



規範ChatGPT ——通用人工智慧與歐盟 人工智慧法草案

周靖媛*

理慈國際科技法律
事務所律師

目次	壹、通用人工智慧的發展與重要性 貳、通用人工智慧的定義	參、通用人工智慧與歐盟人工智慧法草案 肆、結語
----	--------------------------------	----------------------------

人工智慧（Artificial Intelligence，AI）治理不僅是世界各國關注的焦點，亦為我國目前迫切之立法議題，本文將介紹通用人工智慧（General-Purpose AI，GPAI，詳後述）與其他AI系統的差異，以論兩者應作差別規範，並說明GPAIs與目前歐盟發展中之「人工智慧法」(Artificial Intelligence Act，AIA)¹草案風險分級制度不協調之處，以期作為相關法規制定與探討的研究基礎。

壹、通用人工智慧的發展與重要性

眾所周知，AI的發展近期在全球掀起了滔天巨浪，對各行各業以及人類生活的各個層面都造成了巨大的衝擊與變革，更為監理帶來了相當大的挑戰。基於提高效率與有效性的技術進展，AI的正向發展預期能提升人類社會整體的經濟與福利，使人類有望應對氣候變遷²與醫療健康等全球挑戰，並在微觀及宏觀

的層面改善公營和私營組織³。尤其是ChatGPT此等GPAIs的現世與普及，更大幅加速了普遍想像中人類未來世界的來臨。GPAIs簡言之是具有廣泛可能用途的AI系統（包括開發人員有意和無意的用途），可在無需進行實質性的調修的情況下被應用於各領域許多不同的任務（詳後述）。

一、技術的發展

作為GPAIs狂潮的濫觴，2020年研究實驗室OpenAI⁴發布了GPT-3，為一種在大型網路資料庫上訓練的語言模型，能夠執行廣泛的自然語言處理任務；2021年，Open AI發布了DALL-E，為一種深度學習模型，可以從自然語言描述中生成數位圖像；2022年12月，OpenAI推出了基於GPT-3的聊天機器人ChatGPT，使用網路資料以機器學習模型進行訓練而生成各類文本，雖然與真實人類的表現仍有差距，但其表現出優異的指令識別能力與擬真的創造力，正式造成全球轟動，在發布後的前幾個月便擁有1億活躍用戶，被描述為「有史以來增長最快的消費者應用程式」⁵；而後，升級版GPT-4於2023年3月推出，其性能與表現皆較GPT-3更加細緻與精準⁶。2023年2月，Microsoft推出Bing搜尋引擎和Edge瀏覽器的整合⁷，結合聊天機器人使搜尋功能具有更完整的上下文，顛覆過往線上搜尋的使用模式與流程。Google及其子公

司DeepMind亦正在開發GPAI工具，例如對話式AI服務Bard。GPAIs開發人員正以此類基礎模型提供服務⁸，同時，其他開發人員亦透過應用程式介面（Application Programming Interface，API）等方式建構並佈署基於此類基礎模型的應用程式，通常稱為生成式人工智慧（Generative AI）工具。如今，全球至少數百萬用戶仰賴這些工具生成人類智能級別的文本（例如GPT-4、ChatGPT、Bard、Bing）、圖像（例如Stable Diffusion、DALL-E 2、Midjourney）、視訊（例如Synthesia、D-ID）或音訊（例如MusicLM、AIVA）等，並有更多的替代方案及整合應用程式正在持續開發與提供⁹。

GPAIs的顛覆性質具有從根本上取代現有技術、產品或服務的潛力¹⁰，雖可預期刺激各個產業的創新，為人類社會帶來巨大的利益，然其隨之而來引發了有關隱私權與個人資料保護¹¹、誤導或虛假資訊的放大¹²、偏見歧視與仇恨言論¹³、勞動替代的隱憂、使用倫理、獲得有效司法救濟與公正審判的權利、消費者保護與智慧財產權的風險¹⁴、反競爭¹⁵、人類過度依賴¹⁶、網路安全¹⁷以及允許GPAIs僅掌握在少數公司和政府手中的危險¹⁸等問題，其中傳播虛假和錯誤資訊的可能性更使輿論及心理操縱、假新聞及有害言論的風險上升到一個全新的高度¹⁹，一些GPAIs已因傳播極端主義

內容²⁰、表現出反穆斯林偏見²¹、無意中洩露個人資料²²、使用GPT-3搭建之聊天機器人指導用戶自殺等而引起恐慌^{23、24}。雖然AI技術長期皆在持續發展，但由GPAIs引發之風險與恐慌遠超以往，2023年5月數百名AI科學家和研究人員連署警告該技術對人類構成滅絕危險²⁵，在2023年6月上旬Yale之CEO峰會上，超過40%的商界領袖（包括Walmart執行長Doug McMillion和Coca-Cola執行長James Quincy）表示，AI有可能在未來5到10年內毀滅人類²⁶。

除了機器對人類滅絕式的超大型安全性風險，GPAIs對於人類社會各層面皆具有細微且深遠的影響，相關（潛在）侵害法益事件亦層出不窮，皆一再顯示了對於整體策略規劃之監理的迫切需求。

二、世界強權的角力中心

事實上早在chatGPT走紅之前，AI的治理與倫理問題已是各國探討的主流。由於缺乏明確規範AI之國際法，大型國際組織所採取之相關行動在相關領域便顯得非常重要²⁷。經濟合作暨發展組織（Organisation for Economic Co-operation and Development, OECD）之「人工智慧原則（AI Principles）」構成了G20之非約束性指導原則的基礎，包含「永續成長與福祉、人類價值與公平、透明度與可解釋性、穩健性與安全、問責性」²⁸；

