# AI 系統運用與基本人權之衝突及治理

東海大學　法律學院　范姜真媺

## 目錄

## 壹、源起及背景

### 一、AI 系統之概念及運用

近年大家在日常生活，各類論文中常聽聞「人工智能－AI」（Artificial Intelligence）一詞[1]，但對 AI 系統定義卻仍是各有主張，本文依 ISO 所揭示之定義，將研究開發 AI 系統之學問領域歸類為「計算機科學」，亦即 AI 系統乃係「遂行推論、學習、自我能力提升等，類似具有人類智慧功能之資料處理系統」，及「設計結合人類知識，進行學習、推論等機能之電腦演算模式及系統[2]。」而在建立 AI 之運用系統前，必須讓電腦進行「機器學習」或深度學習，即由人輸入大量資料、資訊或知識至電腦中，依設定之演算法（Algorithm），將集結各類資料之資料庫等分析出其間有共同特徵及關聯性者予以類型化，並建立歸納為各類型之演算模式，之後將須作判斷之資料等，輸入電腦內依建立之演算模式，即可作出推測、判斷，甚至行動決定。

因 AI 系統之機器學習需蒐集大量資料作為學習材料之基底（Data Pool），因此常與網路結合，快速大量蒐集各開放平台之資料，利用大數據演算後，彙集成各類資料庫作為學習材料之基底。故而具有下列特色：其資料之量（Volume）、多樣性（Variety）傳輸處理之速度（Velocity），真確性(Veracity)及經過機器學習，分析後所得推測結果之利用價值（Value）[3]。

且因 AI 系統演算模式之建立至分析、推測結果之作成，係由電腦依設定之演算程式完成，其間未有人為介入，外觀上排除人主觀、感情等因素，形成判斷之結果看似客觀性且正確性高。又因其資料處理速度快，節省時間及人力成本，於少子化、高齡化之現今社會自是被廣泛運用。政府機關可預測交通流量，進行

---

[1] 有人稱為「人工智慧」，亦有人稱為「人工智能」，本文則一律以原文之略稱「AI」為用詞，於此合先述明。

[2] ISO/EC 2382:2015 （Information Technology Vocabulary）

[3] 參閱邱文聰，初探人工智慧中的個資保護發展趨勢與潛在的反歧視難題，收錄於人工智慧相關法律議題芻議，元照出版公司，2019 年 3 月，頁 160；劉靜怡，人工智慧潛在倫理與法律議題鳥瞰與初步分析－從責任分配到市場競爭，收錄於人工智慧相關法律議題芻議，元照出版公司，2019 年 3 月，頁 35。

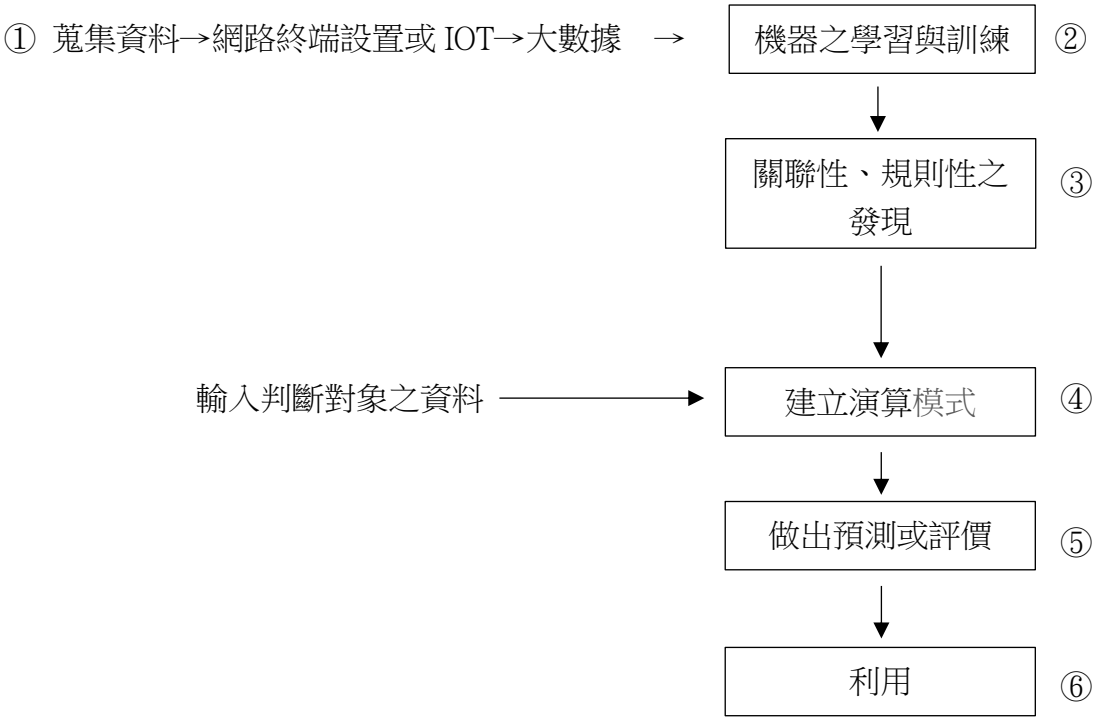
號誌調控，預估糧食需求量調整生產計畫以平抑價格等，甚至預測犯罪之熱區，進行警力之配置，或對人民社會補助之申請許可自動作出決定等。另在民間企業利用具 AI 系統功能之機器人擔任生產線之作業，以具人臉辨識功能之監視攝影器管控辦公大樓人員之進出，或金融機構以 AI 系統對申請貸款人進行人物剖析（Profiling）評定信用降低貸款風險。醫療上則利用作疾病判斷、處方開立，甚至手術之施行、病人之照護等。事實上自駕車之道路行駛亦為先蒐集各種道路環境、號誌、路上人物動靜行止、事故原因等資訊，由機器學習訓練後建立演算模式，依自動化系統操控車輛之行駛。AI 系統之運用已浸透在人們生活當中，依 IDC（International of Data Corporation）對世界 AI 系統市場之調查，2021 年 AI 系統市場包含軟硬體服務之市場約有 35 兆日幣規模，至 2024 年預估將有 59 兆日幣[4]。但在 AI 系統為世界帶來許多生活上、交通上、醫學上、產業上等便利與進展之同時，卻對個人基本權造成莫大之威脅而有許多應檢討之課題。

二、AI 系統運用方式之概觀

如前所述，AI 系統之運用需大量資料作為學習材料，因此常藉各種電腦網路終端設置或物聯網（Internet of Things：IoT）集結成大數據（Big Data）快速、大量蒐集存於世界各處之資料，而經機器學習訓練後，發現人類過往無法察覺之各種資料或事象間之相關性或規則性，並分析或歸納出得歸屬於特定集團之事或人常具有一定性格特質或行為方式，建構 AI 演算模式，之後再將標的之人、事、物等資料(raw data)，鍵入 AI 系統後自動做出判斷並下達動作指令。

[4] IDC Forecasts Improved Growth for Global AI Market，IDC 為經營有關 IT 或 AI 與有關調查、分析、諮詢有關業務之全球性企業。參閱石村耕治，EU の包括的 AI システム規制始動，TC フォーラム研究報告 2021 年 5 月號，頁 17。

運用流程以下圖示之：




**貳、對人性尊嚴之威脅**

依上述之 AI 系統建立及運作流程至最終之利用，每個階段所隱藏之對人性尊嚴、基本權等形成侵害之疑慮。

一、資料庫建置階段

1、過剩資料之搜集—對個人資訊自主權之威脅

AI 系統之研發於最初階段因進行機器學習須建立資料庫供作學習材料，故常結合 IoT 或各種電腦端末裝置大量蒐集各種資料形成大數據，作為機器學習之材料，經機器反覆訓練學習後建立演算模式 。且因 AI 系統之預測準確度與蒐集資料量呈現比例關係。在學習資料越多，演算模式所作之分析及判斷越準確之誘因下，往往大量蒐集資料，致使個人之資訊自主權、隱私權遭受嚴重威脅[5]。例如為甄選職員而蒐集其在社群網站（SNS）與朋友交往之資料。再如自駕車自動操

