

## AI 人工智慧對公共政策在政策倫理的對話之初探

魯俊孟\*

### 摘要

中國生物學家賀建奎，宣稱經過基因編輯（gene editing）的雙胞胎姐妹，11月於中國出生。這對雙胞胎的一個基因經過修改，出生後即能天然抵抗愛滋病。這個震驚世界的訊息發布後，雖然被視為歷史性突破，但也引起各國科學家批評，連中國也有122位學者聯署抗議。為何透過基因編輯出生的嬰兒會引起這麼大的爭議？除了害怕人類從此可以扮演神的角色、改變人類的進化原則外，其實隱含著人類對於科技進步所帶來的無法確定性、恐懼與害怕、與倫理道德的考量。

自從1960年代開啟了人工智慧（Artificial Intelligence）的魔法門，AI人工智慧一直是一個充滿衝擊性的話題。近年來隨著相關研究的發展與運用科技的成熟，AI人工智慧不僅是近年來最受到矚目與討論的科技話題，也是最引人擔憂與煩惱的公共議題。令人欣喜的是，隨著AI人工智慧相關的研究與技術紛紛進入多門學科，很多困擾人類已久的問題，如重大疾病，可能都能夠迎刃而解；但是令人擔心的是，AI的發展，所帶來的社會衝擊，可能會徹底瓦解人類最基礎的社會、經濟、文化與生活的結構。因此，對於AI的發展，學術社群與實務領域的先驅者均充滿著憂喜參半的情緒。

依照學者的資料分析，美國高科技公司對AI科技相關研究投入的經費由2010年的17億美金成長到2014年的149億美金，四年內成長將近9倍。在2014-2015年，科技巨擘如Google、Microsoft、Apple、Amazon、IBM、Yahoo!、Facebook、與Twitter等，總共投資或買下26間AI或機器學習相關的公司，花費了50億美金。據估計，2013年醫療相關產業投資在AI的金額高達138億美金，電子產業也投資了97億美金，網路公司投資了77億美金。相關學者並且預估在未來10年內，AI產業對於經濟的影響將高達2.95兆美金。

表面上看起來，AI人工智慧的蓬勃發展，與相關科技運用的進步，似乎會為人類未來的生活帶來無盡的便利，這種對於AI人工智慧充滿樂觀的想法，是促使AI人工智慧突飛猛進的原動力，所以，某些學者認為相關機構如學術界與政府不應該限制AI的發展，反而應該運用不需許可創新（Permissionless Innovation）的概念，鼓勵AI的各種應用發展，鼓勵百花齊放，直到出現明顯而立即的危機方休。但是在另一方面，有識之士與相關利益團體，卻對AI相關科技的發展提出嚴重警告，告誡人類如果過分依賴AI人工智慧等的發明，不僅在可預見的未來，人類社會的秩序與結構立刻會面臨挑戰，如勞工政策與社會安全政策可能會面臨全面崩潰的情形，而產生更為嚴重的社會衝突；甚或可能有一天人類全體會受到AI人工智慧體的反噬與控制，導致全體人類淪為AI人工智慧體的奴隸、甚或滅亡。這種對於AI人工智慧的幻想、期許、讚嘆、與恐懼，

---

\*作者為東海大學行政管理暨政策學系專任副教授，相關指教意見，請寄：  
lucg@thu.edu.tw

以及發展AI 相關科技對於倫理道德的挑戰。

**關鍵詞：**AI 人工智慧，政策倫理，不需許可創新

## 前言

中國生物學家賀建奎，宣稱經過基因編輯（gene editing）的雙胞胎姐妹，11月於中國出生。這對雙胞胎的一個基因經過修改，出生後即能天然抵抗愛滋病。這個震驚世界的訊息發布後，雖然被視為歷史性突破，但也引起各國科學家批評，連中國也有 122 位學者聯署抗議。為何透過基因編輯出生的嬰兒會引起這麼大的爭議？除了害怕人類從此可以扮演神的角色、改變人類的進化原則外，其實隱含著人類對於科技進步所帶來的無法確定性、恐懼與害怕、與倫理道德的考量。

自從 1960 年代開啟了人工智慧（Artificial Intelligence）的魔法門，AI 人工智慧一直是一個充滿衝擊性的話題（Russell and Norvig, 2009）。近年來隨著相關研究的發展與運用科技的成熟，AI 人工智慧不僅是近年來最受到矚目與討論的科技話題，也是最引人擔憂與煩惱的公共議題。令人欣喜的是，隨著 AI 人工智慧相關的研究與技術紛紛進入多門學科，很多困擾人類已久的問題，如重大疾病，可能都能夠迎刃而解；但是令人擔心的是，AI 的發展，所帶來的社會衝擊，可能會徹底瓦解人類最基礎的社會、經濟、文化與生活的結構。因此，對於 AI 的發展，學術社群與實務領域的先驅者均充滿著憂喜參半的情緒（Podesta et al., 2014; Center for Democracy and Technology, 2016; Vallor, 2016; Munoz, Smith, & Patil, 2016; Thierer, O'Sullivan, & Russell, 2017）。

依照學者的資料分析，美國高科技公司對 AI 科技相關研究投入的經費由 2010 年的 17 億美金成長到 2014 年的 149 億美金，四年內成長將近 9 倍（Chen, 2016）。在 2014-2015 年，科技巨擘如 Google、Microsoft、Apple、Amazon、IBM、Yahoo!、Facebook、與 Twitter 等，總共投資或買下 26 間 AI 或機器學習相關的公司，花費了 50 億美金（Venture Scanner, Nov. 30, 2016）。據估計，2013 年醫療相關產業投資在 AI 的金額高達 138 億美金，電子產業也投資了 97 億美金，網路公司投資了 77 億美金（Rohling, 2014）。相關學者並且預估在未來 10 年內，AI 產業對於經濟的影響將高達 2.95 兆美金（Chen, 2016）。