聯合國及其組織亦持續展開相關活動，其中較為重要者，2021年11月，聯合國教育科學文化組織（United Nations Education Scientific and Cultural Organization, UNESCO）大會的193個成員國通過了《人工智慧倫理建議書（Recommendation on the Ethics of Artificial Intelligence）》，為AI倫理主題的第一份全球標準制定文書²⁹；2020年6月，包括歐盟、英國、法國、德國、義大利、斯洛維尼亞、美國、加拿大、墨西哥、印度、日本、韓國、新加坡、澳洲、紐西蘭等國，成立了人工智慧全球夥伴聯盟（Global Partnership on Artificial Intelligence, GPAI），為一項國際和多方利益相關者倡議，旨在指導以人權、包容性、多樣性、創新和經濟增長為基礎，負責任地開發和使用AI³⁰。

目前歐盟尚在立法程序中的AIA草案在2021年提出，並隨著2022年ChatGPT的發布加速了立法進程。AIA是全球第一個概括性、對非特定商品或特定領域，針對AI技術之現象與風險所為之立法嘗試，採取以風險為基礎的分級管制方法（risk-based approach）³¹。正如歐洲議會主席Roberta Metsola所稱，「AIA是世界第一套立法，無疑將為AI設定全球標準」³²，其發展為全球眾所矚目，惟其目前的結構是否能適當地因應GPAIs的技術與商業模式，仍有相當大的疑義。

AI在國際關係與地緣政治上具有相

當重要的戰略性意義，以美國、中國與歐盟為首，各國皆在爭奪主導地位³³。「得AI者，得天下」³⁴，AI大戰中各方科技巨頭早已磨刀霍霍，而法規也成為這場全球競賽中至關重要的一環³⁵。正如其他與科技高度相關之法規對於各國經濟與政治實力之發展將有深遠的影響，AI之相關管制規範自亦成為世界強權的角力中心。近日，有ChatGPT之父之稱之OpenAI執行長奧特曼（Sam Altman）在美國國會聽證會上呼籲政府監管AI科技潛在的危險性³⁶，卻也表示OpenAI可能因無法完全遵守AIA的規範而停止在歐洲的營運³⁷。關於應如何監理GPAIs的辯論愈演愈烈³⁸，可以肯定的是，在立法與監理層面上必須認清GPAIs與其他AI的不同，進而調整作出差異的規範。尤其是依照目前AIA採取的風險分級管制模式，AI系統的「預期目的」為其風險分析時重要的判別標準，若GPAIs未得到適當的區分與定位，將可能導致其責任脫逸或是落入無法或不適於承擔者的風險³⁹，進一步造成經濟及政治上的劣勢。

貳、通用人工智慧的定義

認識GPAIs的特徵及其特殊風險對於監理議題的探討具有重要意義。事實上，定義AI本身即充滿挑戰，這也展現了監理技術的困難，不僅是因為各學科領域對於AI的定義有所差異，AI與不透

明、複雜、可能具有偏見和快速變化的特性相關，與法律須確定性、透明度、可解釋性和平等待遇的要求不能協調地相互作用。技術正以非常快的速度發展，然而民主式的立法建立在辯論和妥協的基礎上，需要進行長時間的談判⁴⁰。AIA的主要起草人之一Axel Voss便指出「AI在兩年前仍未如此先進，而未來兩年則可能會進一步發展，『如此之快』以至於其中大部分法律實際生效時將不再適用」⁴¹。為了避免此種衝突，目標在面向未來的法案，可能採取以廣泛且幾乎與技術細節無關的用語定義AI，以保留更新的彈性；比起予以一個可以涵蓋包羅萬象或是關注技術細節因而可能隨著科技發展而過廣或過狹的定義，更重要的是：理解各種演算法和AI應用程式的不同特徵與其在實踐中的使用方式⁴²。

一、狹隘人工智慧與通用人工智慧

雖然AI沒有全球公認的定義，但科學家們普遍認為，從技術的角度而言，AI技術可分為兩大類⁴³：（一）「狹隘人工智慧」（Artificial Narrow Intelligence，ANI），也稱為弱人工智慧（weak AI）；（二）「通用人工智慧」（Artificial General Intelligence，AGI）⁴⁴，又稱強人工智慧（strong AI）。

ANI技術在標記良好之資料庫上進行訓練，以執行特定任務並在預定義的

環境中運行，例如圖像和語音識別系統；AGI技術是被設計用於執行各種智能任務、抽象思考並適應新情況的機器⁴⁵。近年快節奏的技術突破，包括大型語言模型（Large Language Model，LLM）技術的採用，已從根本上改變了AGI技術的潛力⁴⁶。新一波具有生成能力⁴⁷的AGI技術，稱為「通用人工智慧」（General Purpose AI）⁴⁸或「基礎模型」（foundation model）⁴⁹，部分文獻中亦稱「大型語言模型」（LLM）⁵⁰或「大型生成模型」（Large Generative Model，LGAIM）⁵¹，本文統一以GPAI稱之。

GPAIs接受廣泛之未標記資料之訓練，可用於不同的任務，僅需進行少部分的微調，下游開發人員便可透過API和開源途徑訪問這些底層模型，如今已被許多公司用作基礎設施⁵²而為最終用戶提供下游服務⁵³。GPAIs使用數十億甚至數千億個參數進行訓練⁵⁴，使其有別於僅旨在進行預測或分類或實現其他特定功能之其他AI模型，其應用範圍的擴大是訓練GPAIs需要大量資料和計算能力的原因之一，以實現「自行在資料中找到模式和關係，而無需明確地告知所要尋找的目標；一旦模型學習了這些模式，便可生成與訓練資料相似的新示例」⁵⁵，故能以有限資源做出決策並能自給自足地適應新的環境以及人類複雜的情境化需求⁵⁶，並將人類的創造力、靈巧性和獨創性等才能體現於機器系統