[5] 寺田麻佑，人工知能（AI）技術の進展と公法学の変・公法研究第 82 号，頁 206，有斐閣 2020 年。福岡真之介編著，AI の法律と論点，頁 207、240 頁，商事法務，2018 年。山本龍彥論著，AI と憲法，頁 85，日本経済新聞出版社，2018 年。

縱系統之建立及運用，須蒐集大量公道上行人、車輛影像之行走狀態建置資料庫以供機器學習辨識及為安全行駛模式之建立。如此在公共領域大量蒐集個資，違反個資法之資料最小限原則，對個人之資訊自主權即有侵害之虞[6]。

2、隱藏之偏差：

如前所述 AI 系統依機器學習、訓練建立演算模式時，係分析資料後篩選出各種資料或事項之相關性或規則性，歸納出類型，並建立判斷之演算模式。但被資料庫所蒐集之學習資料，往往包含社會長年累積下來對特定之族群或事物之觀感或刻板之表徵，其中隱藏偏見或差別意識，而以此做為學習材料給機器進行學習訓練後所建立之演算模式，將偏見或差別待遇包含於其中，且因依 AI 系統演算不亦被查覺，而加重其嚴重性[7]。例如：醫學院在甄選入學新生時依畢業生及在學生之資料所建構之演算模式，以對申請學生之學習能力、可塑性及從事醫療工作適性做出是否合格之決定。如此，過去甄選入學新生時可能存在之偏見依然存在且被沿用，例如出身之高中、父母之職業、性別等之要素[8]。再如少數族群（Underrepresentation）之特質，因資料庫中無法顯示其存在之分量而被忽略，據此進行機器學習後建立之演算模式即很容易帶來差別待遇[9]。故在建立 AI 系統之演算模式時，如何蒐集正確完整之學習資料？如何檢視資料庫所蒐集資料具有足夠之代表性，未隱藏偏見，常為有識者所指摘之問題[10]。

二、演算模式建立階段

1、黑盒子之形成

---

[6] 松村英寿，AI と個人情報· プライバシー，同前揭福岡真之介編著，AI の法律と論点，頁 239-240。

[7] 鈴木悠介，AI に関する倫理の問題，同前揭註福岡真之介編著，AI の法律と論点，頁 383-384。粟原聡，AI 威脅論の正体と AI の共生，総務省，学術雑誌，情報通信政策研究，第 4 卷第 2 号，頁 1-46。福田雅樹、林秀弥、成原慧編著，「AI がつなげる社會」，弘文堂，2019 年。山本龍彦，AI と「個人尊重」，頁 327。

[8] Federal Trade Commission, FTC,2016 年，REPORT, big Data: A Too1 for Inclusion or Exclusion? Understanding the Issue at 28, 2016, JANUARY

[9] FTC Report，Big Data: A Tool for Inclusion or Exclusion？27（January）2016.

[10] 栗原聡，前揭註 7。鈴木悠介，前揭註 7，頁 385。中川裕志，AI 倫理指針の動向，とパーソナル AI エージェント，総務省学術研究雑誌，情報通信政策研究，第 3 卷第 2 号，頁 1-6。

如前圖 AI 系統在運用係由人鍵入大量學習資料，經機器學習及訓練以發現蒐集之資料庫中各種資料之關聯性、規則性，將共通之要素篩選出作為判斷基準而建構演算模式。故 AI 系統機器學習及建構演算模式過程中必然具有「不可預測性（Inscrutability）」及「非主觀性（Non Intuitiveness）」[11]。此過程中確實免除人主觀之介入，且能發現眾多資料中人所未能意識到之「關聯」或「法則性」，然因演算模式之邏輯推演過程極為複雜，很難對一般人說明，未具有專業知識之人亦無法對其準確率作檢驗。再被選出構成演算模式之判斷基準，其被選出之理由及其對判斷之決定發揮如何之作用，亦難以被人理解，而形成所謂「黑盒子」。行政機關對人民申請之許、認可案件為否准之決定時，必須對受不利益處分之相對人儘可能說明其理由，甚至應給其陳述意見之機會。但如依 AI 系統作出決定時，因演算模式之複雜性及理解困難性，行政機關應如何充分說明其理由？確保行政程序之透明化，以維護正當程序原則，為學者所指摘之問題[12]。

2、陷於功利主義迷思之危險

如前所述現代民主國家係建立在維護人性尊嚴，保障個人得依自律、自決之權，決定自己生活方式、追求自己之幸福之以人為本之理念上，故而法制之建構，均以人為權利、義務之主題，個人因自己之主觀意思、客觀行動，而負擔法律上權利、義務之效力。於具體實務案例判斷權利、義務歸屬時往往因人、事、時、地之差別，又彼此交互影響，其間因果關係錯綜複雜須進行耙梳，在許多案例尚須因社會正義或公益之保全進行利益衡量以符合比例原則。AI 系統於學習所後建立之固定演算模式，在錯綜複雜之因果關係中能否衡量各種利益做出妥適之判斷？換言之，由 AI 系統對人依事前、定型化之演算模式決定權利、義務之歸屬，判斷其行為價值、責任輕重，對以人為本、尊重個人之理念，恐有未合。再者法

11 山本龍彥，「完全自動意思決定」のガバナンスー行為統治型規律からからガバナンス統治型規律へ，総務省，学術雑誌，情報通信政策研究，第 3 卷第 1 号，頁 1-30。此說法為山本教授引用 Andrew D. Selbst & Solon Barocas, The Intuitive Appeal of Explainable Machines, 87 FORPETAM L. REV 1085,1100 (2018）一文。

12 中川裕志，前揭註 10，頁 1-6-9。鈴木悠介，AI の基礎知識，福岡真之介編著，前揭註 6，頁 11-12。山本龍彥，前揭註 5，頁 69-70。

律立法時，常為能在個人與個人、個人與社會、國家間發生利益衝突時適當調整，往往保留給執法人有裁量空間，以貼合社會之價值觀。在 AI 系統預先設定之演算模式，能否適當反映各種價值觀，做出符合人性及社會正義之判斷？且因 AI 系統演算模式之建立，常係依數與量為基礎之功利主義判斷方法，少數族群應被尊重更有價值之利益因此被無視。

此亦為許多學者在討論 AI 系統之開發利用之議題時，對涉及人之生死，及行為價值判斷時，提出之質疑且要求應慎重以對者[13]。

三、判斷或預測階段

1、人性尊嚴之侵蝕

人性尊嚴為憲法之核心價值，基本權之規定在為實現此價值來看。基於人性尊嚴，每個個人均為獨立之權利主體，有其生存之價值與意義，應被平等對等並尊重。不受國家或他人之支配。而個人自律及自決權即為人性尊嚴之核心內涵，亦即基於人格自由發展下，得自主決定自己生活方式，人生未來之發展方向及採取行動。因此始能形成多元化之社會，並維護民主政治之運作[14]。

但 AI 系統運用演算程式如前述係分析資料庫各類不同資料以發現其關連性及規則性並選擇作為判斷是否歸屬某特定集團，而預測或判斷（Profiling），特定個人之能力、行動傾向、健康狀況、信用度甚至再犯之可能性。亦即 AI 系統之判斷，其實隱含著個人因其與某特定集團被認定之特定表徵或特質具有共同性，而被粗暴判斷或預測其能力、信用未來發展等。抹殺了個人其他存在應被尊重之人格特質，日本學者稱此為因 AI 系統產生之新集團主義[15]，侵害個人應被公平獨立尊重之人性尊嚴。

2、衍生性個資之取得

一般而言個資之取得必須有特定目的並符合一定要件始得合法蒐集，尤其是

[13] 寺田麻佑，前揭註 5，頁 208。中川裕志，前揭註 10，頁 1-14～15。
[14] 參閱許志雄，憲法上個人尊嚴原理，「憲法之基礎理論」，頁 45，稻禾，1993 年初版。李震山，人性尊嚴與人格保障，頁 8-16，元照出版公司，2000 年，2 月。
[15] 山本龍彥，前揭註 5，頁 67。