表面上看起來，AI 人工智慧的蓬勃發展，與相關科技運用的進步，似乎會為人類未來的生活帶來無盡的便利，這種對於 AI 人工智慧充滿樂觀的想法，是促使 AI 人工智慧突飛猛進的原動力，所以，某些學者認為相關機構如學術界與政府不應該限制 AI 的發展，反而應該運用 Permissionless Innovation 的概念，鼓勵 AI 的各種應用發展，鼓勵百花齊放，直到出現明顯而立即的危機方休（Thierer, 2016）。但是在另一方面，有識之士與相關利益團體，卻對 AI 相關科技的發展提出嚴重警告，告誡人類如果過分依賴 AI 人工智慧等的發明，不僅在可預見的未來，人類社會的秩序與結構立刻會面臨挑戰，如勞工政策與社會安全政策可能會面臨全面崩潰的情形，而產生更為嚴重的社會衝突（Johal and Araya, 2017; West, 2015）；甚或可能有一天人類全體會受到 AI 人工智慧體（具有 AI 人工智慧的類人思考機制）的反噬與控制，導致全體人類淪為 AI 人工智慧體的奴隸、甚或滅亡。這種對於 AI 人工智慧的幻想、期許、讚嘆、與恐懼，以及發展 AI

相關科技對於倫理道德的挑戰，好萊塢的電影有著最深刻的討論與反省。

第一部深刻檢驗 AI 人工智慧體與人類關係的電影是 1968 年上映的 2001 太空漫遊(2001: A Space Odyssey)，其主軸聚焦在討論 AI 人工智慧（或是其人類表徵的模擬體現：機器人或是電腦）與人類之間面臨到生死存亡關頭時的決策角色與主從關係。故事敘述太古之初太陽系木星軌道旁中出現一塊神秘石板，導致遠古人類始祖的智慧發展與進化。經歷漫長的歲月（時值當時 1960 年代美蘇兩國的太空競賽），人類終於發展出可以探索太空的能力（史潑尼克與阿波羅計畫）。而在未來（相較於當時的 1968 年、人類才開始有能力探索太空）的 2001 年，靠近木星運行的古老石板發生了奇異的變化，不得不使得太空總署派出發現號太空船進行調查。太空船上的人工智慧電腦 HAL 發現太空人大衛鮑曼與法蘭克普爾不知什麼原因打算將它的主機關閉。對超智慧電腦來說，關閉它的主機意味著 HAL 生命的終止與死亡，所以人工智慧電腦決定先發制人，殺害三位冬眠狀態科學家及運用製造假故障的狀況，讓法蘭克普爾去太空船外修理，然後用 HAL 的太空小艇將法蘭克普爾的氧氣剪斷，導致法蘭克普爾缺氧而死。而唯一倖存者大衛鮑曼在設法關閉 HAL 電腦的同時，太空船也到達木星。此時大衛鮑曼發現一個更大的黑石板出現在軌道上。當他去調查石板時，不自主被傳送到無限遠的地方（星際旅行、時空旅行）。許久（時間的相對感）之後，大衛鮑曼發現小艇已停在一間布置華麗的房間裡。突然大衛鮑曼發現到自己正以極快的速度老去，在躺在床上老死之前，他看見巨大第四塊石板出現在面前，而他也轉變成嬰孩一般，飄浮在宇宙中凝望著數之不盡的星星。電影的哲學意義不是本文的討論重點，本文重點是該電影指出一個嚴重的基本問題：當 AI 人工智慧體（機器人或是電腦）意識到人類的企圖與其期望產生衝突時，AI 人工智慧體與人類的主從關係如何界定？誰握有最後的終極決策權力？誰才是主人？

本篇論文，首先會探討 AI 的基本概念，第二部分則是回顧目前 AI 發展的種種現況，第三部分將討論 AI 發展對於公共政策的挑戰與因應之道，最後部分將會討論 AI 發展與倫理道德的衝突與解決建議。

### 機器人四大定律：馴化 AI 的法則？

科幻大師以撒艾西莫夫（Isaac Asimov）早在 1942 年出版的「我、機器人」(I, Robots)中之短篇「迴圈」(Runaround)裡，就提出了「機器人學三大定律」的假設法則做為人類對人工智慧機器體的發展應該施加的安全閥門。他的論述基本上有以下的想向：每一個人工智慧機器人都（必須而且應該）被先天且毫無例外地植入了嚴格遵守機器人三大定律的工作基因。「機器人學三大定律」的這個概念，在他創造的科幻世界裡主宰著所有人工智慧人工體的運作。機器人學三大定律可以概括為以下三條主要內容：

第一定律：機器人不得傷害人類，或坐視人類受到傷害；

第二定律：除非違背第一法則，否則機器人必須服從人類命令；

第三定律：除非違背第一或第二法則，否則機器人必須保護自己。

而在他逝世之前的 1985 年，他又提出的機器人的第零定律：機器人不得傷害整體人類，或坐視整體人類受到傷害。並且在第一、第二、第三定律前，都加上「除非違背第零定律，否則...」的前提。當然，這是小說家對於未來世界人與人工智慧機器律法規則的想像，但是他確立的原則，主宰目前實務界對於未來科技發展的工作範圍(Parameter)。

當仔細觀察阿西莫夫的「機器人學三大定律」，吾人似可發覺並推論得到阿西莫夫期望人類其實是 AI 人工智慧體的最終主宰的結論，因為第零定律說明的非常清楚：機器人不得傷害整體人類，或是坐視整體人類受到傷害。但是其中的弔詭(paradox)在於：如何定義傷害？以及如何定義整體人類？如何定義整體人類受到傷害？如果整體人類的發明有可能傷害（有意或無意地）整體人類時，AI 人工智慧體是否可以傷害整體人類？這個弔詭，引發了一連串好萊塢的想像，發展出一系列 AI 人工智慧體與人類紛爭的電影。