上⁵⁷。例如，一個用於語言處理的GPAI系統可為數百個應用模型（如聊天機器人、決策助手、垃圾郵件機器人、翻譯等）之基礎，還能進一步細化調整成為多種客戶量身定制之應用程式⁵⁸。

二、以任務為核心的通用人工智慧定義

歐盟官方日前就GPAI釋出之定義為「『能執行普遍適用的功能，例如圖像／語音識別、音訊／視訊生成、模式檢測、問答、翻譯等』⁵⁹和『可在多個環境中使用，並集成到多個其他AI系統中』⁶⁰之AI系統」，引發定義過於廣泛的批評⁶¹。

本文側重以監理為目的之AI分類及定義方式，除了常見的以固定目的（fixed-purpose）或廣泛目的（general-purpose）的角度觀察分類，尚可進一步從AI系統的功能（ability）、領域（domain）、任務（task）和輸出（output）四個面向⁶²進行關注⁶³。本文認為，以任務為核心，同時強調能力（capability）與通用性（generality）的建議似乎更為可採，以下說明之。

首先，能力係指向系統的勝任力、準確性或有效性；通用性則表示能力分佈的廣泛和均勻程度⁶⁴。在此認識下，以下進一步分就功能、領域、任務和輸出進行說明。

(一) 功能

功能通常可透過目的之普遍性來表現，亦可被描述為可影響的動作或過程之類別，例如語言、視覺、交互、理解、推理和搜尋⁶⁵。部分系統僅具有一種核心能力（如圖像識別）；部分系統則包括功能的組合，例如文本處理加圖像生成，具功能組合之類型在本維度上通用性更高⁶⁶。

(二) 領域

領域係指技術提供解決方案的經濟部門（如醫療、教育、國防等），代表了系統滿足特定部門利益相關者需求的能力。就此維度而言，服務於多個領域的AI系統將被認為比服務於一個或部分領域的AI系統通用性更高⁶⁷。

(三) 任務

對於不同類型問題之說明、目標的定義或系統必須實現的操作稱為任務⁶⁸。注意區分任務與功能的不同，例如，大多數LLM都具有處理和生成文本的功能，但處理文本並不是一個明確指定的任務；可能的任務將包括總結文本、標記資料、回答問題類別等。在此維度上，通用性更高的AI系統將可執行更廣泛的不同任務⁶⁹。

(四) 輸出

輸出是AI系統應用過程中的最後一部分，用於識別任務的結果⁷⁰。例如，從同一提示中產生出的獨特圖像或文章。在此維度上，輸出的多樣性將可用以衡量通用性⁷¹。

(五) 以任務為核心的通用人工智慧定義

定義的提出者認為，就本質上而言，任務是AI系統實現目標的基石。開發人員可以創建任務以專注於一種功能和領域，或滲透到多種功能和領域；並且無論輸出多麼獨特，其任務仍是一個可以始終用於衡量和比較不同系統活動的因素。其將AI分為兩種，一種為固定用途系統，被創造出來僅有一個特定的目標，並僅能完成所訓練執行的任務；另一種為GPAI，可以執行最初未受過訓練的任務。依此，將GPAI定義為「一個可以完成或適應完成一系列不同的任務，包括一些未經有意和專門訓練的任務之AI系統」（An AI system that can accomplish or be adapted to accomplish a range of distinct tasks, including some for which it was not intentionally and specifically trained）⁷²。

在此定義下，GPAIs可以是單模態或多模態，可包含範圍不同的自主或代理系統，可使用不同方法進行訓練，還可包含在模型參數計數或訓練計算的閾值水準上表現出新的和令人驚訝的功能之

系統⁷³。

雖然此種定義屬於質性，但原則上還可透過任務績效指標使其量化⁷⁴。

其他文獻亦有持類似的見解，即GPAIs必須在功能、任務或輸出等方面表現出顯著的通用性，而非僅是可被集成到各種用例當中⁷⁵。

參、通用人工智慧與歐盟人工智慧法草案

最早AIA草案版本並未特別提出GPAIs，歐盟理事會於2022年底將GPAIs的考量納入AIA草案，歐洲議會於2023年6月14日通過關於AIA的談判立場，隨後將與歐盟成員國就該法律的最終形式進行會談⁷⁶。

一、通用人工智慧於歐盟人工智慧法草案之最新發展

AIA草案結構係透過評估AI系統提供者之「預期目的」以認定風險級別，進而確定哪些系統須進行高風險類別之高度審查，或是禁止⁷⁷。歐洲議會仍遵循基於風險的方法，其「希望在AIA中納入一種區分層級的方法來規範GPAIs。歐洲議會希望對基礎模型的提供者施加義務，以確保有力地保護基本權利、健康、安全、環境、民主和法治。此外，使用LLM生成藝術、音樂和其他內容的生成基礎AI模型（例如ChatGPT）將受到嚴