病歷、健康檢查等敏感性特種資料之取得，原則上是被禁止的，但經由 AI 系統對蒐集所得個人之購物清單，上網瀏覽之網站紀錄或是在 FB 上之點讚或留言，進行分析後往往可推知當事人之政治、宗教信仰傾向，甚至健康狀況等原蒐集資料外之「衍生資料」[16]。如此個資之取得可能迴避了個資法之適用，更使當事人之隱私、窺伺、資訊自主權受到侵害。

四、利用階段—個人自律、自決權之壓縮

1、新弱勢族群之形成

如前述 AI 系統進行人物剖析(profiling)將具有特定屬性或特徵之個人歸類於具有相同屬性或特徵之已被黏貼特定標籤之集團或族群（例如有暴力傾向、偏好之行動方式或隱藏遺傳性疾病、抗壓性低、低成就等），從而在其身上投射所被歸屬之集團或族群之既視感、評價其能力，發展可能性、信用度、健康狀況等，進而決定是否僱用、升遷、准予入學或貸款。基於人性尊嚴內涵之個人應被尊重並平等對待，每個人均有規劃人生發展藍圖，依自己能力、努力，實現自己人生藍圖之自主、自律權。卻由於 AI 系統依過去有意識或無意識留存之資料作出判斷後，個人如原罪般被迫背負被歸屬集團或族群之既定標籤，進而扼殺其將來應有依自己之意志及能力改變人生之機會，侵害人之自決權及重生之權利，形成所謂「新的弱勢族群」[17]，如封建時代之階級制度被重新建立，而違反平等原則。

2、鎖定（Target）行銷之操控

權力業者為更有效率行銷產品或服務，往往蒐集個人在網路上之購買履歷、閱覽紀錄等，經由 AI 系統進行人物剖析後推測出個人之消費習慣、能力，甚至掌握消費者心理狀態，而進行高誘導性之精準行銷。此乃掌控特定消費者心理或性格上弱點（Particular Vulnerabilities）針對性推銷特定商品或服務，且係在個人難以察覺方式下操作消費者選擇環境、影響。消費者得理性考量自己之實際需求，

[16] 新保史生，AI の利用と個人情報保護制度における課題，前揭註 7，福田雅樹等編著，AI がつなげる社会，頁 228-229。

[17] 山本龍彦，前揭註 5，頁 72。矢守亜夕美，AI によるスコアリングが引き起こすバーチャル、スラム化，特集人事の 72 AI 原則，Works No. 156 oct-Nov 2019，頁 15。

及與其他相類似產品價值比較後，自律性決定是否選購之權利，無形中被窄化甚至侵害消費者心理之行銷行為[18]。如此之交易契約是否認為基於自由意思而成立者並非無疑義。

3、扭曲民主政治[19]

民主政治之運作須根植於言論自由市場之成立，人民需於公共公開之場域得自由發表自己之見解，聽取他人之意見，經相互辯論、溝通及省思、得到真理並形成公共意思，再透過選舉組成議會代表制定法律，規範社會生活秩序，此亦為表現自由四個重要價值之一。現因資訊數位化後任何人均得在網路平台上以自媒體身分自由進行雙向大量資源之收、發，截然不同於過往民眾多處於資訊接收方。如此發展下資訊流通之更加快速、繁多，似變得更有助於公共議題之討論。然因各種社群網站提供之操作技術，讓閱聽者得依自己之偏好選取平台上資訊及限定交流對象，輕易形成在認同自己意見之交流平台上汲取互換同溫層現象。尤有甚者，因網路平台蒐集利用人之閱覽後而再代入 AI 系統之分析推測，得掌握網路使用人偏好之資訊，在先為你篩選以方便獲取所欲得之資訊之名義下，提供客製化（Personalization）新聞資訊，將網路利用人置於宛如隔離泡膜（Field Bubble）之資訊環境中，與其他資訊隔絕而孤立，網路利用人在未有其他不同觀點、不同價值觀之資訊做比較、思辨下，更執著於自己所見，而傾向於極端主義，實不利於民主政治之發展。事實上 2016 年美國總統大選時，臉書使用者之資料由當時之網路平台提供給劍橋分析後被利用為操縱選舉之事，殷鑑不遠[20]。

4、監視社會之形成

為維持治安或拘捕逃犯、防止恐怖活動，政府執法機關架設具有進行遠距生

[18] 山本龍彥，前揭註 5，頁 104。古谷貴之，AI と自己決定原理，山本龍彥編著註 5，頁 134。大屋雄裕，ロボット・AI と自己決定する個人，弥永真生・宍戸常寿編「ロボット・AI と法」，頁 68，2018 年 4 月，有斐閣。

[19] 參閱水谷瑛嗣郎，AI と民主主義，山本龍彥論，同前註 5，頁 304-309。寺田麻佑前揭註 5，頁 209。

[20] 湯淺墾道，AI ネットワークと政治参加政策決定，福田雅樹、林秀彌、成原慧編著，前揭註 7，頁 314。

物辨識功能之 AI 系統（Remote Biometric Identification System）之監視攝影器，亦即在事前不知該特定人是否會出入之特定處所或能否被辨識出之狀況下，自遠端將攝錄下該特定人之生物特徵透過與包含於資料庫之資料作比對下而識別出該特定人[21]，即便是因政府機關為執行法律，在公眾得出入之公共空間設置如此監控系統，將會對許多人之私生活造成影響，有常時處於被人監控下活動之恐懼而怯於出入公共場所，其集會活動等表現自由因此萎縮[22]。且被用作比對、組合之資料庫中之資料如存有錯誤偏差時，即可能形成執法之不公平。利用 AI 之即時遠距生物辨識監視系統，如得由公權力機關於不受任何法律對其設置要件及利用目的作限制或作有效之監督時，有可能被濫用為監視全民活動之天網，對民主政治之運作造成嚴重威脅。

## 參、AI 系統之原則及治理

一、緣起

如前所述，AI 系統之利用已浸透在社會各角落，且在此之前法制之定立向以個人為權利、義務之主體，個人基於自由意志決定自己行動，並為其結果擔負責任建構之規範或管制模式，可能因 AI 系統替代人作決定並行動而有所改變，重新建立法規範之對象或要件。

除此之外，如前所述 AI 系統之利用可能涉及人性尊嚴之侵害，壓迫個人自律、自主權等，並可能引發民主政治運作之風險，故在 AI 系統正發展及利用之際，自不得無視上開風險發生之可能，並維護憲法之基本價值，而應慎重考量建立規範降低或防止風險發生於未然。

又，在全球化高度資本主義下國際間產業競爭激烈，AI 系統之開發、利用背後包含有研發者之企圖，企業之野心、國家經濟政策考量及一般利用人之需求等

---

[21] 參閱 EU2021 年 4 月公布之 Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence （Artificial Intelligence Act）and amending Certain Union Legislative Act.之第 3 條 36 號規之 xx，https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_2&format=PDF。

[22] 參閱前揭註前言第 18。

種種不同目的之誘因或趨動力，在功利主義及經濟發展或產業創新至上、效率第一之追求下，可能導致開發者等在有意識或無意識間，挑戰社會價值或人權保障之底線。因此為預防危害或降低風險之發生，訂立一定原則供研發人自發想之初即能知悉紅線之存在，而降低侵害人權或挑戰社會價值之 AI 系統之開發[23]。

二、基本原則提出及發展

在上述之思維下，世界各國及國際組織在近年開始對 AI 系統開發及利用上應有之基本原則進行檢討，希望能預防或降低 AI 系統帶來之風險，建構得到社會信賴，與人類共存得永續發展之 AI 系統[24]。

