真再現」通知後，應迅速移除被標記內容，以避免承擔責任。平臺可在合理時間/緩衝期內展開下架行動，不會因為第三方上傳違法內容而自動承擔責任，即安全港(safe harbor)條款。草案目前獲得音樂創作者的支持，然有亦有批評認為雖然再現著作若授權條款太寬鬆，仍無法真正保障權利人著作與形象，目前也對權利人「本人知情同意」沒有硬性規定。

四、丹麥

丹麥 2025 年 6 月 25 日提出《丹麥著作權法深偽規範修正案》（Danish Copyright Amendment for Deepfake Regulation），此草案為著作權法（Ophavsretsloven）之修正案，目的在於保障每個人「對自己身體與聲音的控制權」（ret til egen krop og egen stemme），防止 AI 或 deepfake 技術未經同意生成、發布逼真的數位模擬影像與聲音。禁止未經同意公開或散布「高度擬真之個人特徵數位模擬」（如臉部、聲音、體貌等）；並禁止未經同意生成或發布表演藝術者演出的「逼真數位再現」，而低擬真或明顯虛構者不在此範圍內。草案保留對諷刺（satire）、戲仿（parody）、誇張藝術（caricature）等表達之合法例外。

若 AI 模擬生成內容屬新聞、評論或藝術表達，且不具誤導或損害他人權益之目的，則不構成違法。若違法發布或散布 AI 生成內容須負民事責任，受害當事人可請求移除不當內容及損害賠償（compensation），此法案目前尚未規定 AI 刑事罪名範疇。且線上平臺收到移

除請求通知後，有義務在合理期間內下架違法 AI 模擬內容。此丹麥草案目標為建立歐洲首批針對「人格特徵 AI 模擬」之保護制度，主要強調「人格權」與「表演權」兩者被 AI 模擬之延伸，補足現行著作權法未涵蓋 deepfake 侵害的漏洞。丹麥文化部表示未來將再推動歐盟層級之類似法案，將對 AI 模仿人格、表演之保護提升至國際層級。

圖 20 AI & Partners 座談：以人為本的 AI

伍、研討議題四、AI 生成內容標示與識別

國際間針對 AI 生成內容的標示與識別管制，正快速發展，如歐盟與法國都已落實或推動相關法規，強調內容透明、消費者保護與技術可追溯性。

（一） 歐盟

歐盟《AI 法案》（AI Act）將於 2025 年 8 月 2 日生效，明文要求所有「重要成分由 AI 生成或修改」的內容（包括文字、影像、音訊、影片、深偽等）必須清楚標示。標示方式分為「機器可讀」（如數位浮水印、元資料）及「人類可見」（如明顯提示、標語）。當 AI 涉入編輯或製作內容是否需標註，以「AI 貢獻是否具誤導性或本質改變」為準，純人為編輯仍可免標示。若有編輯審查（editorial oversight），可免除標示義務，但應留有資料佐證處理過程，落實責任歸屬。

（二） 法國

法國正在推動社群平台 AI 生成圖像、影片強制標示草案（Bill No. 675），使用者與平台須對 AI 生成內容負起審查與標記責任，違者將面臨罰款。

綜上，如歐盟與法國都以法律強制 AI 生成內容兩種標示方式:「人類可見、機器可讀」，並將平台及創作者納入管制對象，兼顧技術透明與用戶權益，顯示全球法規已將 AI 內容溯源與識別列為高度重視的新常規。

圖 21 AI & Partners 生成式 AI 治理

陸、研討議題五、代理 AI 與自動化勞動力

生成式與自治型 AI 代理人（AI Agents）正重塑企業生產力、工作流程與客戶服務，並推動大規模「AI 工作者」部署。大規模「AI 工作者」的實踐指導必須整合法規、組織流程、及強化員工參與與技能轉型，方能實現 AI 與人力協同共存之願景。歐盟以《AI 法》（AI Act）、《平台工作指令》（Platform Work Directive）及歐洲議會關於工作場所 AI 的草案報告為核心，結合法規解釋與政策建議，旨在平衡創新與勞動者權益。

