AI 規範治理觀察
期末報告

主題：

從 OECD 國際合作平台
探討歐盟 AI 與臺灣治理

姓名 Name：林宏嘉

系級 Department：工程科學及海洋工程學系

學號 Student ID：B11505030

日期 Date：12/23

目錄

■ 引言

背景:

隨著人工智慧 AI 技術的快速發展，全球應用逐漸深入各行各業，從醫療、金融、教育到公共管理等領域，AI 的影響力不可忽視。然而，這些技術所帶來的創新同時也引發了各種倫理、法律及社會挑戰，特別是在如何保障人類倫理不受侵犯、維護社會公平正義與安全方面。為了有效管理 AI 的發展及市場應用性，許多國際組織都開始研究及制定 AI 的治理規範和原則，讓我們在促進技術創新的同時，減少潛在的風險與負面影響。

在課程中的最後一堂課，老師有跟我們介紹全球範圍的國際合作組織。其中經濟合作暨發展組織 OECD 就是重要組織之一，他在很早就提出了 AI 治理的「原則」，這也促進國際間的合作討論。OECD 的 AI 原則不僅涵蓋了技術創新對社會的影響，還關注如何「平衡」各方利益。最重要的也是保障「個人隱私」與數據「安全」，並推動倫理與法律框架的建立。

這次的報告希望透過 OECD 的 AI 治理原則為出發點，深入分析歐盟在 AI 治理上的實際案例，也就是最後一堂課中的歐盟 AI 法案「AI Act」作為具體的治理框架。同時探討 OECD 原則如何在全球範圍內得到實踐和落實。此外，也會融入在 16 週課堂內討論到的倫理問題來探究其中。

OECD 簡介:

想要知道為甚麼 OECD 的 AI 原則為甚麼如此重要，就要先了解這個組織。OECD 中文名稱為「經濟合作暨發展組織」，由 38 個成員國組成的國際組織主要是透過政策建議、數據分析與研究，協助各國政府應對全球性挑戰，並推動成員國之間的合作交流，台灣並非成國而是以參與方身分參加三個委員

會。去網站上也能看到人工智慧領域在重要的 Topics，可見 OECD 對 AI 發展的重視。作為重要國際經貿組織，理所當然會成為全球 AI 治理的參考框架，保障 AI 技術發展與應用的可課責性與可持續性。

■ OECD AI 原則分析

何謂 OECD AI 原則?

OECD 提出了五大原則，這也成為世界各國制定 AI 法案或政策的方針。
Values-based principle(五大原則):

1. Inclusive growth, sustainable development and well-being（**包容性增長、可持續發展與福祉**）

強調 AI 應該促進社會的包容性發展，確保它的利益能夠普及至所有社會群體，特別是弱勢群體。AI 的應用要能有助於增進公共福祉，改善生活品質，同時也要考慮長時間環境可持續性。也就是說，AI 不僅要能驅動經濟增長，還應該讓所有人都受益，避免加劇社會不平等 M 型化。

(重點:強化人類能力與創造力、促進包容性、減少不平等、保護環境)

2. Human rights and democratic values, including fairness and privacy（**人權與民主價值，包括公平與隱私**）

AI 的發展必須尊重基本人權與民主價值，包括對個人隱私的保護、公平對待每個人、以及反對任何形式的歧視。這一項原則要讓 AI 系統的設計和運用必須確保不會侵犯基本自由，最基本的平等對待、隱私保護及個人自由選擇權，確保 AI 決策不會帶有偏見與不公。

(重點:法治社會、公平性、個人隱私保護、民主價值)

3. Transparency and explainability（**透明性、可解釋性**）

要求 AI 系統應該具有高度透明度，並能夠清楚的解釋決策過程。當 AI 作出決策時，利害關係人應該要能夠理解背後的運算邏輯。當決策涉及到

影響重大決策的領域像是醫療診斷、司法判決，透明性和可解釋性就非常重要。這也是最能有助於增加大眾對 AI 技術的信任，避免黑箱操作。

(重點:決策過程透明性、可解釋性、提升演算法透明度)

4. Robustness, security and safety（**穩健性、安全性、可靠性**）

在設計和運行 AI 運算的過程中具備穩健性，能夠有效的應對各種情況和挑戰，並能夠保護它不受外部威脅或攻擊(ex:電影關鍵報告的假案件)。這不僅涉及技術層面的安全性，還包括應對意外情況的能力。AI 系統應該經過測試與驗證，確保在實際運行中能夠穩定、安全地運作。因此投入第三方驗證是可以考慮的方向之一。