世界組織或各國於考量 AI 系統發展、運用原則內容及實踐之方式時，雖因各國有其經濟發展需要及戰略上之考量，其主要重點卻並未有重大差異。

以下簡述近年各國或世界組織所討論並揭示 AI 之倫理或原則：

（一）日本與 OECD：

AI 系統之運用因網路等資訊科技之發展早已跨越國界，日本與 OECD 均意識到 AI 系統之研究利用對國家經濟、產業等各方面競爭力提升之重要性；然同時亦因對社會可能帶來各種風險及負面影響，集眾人仍對其抱持不安與不信任感。故而各國企圖在 AI 系統研發上取得領頭羊地位，引導 AI 系統發展方面，同時為能消弭社會大眾對 AI 系統抱持之不信任及不安，故而自 2016 年開始日本積極在國際間倡議 AI 系統倫理原則之策定，於同年 4 月 G7 在日本香川開會時，提案有關 AI 系統倫理課題，應由 G7 之國家及 OECD 等國際組織進行國際討論擬共識，並得到各國贊同[25]。延續上開政策，後總務省資通信政策研究所於同年 10 月舉行「AI 網路社會推進會議」[26]，經各方討論及意見募集，於 2017 年

[23] 座談会，AI、ロボトの研究開発をめぐる倫理と法，久木田水生之發言，同前揭註 15 福田雅樹等編著，頁 105。

[24] 其詳細狀況得參閱鈴木悠介，AI と倫理，同前揭註 5 福岡真之介編著，頁 358-371。

[25] 成原慧，AI の研究開発に関する原則、指針，同前揭註 7，福田雅樹等編著，頁 87。

[26] 「AI ネットワーク社会推進会議」，http://www.soumu.go.jp/menu-news/s-news/oliicpol_02000052.html

7 月公布「推進 AI 網路社會之國際性議題—報告書 2017」[27]，此報告書中揭示「供國際討論議題之 AI 開發指引案」[28]，包含有目的、以人為中心之社會實現、技術中立性確保等 5 項基本理念及透明化原則、安全原則、隱私原則說明原則等 9 項 AI 系統開發原則，並於其指引之序文中說明此指引定位為「G7 或 OECD 進行 AI 國際性討論之基礎文書」，且因 AI 系統技術尚在發展中，所謂「AI 開發原則」及「AI 開發指引」，二者均導入具有強制規制力之規範為目標，性質上為屬於國際共有「非規制性、無拘束力之軟性規則」之指引案[29]。

後 OECD 則於 2019 年 5 月之理事會，接受日本主導之提案並議決通過「人工智能 OECD 之建議(Recommendation on Artificial Intelligence )」[30]。OECD 所建議之五項原則及五個建議，基本之重點與前述日本指引差異不大。

而日本則於 2018 年 5 月開始檢討政府統一之 AI 系統社會原則，以前開總務省之指引為基礎而於 2019 年 3 月立定「人為中心之 AI 社會原則」，由總務省於 2019 年 3 月 24 日公布：「人工智能與人間社會懇談會」報告書[31]。

此外，AI 系統網路社會推進會議將 AI 系統服務提供者、最終利用者及資料提供者被期待應注意之原則，於 2018 年 7 月整合為「AI 利活用原則案」，並以此為基礎於 2019 年 8 月 9 日公布「AI 利活用指引(AI 利活用ガイドライン)」[32]，其中揭示七項基本理念：1、由人與 AI 系統共存，使所有人得享受其利益而實現尊重人性尊嚴及個人自律之以人為中心之社會。2、尊重利用者之多樣性、包攝不同背景及各種價值觀之群眾。3、謀求解決各種課題，實現永續性社會。

---

[27] 「報告書 2017 -AI ネットワーク化に関する国際的な議論の推進に向けて」，http://www.soumu.go.jp/main-content/000499624.pdf

[28] 「国際的な議論のための AI 開発ガイドライン案」，https://www.soumu.go.jp/main_content/000499625.pdf

[29] 同前揭註序文。福田雅樹，「AI ネットワーク」おびそのガバナンス，同前揭註 7 福田雅樹等編著，頁 21-22。

[30] OECD, Recommendation of the Council on Artificial Intelligence，OECD/LEGAL/044914

[31] 日本内閣府，人工知能と人間社会に関する懇談会報告書，平成 29 年 3 月 24 日，https://www8.cao.go.jp/cstp/tyousakai/ai/summary/aisociety_jp.pdf

[32] AI ネットワーク社會推進会議，AI 利活用ガイドライン，令和元年 8 月 9 日，頁 1-2。

4、最大限度尊重民主主義社會價值，抑制侵害權益之風險，確保利益與風險之平衡。5、實現各利害關係人間適當分享 AI 系統具有之能力或知識。6、與國際共有非具拘束性、軟性之 AI 系統活用指針。7、追隨 AI 系統之發展持續性修正本指針。又關於活用原則有:①適當利用原則:人與 AI 系統間及利用人間，適當之分配任務，在正當之範圍及方法下利用 AI 系統、②適當學習原則：應注意供學習用資料之品質、③聯繫原則：注意 AI 系統服務提供者、利用者、資料提供者相互間之聯繫，及可能引發或增高風險之可能性。④安全原則：AI 系統應透過技術保護利用人、第三人之生命、身體及財產之安全。⑤系統安全維護原則：利用人及資料提供人應注意 AI 系統之安全性。⑥隱私原則：利用人及資料提供人應注意不得侵害自己、或他人之隱私。⑦尊嚴、自律之原則：AI 系統應尊重個人尊嚴、個人之自律性。⑧公平性：AI 系統服務提供人、利用人、資料提供人，應注意 AI 系統或 AI 系統服務提供之判斷上包含偏見之可能性，不得對個人或特定集團為不當之差別待遇。⑨透明化原則：AI 系統服務提供人、利用人應注意對 AI 系統服務進入及輸出之檢證可能性及判斷結果之說明可能。⑩說明責任：利用人應致力對利害關係人說明 AI 系統利用之事實、被利用資料之取得或使用方法，擔保 AI 系統運作結果正確性之機制等以提升對 AI 系統技術之信賴度。[33]

（二）歐盟：

2019 年 4 月 8 日歐洲委員會（European Commission）公布「得信賴 AI 之倫理指引」（High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI）[34]，此指引主旨在為實現得信賴 AI 系統，而由利害關係人自主性遵循者。而 AI 系統之倫理原則為奠基於對基本之維護，其包含下列 4 個基本項目：①人自律性之尊重：AI 系統對人不得以非正當之從屬、強制、欺瞞、操

[33] 參閱前揭註，頁 8-9 頁。

[34] European Commission Ethics Guideline for Trustworthy AI, Ethics Guidelines for Trustworthy AI, https://www.aepd.es/sites/default/files/2019-12/ai-definition.pdf

作、附加條件或當成傀儡方式運作，應本以人為中心之設計原則，給予人有意義之選擇機會。②防止危害：AI 系統不得對人帶來危害，及負面效應。在技術上需具備堅固性，並應注意起因於資訊不對稱而所帶來之影響。③公平性：AI 系統之開發、利用應公平，此公平性包含實質上及程序上兩者，前者包括成本及利益之公正分配，確保個人及特定集團不受歧視、差別待遇、被貼標籤；後者為對 AI 系統做出決定得聲明異議及為有效之救濟。④說明責任：說明責任為 AI 系統獲取利用人信賴不可或缺者。在此原則下，建立機制與直接或間接受影響人，交換 AI 系統利用利弊進行公開對話是必要的。

對應上開倫理原則，指引並列出實現得信賴 AI 系統之七要件：①AI 系統應支援人之自律與自決之權，促進基本權之保護並得受人之監督。②具有信賴性之 AI 系統，其技術之堅牢性為不可欠缺之要件。開發 AI 系統引發不可預期風險應最小限化，防止社會所不可容忍之危害發生。③隱私權為特別受 AI 系統影響之基本權，為防止對隱私權之侵害，包含對資料之品質與完整性，必須進行完善之管理。④對資料、系統及商業模式，應予以透明化。⑤AI 系統之生命週期必須包攝性及多樣性。運用程序中考量所有受影響之利害關係人得為參與，確保其平等之近用及處理。⑥AI 系統應將社會、有感覺之生物環境亦視為利害關係人，獎勵系統之永續性及對環境之責任。⑦必須確保 AI 系統之開發、推展及利用結果之可責性及說明責任。