格的透明度義務約束。最後，所有基礎模型都應為下游供應商提供所有必要的資訊，以便其能遵守AIA規定的義務⁷⁸，「評估和降低風險，遵守設計、資訊和環境要求，並在歐盟資料庫中註冊。……必須遵守額外的透明度要求，例如揭露內容是由AI所生成，設計模型以防止其生成非法內容，以及發布用於訓練之受版權保護資料之摘要。⁷⁹」要求「基礎模型的提供者—AI領域之一個新的和快速的發展—必須評估和減輕可能的風險（對健康、安全、基本權利、環境、民主和法治），並在歐盟市場上發布之前於歐盟資料庫註冊。基於此類模型之生成式AI系統，如ChatGPT，須遵守透明度要求（揭露內容係由AI所生成的，有助於區分深偽圖像與真實圖像）並確保防止生成非法內容。用於培訓之受著作權保護資料之詳細摘要也必須公開。」⁸⁰

二、風險分級之規範模式與通用人工智慧

歐洲議會對於GPAI之分類與定義方法似與本文參考多篇文獻及建議似有些微不同，其將基礎模型與生成式AI系統分層（ChatGPT屬後者）規範之方法或許更為細緻。然有文獻指出，任何根據預期目的將AI系統歸類為高風險或非高風險的方法，都會造成關於GPAIs的漏洞⁸¹。蓋GPAIs不存在預期的目的和採用

規模，因而至少在三個重要方面挑戰了當前AIA的方法：（一）將GPAIs分類為高/無高風險類別的可行性；（二）未來風險的不可預測性；（三）對私人風險安排（通常可能透過契約）的擔憂⁸²。

GPAIs難以特定出預期的使用目的，或是要進行此種預期之成本過於高昂以至於幾乎不可能達成。除了GPAIs的複雜性以及經常表現出令其開發人員驚訝的行為⁸³，決定如何使用GPAIs者為其用戶或最終用戶而非其提供者，因此，除非能在技術上可驗證地排除任何濫用，否則GPAIs通常將被視為高風險系統⁸⁴，必須遵守高風險義務，特別是建立全面的風險管理體系，考慮到GPAIs的多功能性，這將迫使提供者識別和分析所有已知和可預見之健康、安全和基本權風險，以及涉及GPAIs所有可能的高風險使用，並在此基礎上制訂和實施針對所有這些風險的緩解策略，而且還須對所有可能的高風險用途進行性能、穩健性和網絡安全測試，成本之高幾乎不可行，蓋其分析須基於抽象、假設性的調查，並與同樣係基於假設的風險緩解措施相結合⁸⁵。又與目前歐盟擬議的責任制度⁸⁶相結合，將使GPAIs的提供者面臨重大的責任風險⁸⁷。

另一方面，由於開發者和部署者、任務以及產品目的之間可能存在複雜的映射，GPAIs的提供者可能會透過契約安排等方式規避或轉嫁AIA的措施⁸⁸。若

GPAIs之合規義務直接轉嫁用戶承擔，不僅更加不可能執行，亦存在扼殺新創的風險，導致AI創新進一步反競爭集中⁸⁹（對不論資金規模的GPAIs提供者施加相同義務，亦可能導致反競爭⁹⁰，並使歐洲與美國或中國相比處於競爭劣勢⁹¹。

從權利保護的角度而言，若不能促使GPAIs的提供者考量系統安全性而採取保護措施，任何潛在的偏見、隱私侵犯、非法使用的內容或資料都將滲透到無數可能的未來應用中，加上跨市場部門運作的能力，說明了其缺陷甚至可能對經濟產生系統性的負面影響⁹²。

肆、結語

本文整理並說明了GPAIs的定義、與其他AI系統在應用及風險上的差異，並主張應依照平等原則與比例原則作不同規範。至於應如何應對GPAIs之監理挑戰，目前有論者提議應以基本權作為分類規範基準⁹³、以獨立風險類別規範⁹⁴，以及關注AI價值鏈的各方角色以進行責任義務分配⁹⁵等學說。

本文認為，價值鏈的責任義務分配值得相當重視，其係自商業實務考量產業鏈上各角色對資料與用戶等各種面向之風險控制程度以進行責任義務分配，相較於僅考量基本權利的風險程度，更能符合監理需求與實際可行性，應能減少權責錯位情形，並提供業者更明確的

遵循方向。此外，監理沙盒亦可提供相關業者充足的過渡期以推進風險分類過程⁹⁶。礙於篇幅限制，將另行為文說明前述學說，並就我國法制提出建議。

對GPAIs的積極監理刻不容緩，然而，不合理的規範將對價值鏈中的各方

造成傷害，並破壞對使用GPAIs之機構的信任，最終阻礙其發展⁹⁷。GPAIs對人類的未來影響極為重大，如何平衡基本權利保護與技術、經濟之發展，端賴各國立法者智慧的展現。♣