一、AI Agents

（一） 打造 AI Agents

近年生成式 AI 與自主 AI 代理（AI Agents）技術成

熟，可支援公文撰擬、知識檢索、民眾諮詢與決策輔助。AI Agent 可想像為製造一個幫助達成任務的 AI 員工。具備主動任務執行與語義推理能力之智能代理，能依開發者授權的任務自動檢索資料、執行流程與提供建議。

打造 AI Agent 可使用 Python 等程式語言和專業 AI 函式庫來編寫所有元件（如推理引擎、感測器、執行器），也可以利用開源或商業模型（例如 LangChain、Semantic Kernel），提供用於提示管理、記憶體、工具選擇和協調的預建模組。企業導入 AI Agent，通常會遵循企業內部工作流程來建構、設計代理工作流程，將複雜問題分解為較小的任務，然後自主處理每個步驟。另外，也能夠集合多個 Agents 的代理系統，處理專案的不同面向，將結果傳遞給負責協調的 Agent 進行彙整和監督。

圖 22 World Summit AI：代理 AI 與自動化勞動力專場

（二） AI21 Maestro

各大供應商都在大力宣傳「AI Agents」系統，據稱

該系統能夠以最少的人工投入處理複雜任務。然而，在實際應用中，企業對 AI 代理的真正應用卻停滯不前。事實上，Gartner 預測，到 2027 年，超過 40%的 AI 代理專案將被取消，原因是成本飆升和價值不明確。「代理」一詞本身就暗示了這項工作本質上是之是工作的代理。然而，我們所見的大多數 AI Agents 仍然僅能檢索事實，但還不能執行需要使用即時專有企業知識進行判斷、綜合和規劃的工作流程。如果 AI Agents 要成為企業營運的核心，它們必須專為企業級的知識工作而建構。這意味著它們不僅要回答常見問題或查詢資料庫，還要處理高風險、多步驟的任務，例如產生併購盡職調查摘要、策略供應商評估報告或合規風險報告。

AI21Labs 是 2017 年由 Mobileye 創辦人 Amnon Shashua、史丹佛教授 Yoav Shoham、CrowdX 創辦人 Ori Goshen 所設立。2025 年推動 AI21 Maestro，用於建構和部署知識代理的系統，專門用於跨公司特定資料執行多步驟推理和綜合的人工智慧系統。它並非像典型的代理那樣猜測下一步，而是系統地規劃和執行複雜的工作流程：選擇正確的檢索策略、驗證中間結果並產生結構化、準確的輸出。

Maestro 動態規劃每個任務，平衡成本和延遲等權衡，並在工作時自我糾正，產生清晰、透明的輸出，且 Maestro 與簡單的 RAG 流程不同，其執行知識工作的整個過程為：分解目標，確定所需信息，檢索信息，解讀信息，並產生連貫且有價值的結果。無論是盡職調查備忘錄、合規摘要，還是多頁研究文件均可完成任務。

圖 23 World AI Week：AI21 Lab 攤位

二、自動化勞動力

（一） 歐盟 AI 法依風險將 AI 系統應用分為「禁用」、「高風險」、「有限風險」及「低風險」，其中客戶服務與企業流程中常見的 AI 代理人與聊天機器人多屬「有限風險」，須履行透明度告知義務，使用者須知其與 AI 互動。高度自動化的員工監控、績效評估與人力資源決策系統則屬「高風險」，須進行風險評估、建立品質管理體系並接受第三方合規檢驗。

（二） 平台工作指令（the Platform Work Directive,PWD）第 III 章要求數位勞務平台對「自動化監控與決策系統」以明確易懂語言揭露算法機制與決策依據。另，應建立人為監督與人工覆審重大決策之機制；再者，應確保工會或代表能取得風險評估報告與影響評估報告。

（三） 2025 年 6 月，歐洲議會就業與社會事務委員會進一步提

出草案報告，呼籲在傳統就業領域修補《AI 法》與平台指令之規範空缺。特別針對 AI 做出招聘、晉升或解僱等決策時，呼籲應提供員工申訴與覆議機制；針對演算法影響就業評估與事後補救機制，要求企業提交合規報告並接受監管機構審查。