歐盟於 2019 年公布倫理指引後，先確立 AI 系統適當發展之方向，後繼續朝向制定具強制規範效力之規則推進，主要是鑒於中國近年施行之 AI 系統社會信用評價系統，對人精神上、身體上形成嚴重之威脅，為確保歐盟人民之人身安全及基本權，且歐盟希望對 AI 系統發展及運用上能發揮其世界主導地位，因此積極推動定立具有強制力之規範。2019 年底就任 EU 委員會主席之 Ursula von der Leyen 即表 AI 系統規則之訂立為其最重要、優先之事項；強調 AI 系統為得信賴、安全、無差別之必要性。之後於 2020 年 2 月 19 日公布 AI 白皮書（White Paper on AI—A European approach to excellence and trust），白皮書表明其政策重點

為：如何實現促進 AI 系統之運用及解決隨著技術之利用而帶來之風險兩個目標[35]。白皮書公布後，歐盟委員會，即與廣泛之利害關係進行協商。後於 2021 年 4 月 21 日歐盟委員會為實現上開第二個目標，提出值得信賴 AI 系統之法框架，以推展信賴系統。並公布 AI 系統規則草案（Artificial Intelligence Act）[36]。此提出之草案係基於歐盟之價值及基本權，目的在給予民眾及利用者對接受 AI 系統之信心，並鼓勵企業之研發推展。在草案說明備忘錄（Explanatory Memorandum）指出：AI 系統應成為人類之工具及社會向善之力量，最終目的是增進人類之幸福。故而，通用於歐盟市場上之 AI 系統或其他方式影響歐盟境內人民之 AI 系統，應以人為中心，如此人們可以相信該技術係被安全利用且符合法律，並包含了對基本權之尊重。

三、基本原則之實現

如前所述，AI 系統之開發及利用涉及許多利害關係人（Stakeholders）間錯綜複雜之利益，且須考量現今社會對 AI 系統之容受度及憲法基本權之保障等，因此在將原則法制化明確訂出規範之實現上須作多方利益之折衝及協調。

又 AI 系統之運用不僅在產業界、企業、醫療等民間，政府之行政管理、服務提供、政策規劃、犯罪預防等亦得利用為有效率之執行，為橫跨公、私部門、各行業領域之廣泛利用狀況，因各有不同之利用目的方法，而所帶來風險高低程度不一，得否不問規範對象建構一包括性、單一性法予以規範為必須面對之難題且不易解決。

再依電腦網路連結 AI 系統運用早已無國界之分，故而僅一國立法管制 AI 系統之利用必然力有未逮，必須透過國際間合作始能為法之有效執行。然因各國科技水準、經濟規模及產業發展策略均有不同，對 AI 系統管制之密度、方式必然

[35] WHITE PAPER on Artificial Intelligence-A European approach to excellence and trust, https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[36] European Commission Website “Proposal for a Regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act)”, https://digital/strategy.ec.europa:en/en/library/proposal-regulation-lay-down-harmonised-rules-artificial-intelligence-artificial-intelligence

存有差異考量，如何在國際間進行協調在法規範及執行上採較一致性之腳步亦為重要課題。

且 AI 系統之運用隨資訊科技之快速進步，至今尚難掌握所有可能發生之風險及影響之全貌，如現即急於以法律作硬性規範管制，可能因欠缺前瞻性及得因應時勢之所需做適當應對之柔軟度而無法解決問題，反而對產業等之創新形成阻礙。

基於上述之思考角度，應以如何方式實現 AI 系統之基本原則，國際間即有不同之選擇。

## 肆、日本 AI 治理（ガバナンス）（Governance）之模式

日本一直以實現其世界 IT 大國為目標，而意圖在 AI 系統發展上能發揮一定主導力，故在世界七大工業國高舉會議 G7 及 OECD 理事會中提出 AI 系統原則，主張以非具有拘束力之指針方式實現 AI 系統「治理」，並未用立法管制或規範一詞，因。所謂「治理」係用在包含公務機關或非公務機關或團體對所管事務執行之管理作用。日本之產業界多主張以軟性規範方式進行 AI 系統之規制，且日本政府基於前參、三所述基本原則上問題之考量，自始即立於由產業界自主性建立 AI 系統開發、運用之規則，機關僅依指導方式從旁予以建議、協助，即不居於主導地位而係依官、民合作方式進行 AI 系統之治理模式[37]。如此得將利用 AI 系統所產生之風險依產業團體自律方式控制於各方利害關係人可容忍之程度，最大化其所帶來之正面功能及影響，建構能顧及技術、組織面及社會實用性易於調整之 AI 系統治理制度[38]。

一、治理之基本理念

（一）軟性非跨領域指針之定立

在過往因新的技術之開發或運用，在對人類生活方式、社會環境、秩序或安

[37] 參閱前揭注 32AI ネツトワーク社会推進会議，AI 活用ガイドライソ，頁 1-2。
[38] 泉卓也，AI 原則実践のためのガバナンス・ガイドライン ver. 1.0 の概要，NBL.NO.1199，2021 年 8 月 1 號，頁 72。

全有造成危害之風險時，國家常啟動立法權制定一套法制以進行規範、管制，惟在對於 AI 系統之發展、運用，日本則選擇以軟性非橫跨各領域方式進行 AI 系統之治理。換言之，為達到 AI 系統原則之遵守、實踐以人為中心之 AI 系統社會原則，並促進 AI 系統得實用於社會，僅由經濟產業省，依 2021 年 7 月 9 日公布「AI 治理指引（ガバナンス.ガイドライン）」，AI 系統實際運用上治理規劃及實行，則委由各領域事業自主定立其 AI 系統運用上應達成之原則作為治理之目標及其執行之模式。所建構治理程序中，主要應設計出確保運用 AI 系統之網路空間及物理環境之安全性、透明化其利用目的、方法並持續評價其風險，建立利用之團體、個人均得積極參與、表達意見之機制[39]。

（二）「終極目標為基礎」之治理方式

AI 系統運用上為跨各行業領域且涉及多方利害關係人間複雜之利益，所帶來之風險性質、大小不同，難以依統一立法方式定出 AI 系統開發、利用上應遵守之原則，故由開發、利用之事業依其利用目的、方法自行宣示其 AI 系統運用最終所維護之 AI 原則作為其「終極目標（Goal Base）」，再基於此目標之達成，事業應先進行該 AI 系統開發利用有無悖離社會價值及帶來風險之評估，其次對 AI 管理系統之設計及運用評價其妥當性，並衡量社會民眾對該 AI 系統運用之容受度為何，在得持續性評價進行修正下、之機制下建構 AI 系統治理架構及運作模式。依建立在達成自主選定並宣示之終極目標基礎上執行 AI 系統治理方式，得降低一般人因 AI 系統開發利用帶來之負面影響所產生之不信任感，亦得避免事業在開發、利用 AI 系統時，僅追求 AI 系統所帶來之利益，違反應固守之社會價值而侵害人權[40]。