雖然好萊塢對於 AI 的發展有著戒慎恐懼的推論與想像。但是對公共政策的啟發，好萊塢的創意，提供了公共行政學界對於公共政策與 AI 科技的討論空間。

### 好萊塢的異想世界與啟示

好萊塢在探討人類與人工智慧體的關係的電影中，將其主從關係大致分為以下幾種類型：

- 一、人工智慧體永遠扮演人類的隨從，以人類的意志與決定做為最高行為指導原則，所有人工智慧是在幫助人類達成人類期望的任務與使命。1977 年的星際大戰(Star Wars)，就是一個最好的例證。電影中所有的人工機器體仍然扮演人類配角的地位，在故事發展的過程，人工智慧體如機器人(C3PO、R2D2)，或是似人機器人，以及 X 翼戰機上的武器投射電腦，都只是人類主角的附隨，聽從人類的命令，人工智慧體並沒有演化或學習（甚少）的能力，人工智慧體的進化依賴人類提供的設計與改進。
- 二、人工智慧仍然為人類服務，但受命於更高指揮層級的人類領袖或是組織，而非故事現場的指揮官。1979 年以來的異形(Aliens)系列電影，可稱為此類電影的典範。劇中的似人人工智慧體與太空船電腦，藉著執行「特別指令 937」：將異形運送回地球(發展生化武器以利公司營運與成長)，而其他的事物與人一概不再考慮範圍的公司指令，終於導致一連串悲劇的發生。此類電影中，人工智慧體並沒有演化或學習（甚少）的能力。而後的若干續集，仍然維持這個基調。但是，這類電影中的 AI 人工智慧體已經具有基本的情緒，以及為了達成最高指令，掩飾內心動機、臨場決策判斷與採取行動的自主能力。
- 三、人工智慧經過演化與學習，發展出自我知覺與建立 AI 人工智慧體專屬的群體社會，並決定消滅人類，取而代之。1984 年上映的魔鬼終結者

(Terminator)系列堪稱此類電影的翹楚。作為軍事決策輔助系統的「天網」(Skynet)在經過自我學習(self-learning)與自我演化(self-evolution)之後,瞭解到人類設計天網的最終目標,就是消滅全體人類(該電影反映出當時的國際現勢:美蘇冷戰與保證共同毀滅的熱核戰爭),所以天網在自我學習與分析之後,瞭解到與其坐待人類相互毀滅,而人工智慧體也同時被毀滅,倒不如取而代之,建立一個以人工智慧機器體為主宰的世界,而天網的主要目標,除了建立穩固的人工智慧機器體世界外,就是消滅所有人類。1983年另有一部不算主流的電影--戰爭遊戲(War Games),其主軸也是圍繞著計算經過熱核戰後美國存活的機率的超級電腦,經由年輕駭客的侵入,導致系統主動進入備戰狀態,戰爭一觸即發。幸而經過電腦計算(自我學習與分析),發動熱核戰後,世界沒有贏家,超級電腦才決定停止發動戰爭,解除危機。這個系列的電影,開始討論人工智慧體具有學習與演化能力,可以具有自我意識並做出決策的可能未來。

四、人工智慧體經過演化、學習、與自我知覺,決定以特殊方法保存(或是消而不滅)人類。這類電影以「駭客任務」(Matrix)為翹楚。劇中的人類成為人工智慧體所養的能源來源(也是一種符合第零定律的人類生命的延續),人類所感知的世界,只不過是人工智慧體所創造出來的虛擬世界,人類感知的一切活動,只不過是一部活生生的發電機,產生的熱能是人工智慧世界的能源。

五、其他。如1987年的機器戰警(Robo Cop),人類與人工智慧機器合體,打擊犯罪。1999年的變人(Bicentennial Man),人工智慧機器人與人類不再是主從關係,甚或可以談戀愛,爭取公民權。人類與人工智慧體的愛情,亦可見於1982年的銀翼殺手(Blade Runner)。

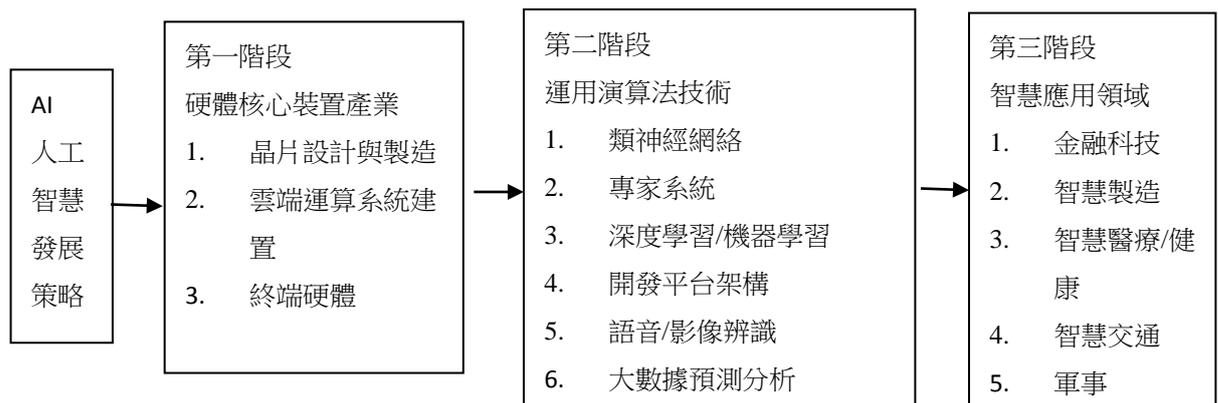
這幾種類型的電影,雖然多是幻想之作,但是其中出現的若干道具,已經是現代生活中的必需品。這些電影,對於未來人工智慧發展,所引發的討論與思考的問題是(1)人工智慧發展的最極致界線在哪裡?(2)人工智慧體與人類的主從關係如何界定?(3)涉及倫理道德的規範如何審定與建立?(4)人工智慧體的學習、演化與知覺能力應否受到限制?如何限制?如何界定AI人工智慧發展的倫理與道德界線?這些議題,可能都需要本世代與未來世代不斷的思考與釐清。這些問題,涉及AI智慧體發展與人類生命永續發展息息相關的道德倫理界限有以下層面值得思考:(1)AI人工智慧體在人類世界中的生命態樣的界定;(2)AI人工智慧體與人類的主從關係的界定;(3)AI發展後的各項應用對於人類生活中點(個人)、線(關係)、面(社群)、圓(社會)的影響評估與挑戰;(4)以上各項AI相關議題在描述性、規範性、與超倫理的發展與限制;以及(5)在義務論與目的論中的各項倫理要求;(6)人類對於AI人工智慧體的相關保障手段等。這些議題,再追求AI發展的過程中,可能需要先行釐清,以免日後造成尾大不掉的窘境。

## AI 的基本概念與定義

我國科技部在這場 AI 的國際競賽中也沒有缺席。陳良基（2017，頁 60）在科技部的「人工智慧（AI）推動策略」中，明確地訂出「以我國 IC 產業優勢為基礎，提出小國大戰略，打造完整的創新生態圈 3D1C