註釋

* 本文僅為作者個人之觀點，並不代表任職事務所之立場。

1. The AI Act, <https://artificialintelligenceact.eu/> (last visited June 11, 2023).
2. Global Partnership on Artificial Intelligence (2021). Responsible AI Working Group Report, at 7-8.
3. Ruschemeier, h. (2023). AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal. ERA Forum, 23, at 361-376.
4. OpenAI, <https://openai.com/about> (last visited June 11, 2023). 微軟為OpenAI投資人之一，請參閱Deutsche Welle (2023). Microsoft makes major investment in ChatGPT-creator OpenAI. <https://www.dw.com/en/microsoft-makes-major-investment-in-chatgpt-creator-openai/a-64492187> (last visited June 20, 2023).
5. Gal, U. (2023). ChatGPT is a data privacy nightmare. If you've ever posted online, you ought to be concerned. The Conversation. <https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283> (last visited June 11, 2023).
6. 關於GPT-3與GPT-4的差異，請參閱Saqib, M. (2023). The Future of AI: GPT-3 vs GPT-4: A Comparative Analysis. <https://becominghuman.ai/the-future-of-ai-gpt-3-vs-gpt-4-a-comparative-analysis-5ac525cb7ae1> (最後瀏覽日：2023年6月11日)。
7. Mehdi, Y. (2023). Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web. Microsoft. <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/> (last visited June 11, 2023).
8. Madiega, T. (2023). General-purpose artificial intelligence. European Parliamentary Research Service. <https://epthinktank.eu/2023/03/31/general-purpose-artificial-intelligence/> (last visited June 11, 2023).
9. Glaese, A. et al. (2022). Improving alignment of dialogue agents via targeted human judgements. arXiv preprint arXiv:2209.14375. Shuster, K. et al. (2022). Blenderbot 3: a deployed conversational agent that continually learns to responsibly engage. arXiv preprint arXiv:2208.03188. Scao, T. L. et al. (2022). Bloom: A 176b-parameter open-access multilingual language model. arXiv preprint arXiv:2211.05100. Hacker, P., Engel, A. & Mauer M. (2023). Regulating ChatGPT and other Large Generative AI Models. Working Paper. <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> (last visited June 11, 2023).
10. Lacy, P., Long, J., & Spindler, W. (2020). Disruptive technologies. The Circular Economy Handbook: Realizing

- the Circular Advantage, at 43-71. Danneels, E. (2004). Disruptive technology reconsidered: a critique and research agenda. *Journal of Product Innovation Management*, 21, at 246-258.
11. Lee, D. & Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International Journal of Environmental Research and Public Health*, 18, 1, at 271. Aung, Y. Y., Wong, D. C. & Ting, D. S. (2021). The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. *British medical bulletin*, 139, 1, at 4-15. Manheim, K. & Kaplan, L. (2019). Artificial intelligence: risks to privacy and democracy. *Yale Journal of Law & Technology*, 106, at 106-188. Naughton, J. (2023). ChatGPT isn't a great leap forward, it's an expensive deal with the devil. *The Guardian*. <https://www.theguardian.com/commentisfree/2023/feb/04/chatgpt-isnt-a-great-leap-forward-its-an-expensive-deal-with-the-devil> (last visited June 11, 2023). Bryant, J. (2023). Generative AI: A 'new frontier'. IAPP. <https://iapp.org/news/a/generative-ai-a-new-frontier/#> (last visited June 11, 2023). 義大利資料保護主管機關於2023年3月31日便考量資料隱私問題採取臨時行動封鎖ChatGPT，OpenAI配合改進安全措施後於同年4月28日恢復服務，請參閱 iThome，OpenAI更新隱私規範及加強資料保護，義大利解除ChatGPT禁令，2023年4月29日，<https://www.ithome.com.tw/news/156629> (最後瀏覽日：2023年6月11日)；Deutsche Welle (2023), ChatGPT: Italy blocks AI chatbot over privacy concerns (last visited June 11, 2023).
 12. Hsu, T. & Thompson, S. A. (2023). Disinformation Researchers Raise Alarms About A.I. Chatbots. *The New York Times*. <https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html> (last visited June 11, 2023). 歐盟執委會相關最新行動另參閱香港經濟日報，【AI監管】歐盟要求科企標記AI生成內容 打擊虛假訊息，2023年6月7日，<https://inews.hket.com/article/3542681/?lcc=an> (最後瀏覽日：2023年6月11日)。
 13. Hacker, P., Engel, A. & Mauer M. (2023). Regulating ChatGPT and other Large Generative AI Models. Working Paper. <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> (last visited June 11, 2023). Garcia-Murillo, M., MacInnes, I., & Renda, A. (2023). The Artificial Creatives: The Rise of Combinatorial Creativity from Dall-E to GPT-3. Queen's University Belfast Law Research Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4350802 (last visited June 11, 2023). Madiaga, T., *supra* note 8. Zuiderveen Borgesius, F. J. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. *The International Journal of Human Rights*, 24, 10, at 1572-1593.
 14. Ebers, M. et al. (2021). The European Commission's proposal for an Artificial Intelligence Act—a critical assessment by members of the Robotics and AI Law Society (RAILS). *Multidiscipl Scientific Journal*, 4, at 589-603.
 15. Hacker, P., Engel, A. & Mauer M. (2023). Regulating ChatGPT and other Large Generative AI Models. Working Paper. <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> (last visited June 11, 2023). Future of Life Institute (2022). General Purpose AI and the AI Act. <https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf> (last visited June 11, 2023).
 16. Taylor, L. (2023). Colombian judge says he used ChatGPT in ruling. *The Guardian*. <https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling> (last visited June 11, 2023).
 17. Winder, D. (2023). Does ChatGPT Pose A Cybersecurity Threat? Here's The AI Bot's Answer. *Forbes*. <https://www.forbes.com/sites/daveywinder/2023/02/03/does-chatgpt-pose-a-cybersecurity-threat-heres-the-ai-bot-s-answer/?sh=1015768f505d> (last visited June 11, 2023).
 18. DIGHUM (2023). Statement of the Digital Humanism Initiative on ChatGPT and a Possible New Online World. <https://caiml.org/dighum/statement-of-the-digital-humanism-initiative-on-chatgpt/> (last visited June 11, 2023).
 19. Marcus, G. (2023). A Skeptical Take on the A.I. Revolution. *The Ezra Klein Show*. *The New York Times*. <https://www.nytimes.com/2023/01/06/opinion/ezra-klein-podcast-gary-marcus.html> (last visited June 11, 2023). Beuth, P. (2023). Wie sich ChatGPT mit Worten hacken lässt, *Der Spiegel*. Bergman, A. S. et al. (2022). Guiding the release of safer E2E conversational AI through value sensitive design. *Association for Computational Linguistics*. <https://aclanthology.org/2022.sigdial-1.4/> (last visited June 11, 2023). Mirsky, Y. et al. (2022). The