柒、研討議題六、數位主權與地緣政治

技術主權（technological sovereignty）指國家或區域在關鍵技術領域擁有獨立研發、生產與管理。近年來，隨著生成式 AI 及大型語言模型（Large Language Models, LLM）對社會經濟與軍民應用的深遠影響，全球主要經濟體皆將其視為關鍵技術主權的核心。目前各國發展模式呈現「公私協力」與「開源與封閉」並存：一方面透過國家基金與基礎設施投入，降低私企研發成本；另一方面推動開放模型以凝聚區域優勢。

一、 歐盟：雖無單一歐盟級 LLM，但透過「Horizon Europe」與「數位歐洲計畫」資助法國 Bloom、德國 Aleph Alpha、歐洲 AI 聯盟開發開源模型，並倡導模型透明度與倫理準則。

二、 美國：OpenAI、Anthropic、Google DeepMind 等私部門領先，並獲得政府與國防部資助推動通用 AI 計畫。

三、 印度與日本：分別由政府主導與企業合作開發針對多語言和特定行業應用的 LLM，以支撐數位經濟與服務出口。

圖 24 World Summit AI：主權 AI 專場

捌、心得與建議

一、 促進 AI 創新應用，推升我國整體 AI 發展

我國現行 AI 發展政策以多元布局架構，透過跨部會協力共同推動，行政院於整體「智慧國家方案（2017-2025）」框架下，跨部會推動重要數位政策，其中包含「臺灣 AI 行動計畫 2.0（2023-2026）」與 AI 新十大建設等，目標將百工百業導入 AI，提升效率、產量並創造新市場，

建議未來我國可參考國際作法，評估搭配適當政策工具，如依各產業需求建立輔導或投資 AI 計畫，以及推動與整合各部會 AI 功能，融合至全國政府機關業務推動，加速效能，並建立內部 AI 使用原則等。透過各部會整體在技術、應用、治理、人才培育、

國際合作、基礎環境整備等面向，共同推動我國 AI 發展。

二、 觀察本次研討活動各國作法，建議我國評估推動可協助企業、使用者落實法遵之指引與工具

我國 AI 基本法草案為我國人工智慧發展與治理奠定基礎原則與方向。草案第 9 條授權數位發展部推動 AI 風險分類框架，並要求各目的事業主管機關依此框架訂定「以風險為基礎之層級管理規範」。第 11 條則進一步針對「高風險人工智慧之應用」，課予政府特定義務，包括責任歸屬、建立救濟及補償機制等，因此，後續可由各目的事業主管機關視需求訂定相關特定義務。

隨者 AI 推廣與使用，不同領域之應用亦可能帶來相關的風險，建議未來相關特定領域主管機關如須訂定 AI 應用規範時，可參考歐盟等國際作法，在硬性規範前，提供指引、工具與宣導措施，並強化 AI 生成與輸出結果應可追溯，以防止假訊息傳播，並使企業能夠在安全 AI 應用環境與不大幅提升成本下，遵循目的事業主管機關之規範，以利發展負責任的人工智慧。

三、 建議借鏡歐盟做法，逐步完善數位與資料主權，發展本土 AI 核心技術，滾動檢討我國配套政策與法規發展環境

隨者 AI 模型應用與技術變化，相關訓練 AI 資料、資料主權、AI 創新技術與 data center 基礎設施等亦日漸重要，建議可評估將政府擁有資料搭配 AI data center，強化數位主權，並透過適當調適各部會相關法規與政策工具，建立發展本土核心 AI 技術，完善 AI 運作環境，使 AI 基本法草案落實 AI 應用的各行各業，並成為優先落實 AI 發展環境之標竿國家。

現行行政院已成立數位政策法制協調專案會議，可參考與借鏡國際 AI 法制趨勢，逐步引導各部會調適政策與法制環境，促進 AI 技術的健康發展，進一步提升臺灣在 AI 領域的競爭力。