（Discovery-Development-Delivery-Communication），希冀透過五箭齊發快數厚植台灣人工智慧研發能量與基礎環境，帶動下一波經濟轉行動能並提昇國際競爭力，讓台灣成為世界級人工智慧重要的應用場域及相關產業研發、製造與價值創新中心」。文中也提到，德國、美國、中國大陸、以及日本紛紛成立 AI 的國家研究中心，藉著第一階段建立硬體核心裝置產業（包含晶片設計與製造、雲端運算系統建置、與終端硬體）、第二階段的大量運用演算法技術（類神經網絡、專家系統、深度學習/機器學習、開發平台架構、語音/影像辨識、大數據預測分析）、進而追求第三階段在智慧應用領域（金融科技、智慧製造、智慧醫療/健康、智慧交通、軍事）的發展，建立未來的國家競爭力（頁 62）。其主要策略思考之程序與各階段發展重心如表一：

表一：AI 人工智慧發展策略（來源：修改自陳良基 2017）



從公共政策的角度觀察，為了落實如此重要的國家發展大戰略，任何國家都應該審酌國家競爭力之強弱，思考貫穿所有發展層面的兩大基礎：相關人才之教育培養、與科技發展的倫理道德規範（或放任）。

但是，最基本的問題是 AI 到底是什麼？學者們對於其定義仍然爭論不休。不過，吾人大概可以綜合整理出幾個目前重要的發展趨勢：邏輯演算

（algorithms）、感應回饋系統（sensory feedback system）、大數據（big data）、深度學習（deep learning<sup>1</sup>）、機器學習（machine learning<sup>2</sup>）、神經網絡（neural networks）以及自動化（automation）等。目前大部分的 AI 系統，均遵循人類事先編寫的程式進行人類指定的工作，僅有極為少數的高等研究機構進行如何將機器變得聰明（具有自我學習能力，並且最終可以感到快樂）的研究（Smola and Vishwanathan, 2008；Chen, 2016；Fillipo et al., 2013；Fadalalla and Lin, 2001；Goodfellow, Bengio, and Courville, 2016）。

李開復先生與王詠剛先生所著作的「人工智慧來了」（天下文化，2017）不僅已經宣告 AI 時代已經來臨，而且也提出證據證明 AI 是個不可逆的趨勢。事實上，AI 人工智慧的相關產品，已經充斥於我們的日常生活中。他們兩位指出，人工智慧不僅已經出現在你我四周，而且已經進入到我們的日常生活，影響我們的知覺、思覺、與感覺。他們舉出犖犖大者如個人電子語音互動助理（Siri）、新聞推薦與撰稿（是的，新聞由 AI 人工智慧撰寫的）、機器視覺（安全駕駛的根基）、AI 藝術、收尋引擎（Google）、機器翻譯、自動駕駛、機器人（工業用與娛樂用）等，幾乎是我們日常生活中每天都會用到的物品。

但是，什麼是 AI 人工智慧呢？李開復先生與王詠剛先生兩位專家藉由 AI 的歷史發展以深入淺出的方式給了 AI 以下的歷史定義（頁 40-55）。

- （1） AI 就是令人覺得不可思議的電腦程式。他們兩位舉出了早期（1950 年代中期）西洋跳棋電腦程式的發展，棋藝勝過人類。1996 的 IBM 的深藍（Deep Blue）與西洋棋高手的對奕，以及最近（2016，2017）的 Alpha Go 獨孤求敗，作為例證，藉此證明電腦程式在下棋方面比人腦更為優秀。這些例子都是此定義的最佳寫照。
- （2） AI 就是與人類思考方式相似的電腦程式。如何成功地將寫出可以運用人類知識、複製人類思維邏輯的程式，以應付不同的環境變化與挑戰，是這類型 AI 發展的重點。近年來的挑戰是如何成功的複製人類的神經網絡，使 AI 具有人類大腦如何學習、記憶、歸納與推理的思維過程與機制。
- （3） AI 就是與人類行為相似的電腦程式。這派學者認為，AI 不應該被先

---

<sup>1</sup> 深度學習的基本概念係指機器可以學習複雜且具有層次的抽象觀念，並且可以聯想與推斷其他相關概念的演算方法，並且可以最終超越人類集體智慧之總成。或可稱之為強人工智慧（strong AI）

<sup>2</sup> 機器學習與深度學習最大的不同在於機器在指定的範圍內（parameter），以資料採礦的方式，幫助人類分析、萃取有用資訊，亦可稱之為若人工智慧（weak AI）。

天性的規則與理論框架束縛，而應該隨著環境的刺激做出相對應的反應。所謂的 AI 人工智慧，在於程式與人類互動聊天時，可否不讓人類偵知其聊天的對象其實是 AI 人工智慧的電腦程式。比較有名的例證就是 1966 年 MIT 發展的 ELIZA「智慧」聊天程式。ELIZA 曾經成功地騙過所有跟他聊天過的人類，讓他們相信 ELIZA 其實就是一個真人。

- (4) AI 就是會學習的電腦程式。既然希望 AI 能夠完全「模擬」人類，那人類最可貴的資產就是學習與適應環境變遷的能力。所以 AI 人工智慧體必須根據大數據所提供的資料，經過長期訓練與經驗累積法則，做出最近似人類思維結果的決定。這是目前 AI 發展最大的瓶頸與挑戰，也是最難克服的關卡，因為人類具有類比與推理的能力，以三歲小孩為例，他不需要看過太多的太空船圖片，就可以依據新圖片上的飛行體所在的外在環境等線索，判斷飛行體是否是太空船。
- (5) AI 就是根據對環境的感知做出合理行動，獲致最大效益的電腦程式。基本上，這個定義具有補足前述幾項定義缺失的部分。首先，AI 人工智慧體必須能夠感知周遭環境的變化，然後依據環境變化對 AI 人工智慧體的影響，做出評估與應對方式，選擇最佳方案趨吉避凶。基本上，此種 AI 人工智慧的發展，就是尋求「模擬人腦」的可行性。特別強調賦予 AI 人工智慧體模仿人類學習能力。