---

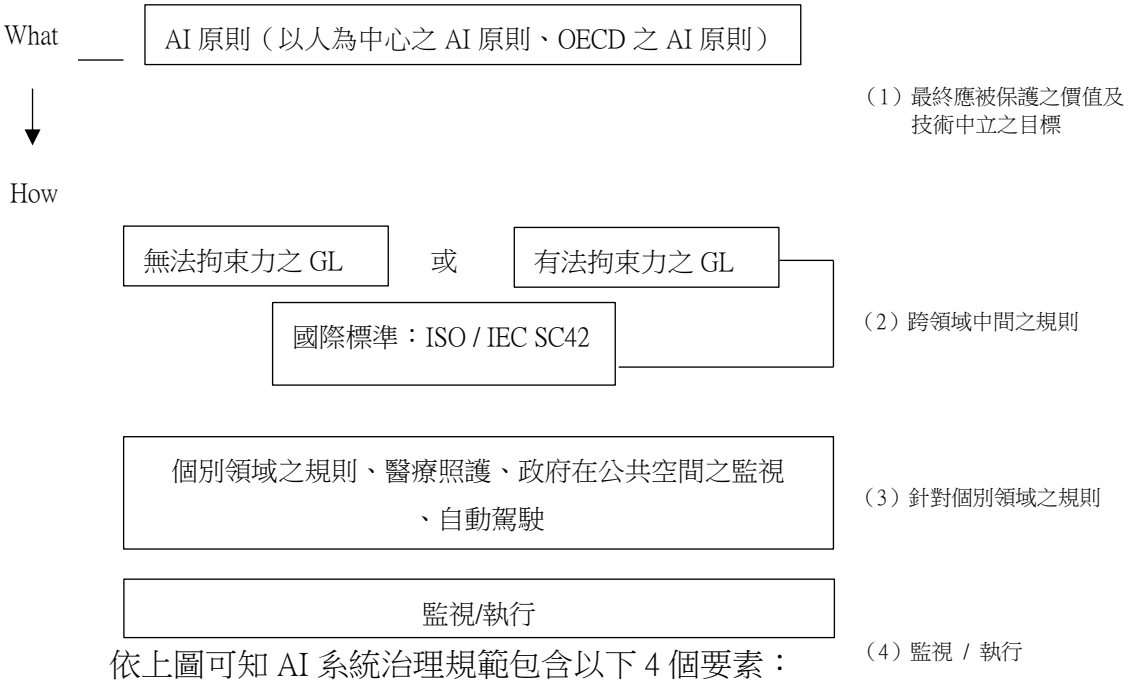
[39] 日本経済産業省，2021 年 1 月 15 日公布「我が国の AI ガバナンスの在り方 ver. 1.0」後於同年 7 月 9 日略作修正再公佈「AI 原則実践のためのガバナンス. ガイドライン ver. 1.0」，頁 22。

[40] 同前註 38，頁 76。

二、治理規範之構成[41]

日本在 AI 系統原則實踐方式檢討會報告書中，指出本文後所介紹之歐盟規則案依風險基礎方式管制，其風險之分類及評價方式未必有國際上共識，且關於實現 AI 系統原則相關之 AI 系統治理方式之討論，在各國或國際上仍處於比較保守之狀態中，未提出清楚勾勒出 AI 系統治理之架構者，因此日本進一步先具體描繪 AI 系統治理之規範構造圖，希望藉此推動此議題在國際間被重視[42]。

以下為日本 AI 治理規範構圖[43]




依上圖可知 AI 系統治理規範包含以下 4 個要素：

（一）最終應被保護之技術中立目標，此為立於 AI 系統治理規範之上位被設定最終應達成之目標，自細項說明有[44]：

1、以人為中心之社會原則：實現尊重個人及個人自律之以人為中心之社會，應以因利用 AI 系統或因其決定受影響之人為中心，保護基本權及包攝多樣性價值

[41]我が国の AI ガバナンスの在り方 ver. 1.1 AI 原則の実践の在り方に関する検討会 報告書 令和３年７月９日 AI 原則の実践の在り方に関する検討会，頁 9
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_1.pdf

[42] 同前註，頁 9。

[43] 本圖改繪自前揭註 41，頁 9 之圖。

[44] 參閱經濟產業省事務局，2022 年 2 月 8 日，AI 開発ガイドライン及び AI 利活用ガイドラインに関するレビュー，以下之說明。

觀。

2、教育、識別能力原則：為解除一般人民對 AI 系統技術之不安及喚起安全意識，因應現存及將來勞動力之課題，應對社會民眾將有關 AI 系統之利用推廣其技術或倫理之教育，在獲得充分知識資訊下做出選擇，而不受不當之影響。

3、隱私保障原則：定立之個資蒐集、保存相關規範，必須保護個人知的權利及選擇之權利。

4、安全確保原則：AI 系統之全部流程均須確保其堅牢性、安全性、可用性避免被駭客入侵，在開發之際必須先評價可能被攻擊之風險並訂出對策，對環境及利用者之安全維護，若為人與機械交互溝通時，情緒上之安全性均必須注意。

5、公平性原則：AI 系統之開發及利用應能反映利用者之多樣性及代表性，將由性別、年齡、學歷、地域、人種、宗教、國籍等而來之偏見及差別待遇最小化。商用化之 AI 系統應公平適用於所有之利用人。確保社會之弱勢族群對 AI 系統技術及服務之近用權。AI 系統所帶來之利益，應平均分配給全體之民眾而非特定族群。

6、透明性及說明責任：AI 系統在實行之可能範圍內，應對包含民眾在內之關係人，公開有關 AI 系統利用及其目的、方法等資訊以確保其透明性。並對各個 AI 系統之演算方式，儘量提供得理解之說明。透明性為指：（1）如何之資料因依何目的（支持基於 AI 系統為決定之用）而被蒐集。（2）支援演算程式作成意見決定之理由為何等資訊之公開。另外應對利用人、利害關係人說明 AI 系統運用狀況，其所作判斷之正當意義及理由。

上述之目標多是為防止前所述 AI 系統開發運用上被指摘或質疑之對人性尊嚴或人之自律、自決形成威脅或侵害，應被遵守之之 AI 原則，而由業者選擇後揭示為其 AI 系統治理上應實現之終極目標。

（二）跨領域之中間性規範[45]：

[45] 泉卓也，同前註 38。

此類規範則為不問事業類別為達成設定目標，關於 AI 系統治理上共通應有之規範或指引，例如學習資料庫之偏差或演算程式上黑盒子問題之解決，即依跨業界領域所訂立中間性之有法拘束原則或無拘束指引供各事業參照、遵循。另外，亦應注意參考國際上共通之如 ISO / IEC 等標準規範，而得與國際規範接軌。

（三）個別領域之特別規範：

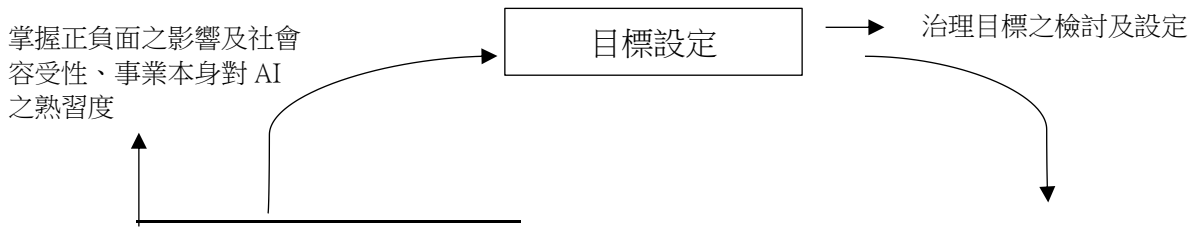
因在某些特定領域 AI 系統之開發及運用上存在其特有個別課題必須另外有特別規範，例如自駕車位置資料之蒐集運用或是設置 AI 系統人體生物特徵辨識系統對民眾之遠端即時監視等，因在實際利用上涉及隱私權或行動自由等課題，則須另訂特別之指引規範。