(重點:穩健性、安全性、應對風險能力、可靠性)

5. Accountability（**可課責性**）

在開發與應用端必須具備很明確的問責機制，當 AI 系統的使用產生不良後果或偏差時，應該有明確的責任方來負責。無論是開發者、運營者還是使用者(利害關係人)，都應該對 AI 的結果負責，並且可以透明化追蹤其決策運算過程，確保任何負面影響都能得到及時性的回應處理。

(重點:確立責任歸屬、透明的問責機制、法律責任)

利害關係人分析

因為這五項原則保護的範圍從個人擴及到所有與系統有關的人，因此我想在這邊分析到底 AI 系統決策的利害關係人有哪些，以及一些例子。以 OECD 定義的 AI Actor 為分析方向。

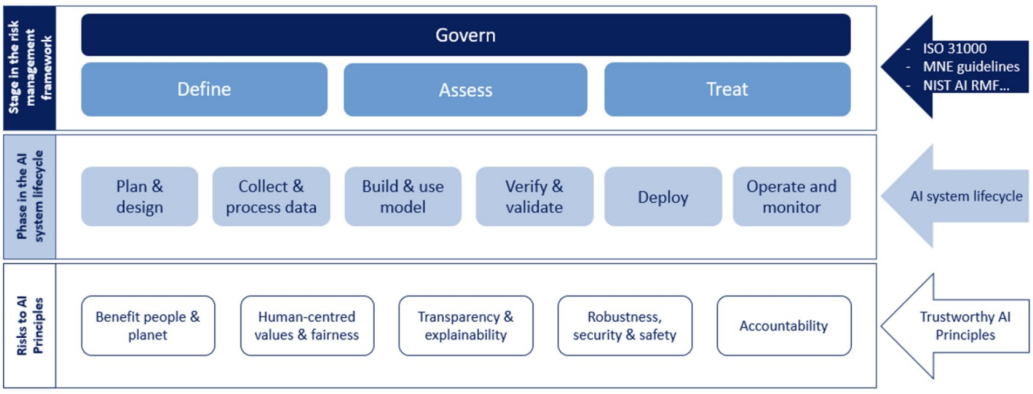

圖 1

圖 1 是 OECD 的 AI 風險管理框架以及 AI 系統生命週期和 AI 原則的風險關聯。

要探討利害關係人就會以圖中間的 AI 生命週期為切入點，上課時也是從這邊入手。AI 系統生命週期的第一個階段是「規劃與設計（Plan & Design）」：定義 AI 系統目標和架構、第二個階段是「收集與處理數據(Collect & Process Data)」：準備 AI 所需的數據，並確保數據質量與隱私保護、第三個階段是「模型建置與使用（Build & Use Model）」：訓練 AI 模型並應用於實際場景、第四個階段是「驗證與驗收（Verify & Validate）」：測試 AI 系統的準確性、安全性和穩定性。第五個階段是「部署(Deploy)」：將 AI 系統正式投入使用。最後一個階段「操作與監控（Operate & Monitor）」：持續監控 AI 系統的運作，確應對可能出現的問題。以下圖表示我整理出來的 Excel 表格。(圖 2)

生命週期	規劃與設計	收集與數據處理	模型建置與使用	驗證與驗收	部署	操作與監控
角色	開發者	數據提供者	開發者、使用者	監管機關	運營商	運營商
利益	提升技術、確保競爭力	數據準確性	提升效率、自動化、增強決策力	信任、系統準確性	大眾可用性、可擴展性	優化系統、問題解決
責任	確保系統符合公平性、透明性、隱私保護標準	確保提供數據正確、隱私性	確保模型公平、安全、準確	驗證系統安全、符合法規	確保順利實施於社會	監控系統表現、問題及時處理
舉例	Google、OpenAI	數據提供商	金融機構	政府主管機關	銀行、保險風險評估	IT技術團隊

圖 2

問題與回應

我喜歡看一件事情的正反兩面，雖然 OECD 的 AI 原則看似很完整，但其實國際社會對於它仍然抱持著正反兩面的看法。我將融入自己的看法並分別從正反兩方各舉兩例我認為比較值得討論的，特別是反面回應來說明。

正面回應:

1. 「全球參考標準的建立」

   OECD AI 原則成為世界各國制定 AI 政策的基礎(像是 AI Act)。特別是「透明性」、「問責性」等原則，為企業和政府提供了清晰的方向，有助於推動跨國合作與協調。