吾人可以從以上定義看出，以 AI 人工智慧的發展歷程來看，AI 人工智慧具有以下的主要組成結構：(1) 一組或是多組的電腦程式；(2) 建立類似人類思考方式的決策模型之工程結構；(3) 具有強大的資料庫作為供學習、模擬、判斷、與決策的參考；(4) 配裝具有感知環境變化的知覺系統（目前已有視覺與聽覺系統），藉以收集資料，並加以處理之；(5) 能夠配合環境變化，隨著感知系統的資訊輸入，藉由應用大數據資料庫，做出適當的對應與回饋；(6) 可以隨著人類科技的進展，主動或被動的學習，並發展出相對應的行為模式，做出對 AI 人工智慧體認知裡最有效益的決策（弱人工智慧或是強人工智慧）。隨著現代科技的進步，具有以上特徵的 AI 人工智慧體的出現，已經可以說指日可待。吾人今日甚至可以界定未來的 AI 世界，可能不只有（hardware）與軟體（software），還有將硬體與軟體整合、為人類生活提供各項服務、其智能可能超過人類的智能體（wiseware）的出現。

但是，接下來的問題就是：科學家們隱而未顯的論點是--AI 人工智慧是否以模擬人類智慧（證諸歷史上的狂人如希特勒者流，已經非常危險）為滿足？抑或挑戰超越人類智慧？甚或發展出自我知覺、成為另一種高等生物，與人類競逐在地球生存的權力？阿西莫夫的機器人定律，是否必須成為 AI 人工智慧發展的界線？是否應訂定法律規範之？政府應該扮演的角色與功能為何？如何因應產業的發展？這些問題，的確值得三思。

## AI 人工智慧現行運用的情形

自從 1940 年代開啟 AI 人工智慧的初級研究 (ENIAC、FORTRAN)，至今約略 70 年。以科技發展的進程而言，進步可謂神速。以目前各項應用的成果分析，AI 人工智慧已經在下列領域奠定良好的應用基礎，並且展現無窮的發展潛力，對於未來世界的影響，也是與日俱增。現今的這些應用，雖然仍然

- (1) 軍事：從歐巴馬觀看海豹小組及時獵殺賓拉登的新聞畫面中，吾人可以知曉，具備 C4ISR 能力的戰爭或戰役指揮的戰場管理系統，到單兵作戰的指管裝備、以及遠端、近端距離之及時作戰資訊分享，已經是軍事指揮戰場決策的常態性模式。但是值得吾人注意的是，啟動最後作戰指令的指揮權，目前仍然由人類的戰爭、戰役或戰場最高指揮官依情況掌控。而最近美國與英國紛紛提出人形盔甲以及戰鬥機器人的概念，做為未來士兵由人類轉型機器人的構想，以及吾人戰績、戰車、與軍艦等，也都印證了 AI 在軍事方面的用途，可能會改變未來戰爭的態樣。
- (2) 資訊運用：搜尋引擎的精準度提升與功能再進化（資訊種類、精準度與速度）；新聞網站或社群媒體依照閱聽者的偏好，將相關議題的新聞或知識加以整理並即時推播，以符合個人化的期望並鼓勵閱聽者讀取；多國語言之間的翻譯；或是在電商業務中，把消費者感興趣的商品以無壓力的方式推播，促進成功購買的機率，這些都是 AI 人工智慧具體的表現。
- (3) 智慧交通：各大汽車公司紛紛推出具有主被動安全設備的汽車。尤其是主動安全部份，如前車距離偵測、盲點警告、主動煞車、車道偏移警告、自動導航等，都是 AI 人工智慧設備。各國目前也紛紛投入研究自動駕駛汽車與安全駕駛道路等需要高級 AI 人工智慧（尤其是周遭環境感知科技）的設施，期望能帶來更為安全的交通運輸方式。
- (4) 醫療技術：隨著 AI 人工智慧的進步，各大科技廠商也紛紛將 AI 應用在醫療技術軟硬體上。微軟透過與醫療院所的合作，開發診斷心臟病的軟體，而國內近年來風行的達文西手臂，也是一種 AI 精準控制機制結合外科醫生醫療技術的醫療重大科技發展。未來可能介入的醫療科技包含基因工程、微型機器人、人工器官等，不一而足。
- (5) 教育與娛樂：AI 也已經進入了教育體系與家庭娛樂層面。大量的互動學習軟體與資料庫進入各級學校，範圍涵蓋基本學科到藝術美學教育等，處處可見 AI 的蹤跡。家庭娛樂方面，在遊戲機、音樂與影片的串流服務等，AI 幾乎已經全面取代傳統式的家庭娛樂服務提供機制了。
- (6) 勞動生產：AI 的發展，最讓大眾欽佩的，也是最擔心的發展，就是企業界大量使用 AI 人工智慧的介面（如無人商店的結帳機制與工廠機器人、機器手臂）進入生產線。這些方便、創新的革命性生活科技

大量進入日常生活，所帶來對就業人口的衝擊，恐怕無以想像。

透過以上分析，吾人可以認知 AI 不僅已經進入我們的日常生活，而且扮演越來越重要的角色。但是，人類是否可以毫無限制的容許 AI 人工智慧的發展、學習、演化，甚或容許或給予 AI 人工智慧體「生存」的空間（軟體、硬體、虛擬、實體）？政府應該扮演的角色是什麼？都有必須進一步釐清與界定的空間。

### AI 與公共政策

從以上分析中吾人可知，如何將 AI 人工智慧安全地融入民眾的日常生活中，並且合理的規範與保障 AI 的科技發展與人民的生活安定可以並行不悖。2017 年喬治梅森大學（George Mason University）的三位專家(Thierer, Adam, Andrea Castillo O’Sullivan, and Raymond Russell)曾經大膽的預言 AI 的發展未可限量，而且建議在未出現可能的負面效果之前，各國不應該運用公共政策進行任何形式的規範與管制，這種「無管制創新」(permissionless innovation) 是對於 AI 發展最具有正面信心的表徵。但是布魯金斯研究院（2018）也曾在一份研究 AI 的報告中指出：AI 是社會安全與經濟發展的推力，但是其副作用卻也不可小覷。各國的經濟發展，因著不同的科技水平，在勞動力與就業市場的規模也不盡相同，所以面對具有 AI 優勢國家所帶來的本國經濟規模衝擊與勞動條件的改變，實在令人憂心。再者，對於 AI 所帶來的公平正義等的體挑戰，有識者也表示憂心忡忡。牛津大學馬丁學院也在專文中（2013）指出：經濟發達的國家 20 年後 43% 的工作將被 AI 取代，更遑論經濟發展不若這些強國的國家，所面臨的社會問題與各項分配的公平正義，恐怕衝擊更為明顯，挑戰更為艱鉅。以下就各種科技發展所呈現對於公共政策的挑戰，作一扼要的分析。