（四）監視及執行：

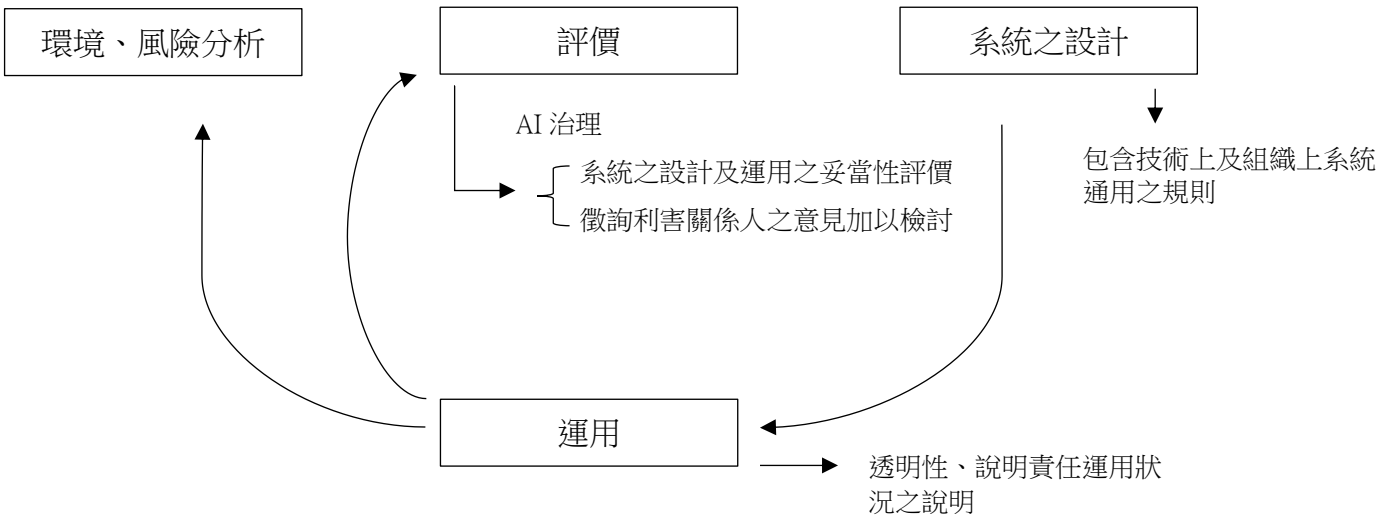
AI系統利用之事後監督，亦即在AI系統正式運用後應設置即時性監視機制，並為設定目標所採行之對應措施作出說明，透過包含政府機關在內之各方利害關係人之意見反映，進行改善與修正，並加強事業、利用人及政府間之信賴關係[46]。又在考量 AI 系統利用所帶來之社會影響或風險之可控性，整備事故發生時對事業較友善之裁判方式，例如對於無論如何均無法迴避之事故，設置原因調查及預防再度發生之特別事故調查委員會，導入延緩訴追程序或緩起訴制度，適度減輕開發利用人之責任，應為今後應考量之事項[47]。

三、AI 系統治理構成框架[48]

（一）構成之概要圖




[46] 同前註 41，頁 19。
[47] 同前註 41，頁 19。
[48] 改繪自經濟產業省，「GOVERNANCE INNOVATION Ver.2：アジャイル・ガバナンスのデザインと実装に向けて」報告書（案）経済産業省 P.55
https://www.meti.go.jp/press/2020/02/20210219003/20210219003-1.pdf




1、目標之定立

其由開發、運用之事業依其利用之目的、利用方法，所蒐集、處理個資之性質、數量，及可能引發風險之嚴重性，選定應達成之 AI 原則為目標，再依為達成設定之目標設計所必要、適當之治理系統，並於實行中隨時檢討有無悖離目標之事情。

2、運用階段之評價及說明責任

事業在運用 AI 系統過程中，對各個 AI 系統運用過程應進行監視，並持續性評價有無悖離設定目標，並記錄其結果，且確保得對外說明其治理系統運作之狀況。有關 AI 系統治理目標之設定、治理系統之整備或運用等資訊應定位為非財務資訊，積極予以公開。[49]

（二）體制之整備與實務上注意事項之結合

自上圖觀之，可知日本之 AI 系統治理運作方式包含運作體制之構成及執行上應注意事項，除適當之治理系統之設計外，在整體運用過程中進行之環境影響、

[49] 參考前揭註，頁 8。

風險之評價分析。自始即須針對治理系統設計及運用依所設定之目標作妥當性進行評價，此際，應由獨立第三人進行之，並請對 AI 系統運用狀況熟悉之有識者，如經營夥伴、消費者等利害關係人提供意見予以檢討；其後則是在治理過程中仍須對所面臨之社會環境及風險進行評價分析。此係為檢驗有無悖離 AI 系統治理設定之目標，AI 系統所帶來之正負影響為何？及 AI 系統運用之社會容受度。又因事業自身對 AI 系統運用之熟習度所形成之環境、風險亦有必要適時進行評價分析以作為改善治理之參考。換言之，繼續性於 AI 系統治理過程中進行必要之評價，為在 AI 系統治理上應具備之敏捷性[50]。

## 四、小結

日本之 AI 系統治理政策，自始至今都維持在不急於以立法作強制義務性規範以為管制，依訂立指針之行政指導方式，輔導企業自主進行 AI 系統之治理。事業自行選定應達成之 AI 系統原則為目標參酌指針設計 AI 系統治理系統依自律方式進行 AI 系統之治理，並將其 AI 系統利用目的、方式及 AI 系統治理之狀況等資訊，以透明化 AI 系統運用過程及其結果，並善盡其說明責任[51]。之所以採此治理方式，此除前所述日本之政策上考量外，另一原因，則是日本之企業文化傳統上其員工與企業間之向心力較歐美企業強[52]；且日本傳統文化之一致性及民族團結特質，在設定以「人為中心」之基本理念下為 AI 系統原則設定為目標，由官民協力訂立指引建構 AI 系統系統治理方式，得因應各事業之不同之需求且經評價後得進行調整，或者為較有彈性且符合日本社會背景之作法。

# 伍、歐盟之立法管制

## 一、背景、目的

歐盟於 2019 年公布本文前述之倫理指引後，實已確立規範 AI 系統發展及

[50] 前揭註 48，頁 75。
[51] 可參閱日本大電器製造企業日立於網頁上揭示之「倫理原則」，及其「行動準則」及「實踐項目」。
[52] 前揭註 44，頁 28。

利用之方向，但仍繼續努力朝向制定具有強制效力之規則推進，主要是鑒於中國近年施行之 AI 系統社會信用評價系統，對人精神上、身體上形成嚴重之威脅，為確保歐盟人民之人身安全及基本權；且歐盟希望對 AI 系統發展及運用上能發揮其世界主導地位，因此積極推動訂立具有強制力之規範。2019 年底就任 EU 委員會主席之 Ursula von der Leyen 即表示 AI 系統規則之訂立為其最重要、優先之事；強調 AI 系統為得信賴、安全、無差別之必要性。後於 2020 年 2 月 19 日公布了 AI 系統之白皮書（White Paper on Al-A European approach to excellence and trust），表明政策重點為：如何實現促進 AI 系統之運用及解決隨著相關技術之運用所帶來之風險兩個目標[53]。白皮書公布後，歐盟委員會，即與廣泛之利害關係人進行協商。之後 2021 年 4 月 21 日歐盟委員會為實現上開第二個目標，提出值得信賴 AI 系統之法規範框架，並公布 AI 系統規則草案 (Artificial Intelligence Act)[54]。此草案之提出係基於歐盟之價值及基本權，目的在給予人們及其他利用者接受及解決 AI 系統問題之信心，並鼓勵企業發展 AI 系統之運用。草案說明備忘錄(Explanatory Memorandom)即指出：AI 系統應成為人類之工具及社會向善之力量，最終目的是增進人類之幸福。故而，通用於歐盟市場上之 AI 系統或其他方式影響歐盟境內人民之 AI 系統應以人為中心之設計，在與歐盟既存之法令調和下，讓確保其安全性及基本權之 AI 系統治理及有效之法執行能以實現可能，而建構關於 AI 系統之包括性、統一性之法律框架。[55]

二、依風險高低為層級化之管制

本草案最受注目、影響重大之部分，為其依從 AI 系統之開發、利用所引發對人權、社會安全等之風險高低，定出寬嚴不同之管制要件及義務，此稱之為：

[53] WHITE PAPER on Artificial Intelligence-A European approach to excellence and trust, https://ee.europa.eu/info/sites/default/files/commmission-white-paper-artificial-intelligence-feb2020_en.pdf

[54] European Commission Website "Proposal for a Regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act)" , https://digital/strategy.ec.europa:en/en/library/proposal-regulation-lay-down-harmonised-rules-artificial-intelligence-artificial-intelligence