- threat of offensive ai to organizations. *Computers & Security*. <https://arxiv.org/abs/2106.15764> (last visited June 11, 2023). Satariano, A. & Mozur, P. (2023). The People Onscreen Are Fake, The Disinformation Is Real. *The New York Times*. <https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html> (last visited June 11, 2023). Hacker, P., Engel, A. & Mauer M. (2023). Regulating ChatGPT and other Large Generative AI Models. Working Paper. <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> (last visited June 11, 2023).
20. McGuffie, K. & Newhouse, A. (2020). The Radicalization Risks Of GPT-3 And Advanced Neural Language Models. <https://www.middlebury.edu/institute/sites/www.middlebury.edu.institute/files/2020-09/gpt3-article.pdf> (last visited June 11, 2023).
 21. Gershgorn, D. (2021). ‘For Some Reason I’m Covered in Blood’: GPT-3 Contains Disturbing Bias Against Muslims. <https://onezero.medium.com/for-some-reason-im-covered-in-blood-gpt-3-contains-disturbing-bias-against-muslims-693d275552bf> (last visited June 11, 2023).
 22. Carlini, N. et al. (2021). Extracting Training Data from Large Language Models, arXiv:2012.07805v2. <https://arxiv.org/abs/2012.07805> (last visited June 11, 2023).
 23. Future of Life Institute, *supra* note 15.
 24. *Id.*
 25. CNN (2023). Experts are warning AI could lead to human extinction. Are we taking it seriously enough?. <https://edition.cnn.com/2023/05/30/media/artificial-intelligence-warning-reliable-sources/index.html> (last visited June 20, 2023).
 26. CNN (2023). Exclusive: 42% of CEOs say AI could destroy humanity in five to ten years. <https://edition.cnn.com/2023/06/14/business/artificial-intelligence-ceos-warning/index.html> (last visited June 20, 2023).
 27. Daly, A. et al. (2022). AI, Governance and Ethics: Global Perspectives (2019). *Constitutional Challenges in the Algorithmic Society*, at 182-201.
 28. OECD (2019). OECD Principles on AI. www.oecd.org/going-digital/ai/principles/ (last visited June 11, 2023). G20 (2019). Ministerial Statement on Trade and Digital Economy. https://g20digital.go.jp/asset/pdf/g20_2019_japan_digital_statement.pdf (last visited June 11, 2023).
 29. UNESCO (2022). UNESCO adopts first global standard on the ethics of artificial intelligence. <https://www.unesco.org/en/articles/unesco-adopts-first-global-standard-ethics-artificial-intelligence> (last visited June 11, 2023).
 30. Government of Canada(2020). Joint Statement from Founding Members of the Global Partnership on Artificial Intelligence. <https://www.canada.ca/en/innovation-science-economic-development/news/2020/06/joint-statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence.html?fbclid=IwAR0QF7jyy0ZwHBm8zkjRQqjblgiLd8wt939PbZ7EbLICPdupQwR685dlvw> (last visited June 11, 2023).
 31. 陳錫平、李建良、陳姿瑄（2022），歐盟人工智慧規則草案之初探——以市場、風險、價值及信賴為核心的管制架構，公共性與AI論壇（十二），<https://ai.iias.sinica.edu.tw/eu-ai-regulation-draft-minutes/>（最後瀏覽日：2023年6月11日）。
 32. EE Times Europe (2023). AI Act: EU Parliament Takes Step Toward AI Regulation. <https://www.eetimes.eu/ai-act-eu-parliament-takes-step-toward-ai-regulation/> (last visited June 20, 2023).
 33. Sharp, A. (2023). EU Lawmakers Pass Landmark AI Regulation Bill. *Foreign Policy*. <https://foreignpolicy.com/2023/06/14/eu-ai-act-european-union-chatgpt-regulations-transparency-privacy/> (last visited June 20, 2023).
 34. 童子賢，請參閱遠見，ChatGPT風潮該追？童子賢：得AI者得天下，2023年，<https://www.gvm.com.tw/article/103388>（最後瀏覽日：2023年6月11日）。
 35. 正如歐盟內部市場執委Thierry Breton所稱，AI在數位社會和經濟中提供了巨大的潛力，歐盟提出AIA並非