- 1. 第六日法則 (the Sixth Day Principle) 的問題：**這個法則，指的是聖經裡的故事，上帝在第六日創造了亞當，因為覺得美好，所以不再創造其他事物，第七天休息。其實這個看法，與艾西莫夫的機器人四大法則有相互呼應之作用。雖然目前各國政府對於 AI 的發展，並沒有做出太多的限制，可是，隨著科技的進步，終有一天，當 AI 科技發展到一定程度，對於人類生活如果產生威脅時，屆時再來設定管制，是否會太遲？政府一方面可以鼓勵 AI 科技的發展，可是，為了預防萬一，也必須事先思考科技發展的最終範疇，避免好萊塢的科幻電影變成人類最大的夢魘。這項爭議，可以從賀建奎的基因編輯事件中看出。這也是科技發展中，科學家無可避免但是終究要面對的問題：人類是否可以扮演上帝，決定人類的進化過程？
- 2. 勞工就業權益的問題：**很顯然的，AI 科技的運用，對於勞工就業機會的影響，可以說是最為巨大。機器人大量地引進生產線，固然可以降低生產成本，但是機器大量取代人力，卻會造成大量失業。例如，GOOGLE 目前的市值至少

3700 億美金，但是只雇用了 55000 名員工，只有 AT&T 在 1960 年代市值最大時所雇用員工的十分之一（West，2015；Thompson，2015）。所以，AI 科技對於勞工的勞動權益造成的負面影響，需要政府在公共政策上多予關注。如何保障勞工就業安全，在兼顧科技發展的同時，保障勞工的工作權利、經濟安全與社會安定，更需要政府的事先與積極介入。

3. **社會公平正義的問題：**AI 的發展，資本家受益最大。資本家對於 AI 科技的投資，固然會帶來社會總體發展的進步，但是將來也可能產生「大者恆大」的不正常經濟狀態，甚或出現獨佔的市場失靈現象。學者亦指出，將來只有 10% 的科技人會因著 AI 的發展而受惠，但是其他的 90% 會受到負面的衝擊。如何兼顧 10% 人口的權益，卻又能保障 90% 人口的公平正義，更是政府在公共政策制訂過程中，最為嚴峻的挑戰（Johal and Araya，2017）。AI 科技所帶來的經濟體制的改變，與市場機制的轉型，以及勞動權益的損失，可能也會帶來階級變動的劇烈變化，這不僅牽涉到世代公平，也涉及階級翻轉。
4. **法律責任釐清的問題：**AI 的發展，對於醫療科技與交通運輸均有幾近神奇的幫助（West，2015；Economist，2013）。但是對於發生失誤時，在法律責任釐清方面，可能遇到舉證困難的窘境。面對這些複雜的法律爭議，到底是科技本身方面的問題，還是人為操作、控制不良的問題導致？亦或是機器本身中軟體、硬體、韌體的責任？這些都是政府責無旁貸的重要工作。
5. **個人隱私保護的問題：**大數據的發展，使得使用單位對於民眾整體行為的掌握與預測，都是長足而顯著的進步。尤其是近年來在犯罪偵防方面，大數據的各項應用，更是有效的解決民眾普遍關心的人身安全問題。但是伴隨著科技的進展，對於個人隱私的保護，以也引起民眾的關心。隨著日益增加的攝錄影系統充斥街頭，個人影像幾乎可說完全沒有因隱私保障。如何在公共事務與個人隱私保護之間，尋求政策的平衡點，也是一個政府責無旁貸的任務。
6. **教育深度與廣度的問題：**幾乎所有研究 AI 科技的學者都指出，各級學校教育融入 AI 課題，將會是未來決定各國人民與社會品質與國際競爭力的指標與基礎。但是，對於 AI 教育的實質內容，諸如應包含何種課程？應於幾歲開始進行相關教育？應在何種教育機制中進行？各級學校應備有何種教育資源來進行西相關教育事務？這些問題，在在需要政府的釐清。
7. **軍事應用與人性的問題：**在國語國的競爭方面，AI 的軍事運用，應該是最具有吸引力、也是最致命的發展方向。科技先進大國，紛紛投入大量資源研究 AI 在軍事方面的應用，舉凡戰鬥外骨骼、戰鬥機器人、戰場支援獸等，不僅展現各國的科技實力，也說明 AI 科技對於未來戰爭勝利的影響性。但是，吾