2. 「促進包容性與公平性」

   五大原則認為 AI 技術應帶來增長與可持續發展，並保護弱勢群體的利益。我認為這部分是科技與社會發展下取得和諧的重點。

反面回應:

1. 「原則過於抽象，缺乏具體指導」

   這一點是我認為 OECD AI 原則比較可能存在的問題也可能存在於文字僅此，我認為原則缺乏具體的技術性細節與操作指南。例如像內容中的「透明性」和「可解釋性」在實踐中如何衡量，仍缺乏很明確的標準。還有透明度要空開到多公開以及開發者保有的權利又有那些都是要進一步藉由制定細則才能去分配的。因此我也認為像歐盟訂定 AI Act 這類法規去透過法制來處理 AI 科技發展我認為仍然有其必要在現階段。因為這些技術都還在發展階段到逐漸應用，我認為越早的建立一套標準、準則能夠比原則性的討論還要有意義。但這種法規要如何項原則一樣能擴及到全球就是

我們進一步要去努力的方向。

2. 「忽略社會文化差異」

國際組織在訂定標準的同時，也需要更貼近社會各個面向。像是 AI 弱勢國家發展中國家都是這些國際組織應該要去重視的群體，如果成員國不能納入更廣泛的族群那可能會讓原則變成有歧視性的原則?不僅如此，AI 的應用涉及不同地區的文化背景與價值觀。例如，對透明性與隱私的重視程度，在歐美與亞洲國家可能有所不同，而原則也未能充分考慮這些差異。另外更嚴重的像是共產國家對人民監控的價值觀，是否會對民主國家造成威脅呢?要如何去防堵?這些都值得進一步討論。

■ 實際案例分析

歐盟 AI 法案(AI Act)

歐盟於今年提出了《人工智慧法案》(AI Act)，這項法案是全球民主國家首個針對 AI 技術的全面立法框架。目的是建立一個統一的 AI 治理體系，確保 AI 系統的使用符合安全、道德、透明且以人為本的原則。歐盟 AI 法案的核心在於將 AI 系統劃分為四個風險等級，並根據風險級別來監管和規範 AI 的開發與使用。我將以四種分類各自的定義、實例、要求規定來看這項法案。

風險分類：

1. 不可接受風險 (Unacceptable Risk)

(1)定義：

這類應用被認為對基本人權或社會價值構成嚴重威脅，因此全面禁止。而且嚴重危害個人尊嚴、安全或隱私，無法被社會接受。

(2)實例：

上課有說過的社會信用評分系統：基於個人行為進行評分，並將評分結果應用於公共或私人決策，例如貸款、健康保險審批或就業機會分配。

(3)要求與規定：

禁止這類 AI 系統的開發、銷售和使用，特定情況下可有例外（例如執法或國家安全相關的應用需獲特殊許可）。

2. 高風險（High Risk）

(1)定義：

在某些對人類安全、基本權利或福祉有重大影響的領域中應用 AI 系統。涉及關鍵決策領域，對個人或社會可能造成重大影響。須滿足嚴格的規範要求以確保系統安全可靠。

(2)實例：

在醫療方面，用於診斷或治療疾病的AI輔助診斷系統，或是在教育上用於評估學生表現或考試分數的系統(頭戴式腦波偵測器)。在面試上，用於篩選或評估求職者的演算法(AI 面試官)。

(3)要求與規定：

在數據治理方面必須使用高品質、無偏見的數據進行訓練。透明性與可解釋性需要建立一套提供完整的系統文檔和決策解釋。此外需通過第三方評估或監管機構的檢查，才能確保系統符合標準。

3. 有限風險（Limited Risk）

(1)定義：

這類系統對人類權利或社會的影響相對較小，但仍需符合基本的透明性要求。像是對個人的潛在影響小，但為了避免誤導，要滿足透明性規定。

(2)實例：

最常見的聊天機器人，用於客服人員的 AI 系統，要去明確告知使用者他是 AI 助理。

(3)規定與要求：

使用者應該要被告知他們正在與 AI 系統作對話，但是不要求進行嚴格的風險管理或監管。

4. 最小風險 (Minimal Risk)、無風險

(1)定義：

這類 AI 被視為對個人權利、社會安全或公共利益幾乎不構成風險，因此不受任何特別限制。應用範圍通常是娛樂或日常工具，影響範圍有限。不需要去額外監管或合規要求。

(2)實例：

像是我們在遊戲 APP 中的角色或動作生成系統，或是文字建議工具(電子郵件中的自動補全或拼寫檢查功能)，這些與公民權利無關之事務，大部分人工智慧的應用皆為此範圍內。

(3)規定與要求：

無需特別的透明性或風險管理措施。

歐盟 AI 法案與 OECD AI 原則

OECD 在《人工智慧原則》中提出了五大指導原則，強調 AI 技術的可持續發展、透明度、可課責性、隱私保護等方面。這些原則與歐盟 AI 法案的創建上在多方面的參考，並為歐盟 AI 法案的設計提供了重要的道德與法律框架。這邊我想要探討的是，歐盟 AI 法案在風險評估這部分參考了那些 OECD AI 原則，那五大原則究竟影響了些甚麼?