- 是想將AI的開發者驅離，而是希望能推動並說服他們在歐洲定下，其更補充說，歐盟不應依賴外國供應商，AI資料應在歐盟儲存和處理。請參閱Deutsche Welle (2023). EU: ChatGPT spurs debate about AI regulation. <https://www.dw.com/en/eu-chatgpt-spurs-debate-about-ai-regulation/a-65330099> (last visited June 20, 2023).
36. 奥特曼 (Sam Altman) 於2023年5月16日與會美國國會聽證會，表示AI可以解決人類最大的問題，但也可以強大到以難以預測的方式改變社會，呼籲政府應立法監管AI科技潛在的危險性。請參閱數位時代，ChatGPT之父：AI很危險！他建議用「三道紅線」，防堵失控的人工智慧，2023年5月18日，<https://www.bnext.com.tw/article/75294/ai-congressional-hearing-chatgpt-sam-altman> (最後瀏覽日：2023年6月11日)。
 37. 數位時代，奥特曼：OpenAI可能退出歐洲！歐盟AI法案怎麼規定？為何業界大佬喊出走？，2023年5月29日，<https://www.bnext.com.tw/article/75409/openai-eu-ai-act> (最後瀏覽日：2023年6月11日)。
 38. Hacker, P., Engel, A. & Mauer M. (2023). Regulating ChatGPT and other Large Generative AI Models. Working Paper. <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> (last visited June 11, 2023).
 39. Gutierrez, C. I., Aguirre, A. & Uuk, R. (2022). The European Union could rethink its definition of General Purpose AI Systems (GPAIS). OECD.AI Policy Observatory. <https://oecd.ai/en/wonk/eu-definition-gpais> (last visited June 12, 2023).
 40. Ranchordas, S. (2021). Experimental Regulations and Regulatory Sandboxes: Law without Order?. University of Groningen Faculty of Law Research Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3934075 (last visited June 11, 2023).
 41. Riegert, R. (2023). EU: ChatGPT spurs debate about AI regulation. Deutsche Welle. <https://www.dw.com/en/eu-chatgpt-spurs-debate-about-ai-regulation/a-65330099> (last visited June 20, 2023).
 42. Ruschemeier, h., *supra* note 3.
 43. Madiega, T., *supra* note 8.
 44. 由於General Purpose AI和Artificial General Intelligence中文皆譯為「通用人工智慧」，除此處之外，本文其他部分所稱之「通用人工智慧」皆指General Purpose AI而言。
 45. *Id.*
 46. Tamkin, A. & Ganguli, D. (2021). How Large Language Models Will Transform Science, Society, and AI. Stanford University Human-Centered Artificial Intelligence. <https://hai.stanford.edu/news/how-large-language-models-will-transform-science-society-and-ai> (last visited June 11, 2023).
 47. Routley, N. (2023). What is generative AI? An AI explains, World Economic Forum. <https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/> (last visited June 11, 2023).
 48. Guidetti, A. (2020). Artificial Intelligence as General Purpose Technology: An Empirical and Applied Analysis of its Perception. Faculty of Economic and Political Sciences Master of Science in Economy and Policies of the Territory and the Company. UNIVERSITÉ DE LA VALLÉE D'AOSTE.
 49. Bommasani, R. et al. (2022). On the Opportunities and Risks of Foundation Models. arXiv:2108.07258v3.
 50. Ganguli, D. et al. (2022). Predictability and surprise in large generative models. ACM Conference on Fairness, Accountability, and Transparency, at 1747-1764.
 51. Hoffmann, J. et al. (2022). Training compute-optimal large language models. arXiv preprint arXiv:2203.15556. Hacker, P., Engel, A. & Mauer M., *supra* note 13.
 52. Küspert, S., Moës, N., & Dunlop, C. (2023). The value chain of general-purpose AI. Ada Lovelace Institute. <https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/> (last visited June 11, 2023).
 53. Madiega, T., *supra* note 8.
 54. Brown, T. et al. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, at 1877-1901. Kim, B. et al. (2021). What changes can large-scale language models bring? intensive

- study on HyperCLOVA: Billions-scale Korean generative pretrained transformers. arXiv preprint arXiv:2109.04650. 許多問題源於GPAIs需求大量資料的特性，蓋其意味著GPAIs之開發人員必須經常依賴網路上公開可用之訓練資料，例如網路上人類社會的歧視，經過系統演算法後，將產生偏移而被放大，並進一步被各種基於GPAIs之應用程式直接投放影響廣大受眾，請參閱Hacker, P., Engel, A. & Mauer M., *supra* note 13。
55. Hacker, P., Engel, A. & Mauer M., *supra* note 13.
 56. Wang, P. (2004). Toward a Unified Artificial Intelligence.. AAAI Technical Report (1). Bieger, J. et al. (2016). Evaluation of general-purpose artificial intelligence: why, what & how, Evaluating General-Purpose AI. Nilsson, N. J. (2005). Human-level artificial intelligence? Be serious!. AI magazine, 26, 4, at 68–68. Meltzer, J. P. (2018). The impact of artificial intelligence on international trade, Brookings Institute. Szegedy, C. (2020). A promising path towards autoformalization and general artificial intelligence. in International Conference on Intelligent Computer Mathematics, at 3-20.
 57. Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on a general problem solving program. IFIP congress, 256, at 64.
 58. OpenAI & Pilipiszyn, A. (2021). GPT-3 powers the next generation of apps. <https://openai.com/blog/gpt-3-apps> (last visited June 11, 2023).
 59. Council of the European Union (2022). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Presidency compromise text. <https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf> (last visited June 11, 2023).
 60. Council of the European Union (2022). Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union - Text de compromis de la présidence - Article 3, paragraphe 1 ter, Articles 4 bis à 4 quater, Annexe VI (3) et (4), considérant 12 bis bis. <https://artificialintelligenceact.eu/wp-content/uploads/2022/05/AIA-FRA-Art-34-13-May.pdf> (last visited June 11, 2023).
 61. Hacker, P., Engel, A. & Mauer M., *supra* note 13.
 62. Hernández-Orallo, J. et al. (2021). General intelligence disentangled via a generality metric for natural and artificial intelligence. Scientific Reports, 11, 1, at 1-16. Gutierrez, C. I. et al. (2022). A Proposal for a Definition of General Purpose Artificial Intelligence Systems. Future of Life Institute – Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4238951 (last visited June 12, 2023).
 63. Gutierrez, C. I. et al. (2022). A Proposal for a Definition of General Purpose Artificial Intelligence Systems. Future of Life Institute – Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4238951 (last visited June 12, 2023).
 64. *Id.*
 65. Hernández-Orallo, J. (2017). Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. Artificial Intelligence Review, 48, 3, at 397-447. Bommasani, R. et al., *supra* note 49.
 66. Gutierrez, C. I. et al., *supra* note 62.
 67. Gutierrez, C. I. et al., *supra* note 62.
 68. Hernández-Orallo, J. (2017). Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. Artificial Intelligence Review, 48, 3, at 397–447. OECD (2022). OECD Framework for the classification of AI systems. <https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm> (last visited June 11, 2023). Thórisson, K. R. et al. (2016). Why artificial intelligence needs a task theory. <https://arxiv.org/abs/1604.04660> (last visited June 11, 2023).