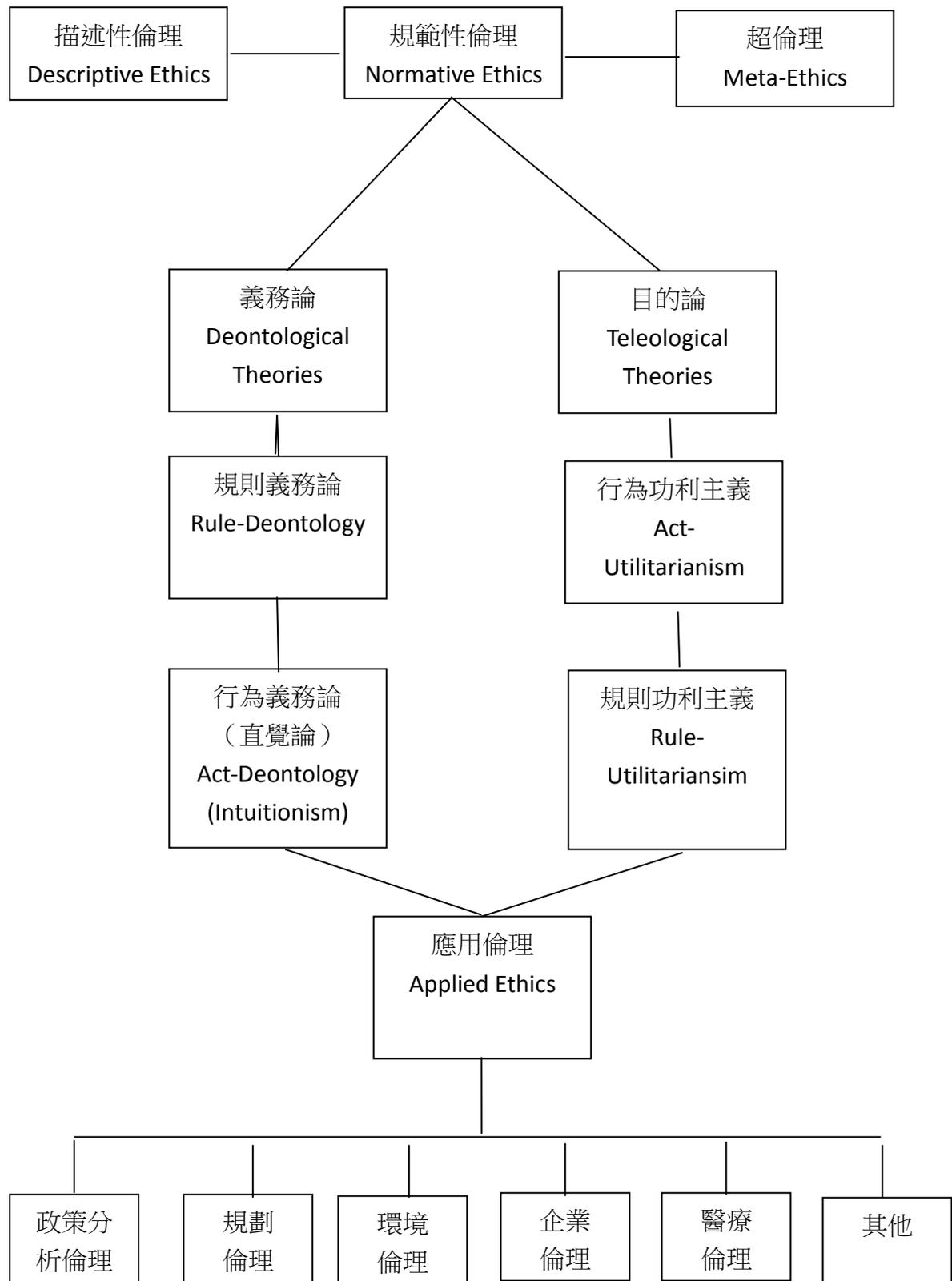
人必須審視這個問題：這些戰爭科技如果脫離負責任的政治領袖的控制，人類全體如何承擔其後果？一個負責任的政府，是否可以容許研發機構無止盡的 AI 為基礎的戰爭武器？。

以上這些問題，都需要公共政策學者與政府、產業界好好研究、思考與論證，提出一整套的政策邏輯，提早規範因應相關議題以及其可能發生的正面與負面效應。

### AI 與政策倫理的思辯

社會科學最為困難的挑戰，就是建立一套具有行為規範效力的倫理道德標準。在公共政策的論辯與發展中，社會科學如何與理工科技為先的學術界進行倫理道德的對話，更是重大的挑戰。AI 的發展，不論是軟體方面的提升，亦或是硬體方面的升級，都無可避免地必須面對科技與倫理道德的爭議、衝突與尋求解決之道。這些可能的倫理道德的爭議大部分與研究方法、決策過程中績效標準的倫理道德要求、以及政策結果評估的倫理道德標準。AI 科技的發展，也必須回答：研究的基礎、研究的過程、與研究的結果是否符合人類的道德倫理標準？但是，比較值得關注的議題是：AI 科技的發展中，人類的道德倫理標準的範疇、實質內容、與作業標準到底為何？不同的學術單位，如公共政策、社會、社工、勞工、政治、教育、法律、管理科學，與企業界對於研究 AI 的道德倫理的期望與要求如何？標準與程序是否有所不同？如何整合差異性？如何落實在研究與產業運用？這些都是必須嚴肅回應的問題。

人類的道德行為，係人類數萬年演化的行為規範與要求，其中包含描述性倫理 (descriptive ethics)、規範性倫理 (normative ethics)、與超倫理 (meta-ethics)。描述性倫理基本上係指合乎人類本性的倫理要求，規範性倫理是指在通則考慮下人類行為的好或壞的評斷標準，超倫理則是指邏輯上、認識論上、語意上的論證後的好壞判斷的結果(Frankena, 1963)。義務論 (Deontology) 關係著政策過程中的倫理道德問題，強調其規則(Rule-Deontology)是否符合倫理道德的規範，以及行為模式 (Act-Deontology) 能否合於規則。目的論 (Teleology) 則強調結果符合良善本意的要求。行為功利主義 (Act-Utilitarianism) 則是要求人類行為的結果唯一必須能夠帶來人類最大多數的最大幸福，不論其分配過程是否公平，而規則功利主義 (Rule-Utilitarianism) 則是強調，不論其結果是否為善，其行為、規則、與權利等必須在道德倫理上是符合公平正義的。其總和之規範，則形成人類行為的應用倫理，再依專業知識分類，分成各項職業專業的倫理規則。人類的倫理道德發展，可以從下圖中看出：



圖二：政策倫理理論（修改自：Patton, and Sawicki，1993）

由圖二吾人可以瞭解，人類在進行 AI 科技的探索與冒險的旅途之中，可能必須先行釐清研究高深科技之描述性倫理、規範性倫理、與超倫理的考量是什麼？在研究途徑的義務論中、與研究結果的目的論理，如何建立有效的行為典範與規則義務，才不會在追求科技的進步過程，迷失在 AI 進步的結果中，而遺忘了 AI 發展後可能帶來的災難。