1. 包容性增長與可持續發展

OECD AI 原則：AI 應該促進包容性增長、可持續發展與公共福祉。

反觀到 AI Act 的應用上：AI Act 將 AI 應用分為風險等級，特別有關注到在醫療、教育、就業等公共福祉領域 or 人民權利的高風險 AI 系統，要求系統對社會弱勢群體的影響進行評估，確保這些群體在技術發展中不被利益衝突犧牲。甚至是鼓勵負責任的創新，確保技術發展與環境保護同時做到，像是使用 AI 減少碳排放、優化資源分配等。

2. 尊重人權與民主價值

OECD AI 原則：AI 系統應尊重人權、公平性和隱私權，避免歧視和偏見。

反觀到 AI Act 的應用上：明確禁止侵犯基本人權的 AI 應用，例如社會信用評分系統、即時生物辨識技術(共產世界監控人民)。高風險 AI 系統必須符合非歧視性要求，並確保數據集的高品質，避免因偏見導致的歧視性結果。這也是在民主國家 AI 立法上比較能看到的部分，也是台灣在未來如果要投入立法可以參考的依據。

3. 透明性與可解釋性

OECD AI 原則：AI 運作過程具有透明性，讓使用者能理解和救濟決策結果。

反觀到 AI Act 的應用上：高風險 AI 系統必須提供透明性文件，詳細說明系統的功能、限制和決策邏輯，確保使用者了解 AI 演算法運作原理。像前面所說的對那些互動型 AI（客服機器人）提出透明性要求，讓使用者知道自己正在與 AI 系統互動而非真人。

4. 穩健性、安全性與可靠性

OECD AI 原則：AI 系統應該在整個生命週期中保持穩健性和安全性，並持續評估風險。

反觀到 AI Act 的應用上：要求高風險 AI 系統需要經過嚴格的安全性測試與認證，並建立風險管理機制。規定 AI 開發者有持續監測系統運行的責任，特別是在醫療或交通這些高風險應用中。

5. 可課責性

OECD AI 原則：開發者、運營者與使用者要對 AI 系統的運行結果負責。

反觀到 AI Act 的應用上：建立監管機構和問責機制，確保系統在歐盟境內的開發與應用符合法規要求。但這部分也是因為歐盟擁有國際市場才會讓法案在實際運作上被遵守。

■ 世界 AI 法案發展與台灣的關係

在 OECD 提出 AI 五大原則後，台灣身為科技強國在 AI 發展領域也是在世界的前段班。究竟在外交處境艱難的條件下，台灣是否能參與到 AI 發展重要會議，又或是台灣要如何自己建立一套符合在地化的 AI 法案呢?就像上客說的以目前台灣立法品質來說並不合適，但如果以長遠的目標來看，我認為還是有一些方向是可以從歐盟 AI 法案當作目標。

台灣在面對 AI 法案的方向

從經濟部國貿署的資訊中能看到，台灣政府在面對 AI 發展其實也是以借鑒 OECD AI 原則為一個大方向框架，再去設立台灣特色的可信賴 AI 框架。

台灣在全球供應鏈中占據重要地位，特別是半導體與自動化工程領域。借鑒 OECD 的建議，台灣推動 AI 技術應用於提升產業效率的同時，制定法規確保環境永續與公平分配，是社會福利的一環。像是課程中說到的「以人為本的價值」，在台灣多元文化背景下，AI 政策需特別注重對原住民、弱勢族群、

無障礙使用等公平性原則，演算法不能去產生系統性歧視。例如將大眾常有的一些不良分類來建立進去 AI 資料庫，這在餵資料的要求上就要更加嚴格。

台灣在推動開放政府及透明化政策方面已經有很良好基礎，可以利用這些經驗推動 AI 原則中的可解釋性與透明度。例如，透過公開 AI 政策文件及審核程序，讓公眾理解 AI 技術的應用範疇與影響，進一步建立社會對人工智慧的信任。不僅如此，台灣在國民教育水準都偏高的情況下，也能透過公開更多 AI 演算法資訊來讓普通大眾閱讀參考。