69. Gutierrez, C. I. et al., *supra* note 62.
70. OECD (2022). OECD Framework for the classification of AI systems. <https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm> (last visited June 11, 2023). Ganguli, D., et al. (2022). Predictability and surprise in large generative models. <https://arxiv.org/abs/2202.07785> (last visited June 11, 2023).
71. Gutierrez, C. I. et al., *supra* note 62.
72. *Id.*
73. Gutierrez, C. I. et al., *supra* note 62.
74. Gutierrez, C. I., Aguirre, A. & Uuk, R., *supra* note 39.
75. Hacker, P., Engel, A. & Mauer M., *supra* note 13.
76. European Parliament Press (2023). MEPs ready to negotiate first-ever rules for safe and transparent AI. <https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rule-s-for-safe-and-transparent-ai> (last visited June 20, 2023).
77. Gutierrez, C. I., Aguirre, A. & Uuk, R., *supra* note 39. Ruschemeier, h., *supra* note 3.
78. Benifei, B. et al. (2023). Proposal for a Regulation on a European approach for Artificial Intelligence, A Europe Fit for the Digital Age. <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence?sid=7001> (last visited June 11, 2023).
79. European Parliament (2023). AI Act: a step closer to the first rules on Artificial Intelligence. <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence> (last visited June 11, 2023).
80. European Parliament Press (2023). MEPs ready to negotiate first-ever rules for safe and transparent AI. <https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rule-s-for-safe-and-transparent-ai> (last visited June 20, 2023).
81. Matthews, D. (2021). A new type of powerful artificial intelligence could make EU's new law obsolete. Science|Business. <https://sciencebusiness.net/news/new-type-powerful-artificial-intelligence-could-make-eus-new-law-obsolete> (last visited June 11, 2023). Gutierrez, C. I. et al., *supra* note 62.
82. Helberger, N. & Diakopoulos, N. (2023). ChatGPT and the AI Act. *Internet Policy Review*, 12(1). <https://doi.org/10.14763/2023.1.1682> (last visited June 11, 2023).
83. Future of Life Institute, *supra* note 15. 例如，經過訓練可以處理自然語言的GPT-3意外地獲得了用編程語言編寫基本程式的能力，請參閱Chen, M. (2021). Evaluating Large Language Models Trained on Code. Cornell University. arXiv:2107.03374v2. <https://arxiv.org/abs/2107.03374> (last visited June 11, 2023).
84. 本文目前所見多數文獻多認為GPAIs將產生高風險義務，惟亦有少數認為文本生成器如ChatGPT等屬低風險或無風險類別，如AFP, AP (2023). EU lawmakers lay groundwork for 'historic' AI regulation. *Deutsche Welle*. <https://www.dw.com/en/eu-lawmakers-lay-groundwork-for-historic-ai-regulation/a-65909881> (last visited June 20, 2023).
85. Hacker, P., Engel, A. & Mauer M., *supra* note 13.
86. 目前為採取AIA、AI Liability Directive (AILD or Directive on adapting non-contractual civil liability rules to artificial intelligence)及修訂後之Directive on liability for defective products (PLD) 三管齊下之框架。請參閱Nawaz, S. A. (2022). The Proposed EU AI Liability Rules: Ease or Burden?. <https://europeanlawblog.eu/2022/11/07/the-proposed-eu-ai-liability-rules-ease-or-burden/> (last visited June 20, 2023).
87. Hacker, P. (2022). The European AI Liability Directives - Critique of a Half-Hearted Approach and Lessons for the Future. Working Paper. <https://arxiv.org/abs/2211.13960> (last visited June 11, 2023). Spindler, G. (2022). Die Vorschläge der EU-Kommission zu einer neuen Produkthaftung und zur Haftung von Herstellern und Betreibern

- Künstlicher Intelligenz. Computer und Recht, at 689-704. Hacker, P., Engel, A. & Mauer M., *supra* note 13.
88. Gutierrez, C. I. et al., *supra* note 62.
89. Muller, C. (2022). AIA in-depth #1: Objective, Scope, Definition. ALLAI. <https://allai.nl/wp-content/uploads/2022/03/AIA-in-depth-Objective-Scope-and-Definition.pdf> (last visited June 11, 2023).
90. Hacker, P., Engel, A. & Mauer M., *supra* note 13.
91. Future of Life Institute, *supra* note 15.
92. *Id.*
93. Ruschemeier, h., *supra* note 3.
94. Helberger, N. & Diakopoulos, N., *supra* note 82.
95. Hacker, P., Engel, A. & Mauer M., *supra* note 13.
96. Felice, A. D. & Chasserieau, J. (2023). 'A text we can work with': Reaction to European Parliament vote on AI Act. <https://www.digitaleurope.org/news/a-text-we-can-work-with-reaction-to-european-parliament-vote-on-ai-act/> (last visited June 20, 2023).
97. Floridi, L. et al. (2022). capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4064091 (last visited June 11, 2023).

關鍵詞：人工智慧、AI、通用人工智慧、ChatGPT、生成式

DOI：10.53106/279069731613

(相關文獻 ◀ 月旦知識庫 www.lawdata.com.tw ;
更多裁判分析 ◀ 月旦法律分析庫 lawwise.com.tw)