[55] 前揭註，EXPLANATORY MEMORANDUM,1.1.參照。

「風險為基礎之方法 (Risk-based approach)」[56]，亦即將 AI 開發、運用依其風險高低、程度區分為四種類型，並分別設置不同程度之規範。

如圖示

| | |
|---|---|
| 被禁止之 AI（unacceptable risk）<br>§ 5 帶來不被容許之險之 AI，例如惡用兒童或精神障礙人之脆弱性之 AI 系統。 | ← 禁止 |
| 高風險之 AI（high-risk）<br>§ 6（1）（2）例如自然人之生物特徵辨識及分類為目的之 AI 系統，或教育、職業等訓練上評價學生、面試過程中評價候選人為目的之 AI 系統。 | ← 符合一定要件或在遵守適合性之評價上而被允許者，且適當保障其透明化，實施風險評估，並由人進行監督等<br>←自主性行動規範 |
| 限定性風險 AI（Limited risk）<br>§ 52 通知與 AI 系統進行交流行動之人，其對象為 AI 系統。 | ← 確保透明化義務下許可之 AI 系統 |
| 少量低度風險 AI（minimal risk）<br>§ 69 | |

上開基於風險分類之 AI 系統類型，而屬高風險 AI 系統所被要求之合法要件，及對提供者或利用者課予之義務等，相當嚴格可謂是本草案最核心之部分。此草案一公布後，歐洲之資料保護官(EDPS)於 4 月 23 日即表示贊同之看法，特別是對公共場所即時遠距生物特徵 AI 系統辨識系統之限制，肯定其必要性。[57]另外歐洲技術產業團

[56] 前揭註，EXPLANATORY MEMORANDUM, 3.5. Fundamental rights。

[57] Artificial intelligence Act: a welcomed initiative, but ban on remote biometric identification in public space is necessary, https://edps.europa.eu/press-publications/press-news/press-releases/2021/artificial-intelligence-act-welcomed-initiative-en

體 Orgalim[58]，發出聲明表示：草案對新技術課予嚴格之規範，將阻礙產業創新之投資。對風險基礎方式雖為歡迎，然認為有必要考量法之安定性，要求應更明確化其定義，並應與業者會商，因應實際而有調整分類之餘地。[59]另外日本經濟團體連合會，亦於 2021 年 8 月 6 日公布其意見，認為歐盟制定 AI 系統草案之目的與日本追求之「得信賴高品質 AI (Trusted Quality AI）生態系統 (ecosystem）」建構及方向是一致的。但草案對禁止高風險 AI 系統之定義等上，殘留有曖昧或待解決之問題，恐降低對歐洲之投資或新興 AI 系統企業等之育成意願，對產業創新或國家安全保障造成影響之虞。於施行前應更明確化定義、作追加說明及提供指引等。另外僅對高風險 AI 系統訂出嚴格之規範，反而形成得信賴 AI 系統生態系統建立之主要阻礙，為擔保 AI 系統之適當利用，不能僅對 AI 系統提供者要求其遵守義務，更期待將 AI 系統利用之生態系全包含在內，同時加強對利用者之義務規範。[60]

三、重要之意義

自草案之備忘錄及規範之內容觀之，歐盟係基於人性尊嚴、基本權維護之原則所建構之法制，且大致亦對應本文前所述之人們對 AI 系統開發、運用所抱持之不安及質疑。例如被禁止 AI 系統即是在維護人身體、精神上之自主，避免因 AI 系統操控而扭曲人之正常行動。再如對特定人之性格特質或社會活動依 AI 系統評分方式評定其信用度，或是原則上禁止公務機關利用遠距即時生物特徵 AI 辨識系統，在公共空間進行執法，均是避免前所述的形成新弱勢族群，或全民監控社會而傷及民主政治運作。另對高風險 AI 系統之所定之要件及對提供者所課予之義務規定，亦可得知其在限制特定領域及特定目的下 AI 系統之利用，係為防止 AI 系統利用對國民生活、交通安全造成威脅。此外關於人之生物辨識及分

58 為歐洲橫跨機械工學、電器工學、電子工學及金屬技術領域，由 77 萬之企業所組成之產業團體。

59 European Regulation on Artificial Intelligence-Orgalim calls for legal clarity and workability, https://orgalim.eu/news/european-regulation-artificial-intelligence-orgalim-calls-legal-clarity-and workability

60 一般社団法人日本経済団体連合会，デジタルエコシー推進委員会，2021 年 8 月 6 日，欧州 AI 規制法案に対する意見。

類，在決定就業、受教育，重要之公共照護、給付等機會之決定上，特別規定 AI 系統利用之要件均係保障人之自決及獨立性。草案對有關高風險 AI 系統開發、運用，自第 8 至第 15 條設置其風險管理上之要件，並要求保障 AI 系統之學習資料品質，此即為回應本文前所述 AI 系統建立演算模式時因資料不完整等造成之偏見或差別待遇之質疑。另外，AI 系統作成技術說明書，設置演算紀錄之機能，提供利用者有關 AI 系統使用方法等，則是解決 AI 系統存在之「黑盒子」問題。

而又其第 16 條至第 29 條課予 AI 系統提供者之義務，則是為保障 AI 系統運作上之安全性、適合性，故要求進行適合性之評估、資料庫之登錄，系統或產品上市亦須持續進行監視。均是為防止風險發生，損害 AI 系統開發、運用過程，建立社會之信賴度所設置之規範。

不同於日本之思維，歐盟係以跨領域、建構統一具有法強制效力之法規範進行 AI 系統運用之管制。而究竟何種方式較理想，因歐盟之規則案未正式完成立法尚有諸多變數，現今一切都難有確論。但至少歐盟之政策方向已明確，其發展亦必定對其他國家或區域有重大之影響力。而且 AI 系統必然與網路連結進行跨國界之利用，其治理涉及國際間之合作與協助，歐盟與日本兩種不同之治理方式，究竟何者能為將來之主流，為往後值得關注之事。

## 陸、結語

在現今各界廣泛利用 AI 系統創新事業、研發各種新產品、提供人生活上之便捷，然在強調得以提升國家之競爭力，發展社會經濟實力，解決高齡、少子化之問題時，應不能僅看到光明面，其背後潛藏之個人資訊自主權弱化，尤以對人性尊嚴造成之威脅，甚至對民主政治造成之傷害，實值得有識者警惕與關切，應嚴肅、謹慎以對。

又我國因有國民身分證制度，許多公務機關及非公務機關長久以來慣用國民身分證號在執行職務或進行交易往來等社會活動時，做特定人身分認證之用。故而利用身分證字號為關鍵紐可串聯許多個人資料，尤其方便在網路上針對特定個

人蒐集特定人住址、工作、活動等資料。加諸大型網路平台如 Google、臉書等之利用人亦不在少數，該當業者往往在提供服務時，大量蒐集使用者之閱覽履歷、購物清單或在群組交流活動之資料，集結成個資檔案後，再利用 AI 系統進行人物剖析做信貸之評估、預測工作能力、嗜好等以決定雇用或錄取，或進行精準行銷，在成本、效率及時間之考量下，實為難以阻擋之運用趨勢。而在防恐、防疫或維護治安之理由下，利用 AI 監視攝影系統進行遠距即時生物特徵辨識之事例，亦時有所聞。凡此行為，我國現行個資法除依其第五條可能認定為未尊重當事人之權益之行為外，實欠缺實質、有效之規範。換言之，現今 AI 系統開發及運用上產生之問題，因非當年個資法立法時所得預想者，故而無法有效規範如上述之侵害個人基本權之 AI 系統之利用，更遑論針對 AI 系統學系資料庫之隱藏偏差、演算方法產生之黑盒子等問題之應對，已超出個資法能規範之行為態樣。因此針對 AI 系統之研發與運作，國家有必要立於經濟、科技產業發展之需求及保障人性尊嚴之高位，集結各方意見進行討論訂出基本方向。對此則期盼將來主導我國運用資訊技術進入數位化社會之數位發展部，或者能參照 OECD 或歐盟揭示之 AI 系統基本原則，即便不立法強制規範，亦能明確給予 AI 系統開發或利用者必要之指導，並推展或宣導 AI 系統治理概念以保障人民之權益，建立以人為本可信賴之 AI 系統社會。