在訂定法案的同時，我們也能去思考到上次李韶曼老師說到的主權 AI 這件事。台灣是否也要透過立法來由政府推動架設屬於自己的 AI 演算法系統呢?如果在短時間內還未開始，這樣國際間的 AI 能力兩極化只會變得越來越嚴重。因此推動台灣 AI 人才培育，針對 AI 可能對勞動市場帶來的結構改變，台灣可以參考 OECD 的建議，提供職業轉型支持與技能再訓練計畫，特別是在製造業、自動化和服務業領域。

AI 被應用在教育也是近年來各方一直提到的，但我認為這方面除了硬體與軟體在學校端做升級之外，AI 倫理與道德問題反而是下一代需要面對道的。從小就要培養起與人工智慧系統共同生活與工作的環境，這也是在技術提升的同時，人民素質也應該配合起來的重要關鍵之一。

最後就是在參與 OECD 與全球 AI 治理上，台灣應該要多做的努力。台灣可透過參與政策論壇或國際平台，分享在數位基礎建設、數據開放與 AI 應用的經驗，並透過民主國家的正當程序來切入與國際間的交流。因為在 AI 系統的運作上，我認為民主體制與共產體制會產生使用上的衝突，雙方對使用的目的、心態都將相差甚遠。這也就是台灣能進一步去參與討論的契機之一。

■ 結論

期末報告:

OECD AI 原則提供我們一個國際公認的框架，強調以人為本的 AI 發展趨勢，倡導透明、可解釋性、問責性及尊重隱私與人權等核心價值，這些原則不僅為 AI 系統的設計與應用設置了道德與法律標準，還為各國政府和企業提供了共同的治理方向。然而，全球在落實這些原則時面臨到的眾多挑戰，包括法律與監管框架的差異導致的跨國協作困難、技術快速發展引發的倫理爭議與監管空白，以及發展中國家因技術能力與資源不足進一步加劇的全球科技與經濟 M 型化不平等。台灣作為全球科技的重鎮，在 AI 技術的開發與應用上擁有很好的優勢，例如半導體產業的領導地位及豐富的數位化經驗，具備參與全球 AI 治理的實力與潛力。我們應該努力地走在領先的地位，讓 AI 不僅是貼近生活的工具，更是我們與世界對話的關鍵技術。

課程回應:

這學期的人工智慧規範治理課程與以往在智財學程的課程有很大的不同,我也是第一次上到這種有很多討論與相互回應的課程,這樣的經驗是在理工教育下比較少能體驗到的。我認為這樣的模式在學習上也讓我更能專注在聽每個人的想法，雖然不一定每一次都認同，但卻可以得到很多值得思考的答案。我也知道這個領域的規範與治理其實都還沒有很明確的答案,但也是因為這種未知性讓所有人的答案都成為了一種可能，這也是這堂課我所學到的想法。課程的安排與內容都很夠，但作業我希望可以減少一個觀影心得改成寫專家演講的心得內容。這學期除了諾貝爾和平獎那場專家演講之外其實另外兩場我都得到了很多的答案,尤其是第二場李韶曼老師的演講讓我知道了更多主權 AI 以及政府機關在資訊工程領域的欠缺驗證所導致的問題,也讓我第一次接觸到一項法律與資訊跨領域整合問題，發現到跨領域所學的重要性。

■ 引用資料

1. OECD 官方網站:https://www.oecd.org/
2. 經濟部國貿署:https://www.trade.gov.tw/Pages/List.aspx?nodeid=251
3. OECD Artificial intelligence(GPAI):https://oecd.ai/en/
4. [圖 1]OECD AI Advancing accountability in AI:
   https://oecd.ai/en/accountability
5. 科技發展觀測平台(促進人工智慧問責性):
   https://outlook.stpi.narl.org.tw/index/tdop/4b1141008683960b0186ed0a6e993443
6. 科技新南向:
   https://nsstc.narlabs.org.tw/NSTC/News.aspx?cate=2059&entry=3140
7. 資策會科法所(歐盟人工智慧法):
   https://stli.iii.org.tw/article-detail.aspx?no=64&tp=1&d=9213
8. 立法院:歐盟人工智慧法與生成是 AI 規範（楊智傑、鄭富源）
9. 中華經濟研究院(美、歐對人工智慧（AI）管理模式異同之初探）
   https://web.wtocenter.org.tw/Page/121/391300