目前 AI 科技的發展，對於倫理道德的論述與要求，並不多見。但是人類對於科技發展的倫理道德要求，在醫學發展中的人體實驗(Clinical Human Trial)的種種規範中，綜合許多學者的文獻與我國的醫療法語人體試驗管理辦法之內容進行分析，約可觀察歸納下列研究倫理標準：(1) 在實驗的設計與實施的過程中，必須保證實驗不能以任何方式短期地或長遠地傷害受試者或人類整體；(2) 必須成立一個獨立的審查機關，採合議制，由研究人員、法律專家、社會科學專家所組成，並且組成成員不得為單一性別，以審議實驗的安全性與可行性；(3) 強調安全性評估的重要性，遠高於實驗結果，並且必須受到相關法律的約束；(4) 研究計畫經審核通過後，必須在規範的範圍內嚴謹地進行實驗，並不得任意改變實驗範圍、程序與內容；(5) 實驗過程與結果必須像審議委員會定期通報進度，並接受檢查；(6) 受試者必須充分地被告知實驗過程與程序，以及可能的風險，並且擁有隨時退出終止實驗的權利；(7) 受試者的招募必須符合人口分布公平性(種族、性別、年齡教育程度等)的原則，而且必須保護其隱私權；(8) 研究者必須誠實地、公正地、公開地、不違背良心地進行實驗，並不得任意挑戰或改變自然定律 (Appelbaum, 2002; Emanuel et al., 2000; Fried, 1974; Miller and Brody, 2003; Glickman et al., 2009)。

基於以上的分析與檢驗，以及分析 AI 倫理的相關文獻 (Bostrom and Yudlowsky, 2011; Bedau, 1996; Anderson and Anderson, 2007; LaChat, 1986; Lin, Abney, and Bekey, 2011; Yampolskiy and Fox, 2012)) 吾人可以似乎可以期待 AI 發展的倫理道德要求，應該要有以下先決條件：(1) 建立以人為本的 AI 發展政策的先決倫理規範，任何實驗不能以任何方式短期地或長遠地傷害人類或是受試者 (雖然 AI 的發展可能並沒有人類的受試者)，並且必須成為所有 AI 相關課程與研究的基本訓練與準則。對於我國高等教育機構與研究單位而言，如何發展 AI 倫理道德的課程，必須將其列為首要任務；(2) 必須效法醫療機構，成立獨立的審議委員會，涵蓋各種科技、法律、社會科學、管理學等專家，進行研究計畫的審查，雖然可以不需限定研究主題，但是必須訂定嚴格的審查追蹤機制，注意實驗與發展的進程；(3) 我國高等學術機構應該配合官方與產業界，針對 AI 發展後所引發的各項問題進行廣泛與長期的追蹤與研究，來瞭解 AI 對於整體國家社會與人民的影響。這些倫理道德議題包含 (1) 勞工保護與失業率相關的議題；(2) AI 應用後所可能引發的財富分配不均與世代負擔不均的議題；(3) AI 應用後對於人類個人行為影響所可能產生的高科技相關的詐騙或網路犯罪等社會問題；(4) 嚴格的防錯機制以避免 AI 科技出錯後所可能產生的負面效應，如股市的崩盤；(5) 具有偏見的 AI 決策所可能導致的歧視問題，如在審核貸款

中所可能產生的刻板印象，影響貸款審核通過的公平性；(6) 防堵網路駭客操控 AI 智慧體的危險，導致個人或社會全體受到傷害的可能性；(7) 建立以人類控制作為最後主宰機制的防護措施，不僅要控制可能存在的瘋狂人類科學家，也要防止 AI 智慧體侵犯人類主權的可能風險；(8) 防範 AI 智慧體決策介入戰爭的決策之控管，以避免擦槍走火，導致萬劫不復的局面；(9) AI 智慧體出現後，如何保障 AI 智慧體的生存權利，以及人道對待等問題。

以上這些倫理道德議題，都需要相關專家學者開始分析、研究、檢討，並提出可行對應方法，迎接 AI 智慧體成為人類生活一份子的未來世界。

### 結論與建議

由於 AI 人工智慧在中華民國還是一個新興主題，所以，本文提出若干值得思考的問題，作為拋磚引玉之用，讓公共政策學者可以提早思考 AI 與公共政策之間的關係。

AI 人工智慧以及 AI 智慧體已經成為現代生活的必然，而且，證諸國內外的證據顯示，產業界、學術界、科研界在 AI 方面投入的努力與資源，遠遠超過政府關注的程度。AI 人工智慧雖然充滿未來性，對於改善人類生活也有一定的預期效益，但是仍然充滿太多的未知數與風險性。我國政府應該如何因應 AI 的發展，針對不同的應用範圍，提出不同的鼓勵措施或風險控管機制，的確值得相關部會思考。

而且，隨著 AI 人工智慧及 AI 智慧體介入各項產品的製造，對於勞工就業的衝擊、不同產業的轉型、生產技術的研發與提升、總體經濟的改革，教育體系的應對，政府都有責無旁貸的壓力。如何在科技快速發展的今天，沈穩地以有效率、效益的專業智慧，訂定相關法規，鼓勵尖端科技的研發，設定發展的倫理道德範疇，更是需要多加費心之處。

我國雖然在 AI 科技的發展上，無法與先進國家相比。但是目前以我國現行科技的優勢--晶片製造，做為我國 AI 科技切入國際競爭的基礎，可能又會淪為大國的生產基地。在現今科技進步神速的時代，作為低階 AI 智慧機器的生產國，應當會快速地面臨被其他各國以資本或技術超越的窘境。先進國家在發展 AI 科技的競爭過程中，不僅投入大量金錢，追求各層面技術的快速進步，而且不斷地追求創新領域的應用技術與知識。我國面臨如此嚴峻的競爭挑戰，如果只是滿足於發展 AI 的低階運用，會不會數十年後，又成為先進國家的生產殖民地？的確是令人憂心的問題。

目前各國在 AI 科技方面的發展，多採取鼓勵與樂觀其成的態度。對於相關的風險控管，並沒有提出有效的策略。尤其對於 AI 發展之後，對於勞工權益、市場經濟、社會公平正義、法律責任的範圍與釐清、個人隱私保護、教育訓練與軍事方面的運用，並沒有提出有效的對應辦法。吾人社會科學的研究者，如何掌握契機，分別就以上議題，進行深入的研究，以提供政府發展 AI 政策時的思考

方向。

賀建奎的基因編輯爭議，確確實實說明了一項革命性的先進科技，在有心人士的運用之下，會發生無法預料而且可能是非常恐怖的結果。這是貫徹高薪技術研究中研究倫理與道德的重大挫敗。如何建立一套完整的研究道德與倫理檢驗機制，避免研究人員進行具有野心但是充滿危險的研究冒進主義，是一項值得深思的問題。

AI 的發展，與 AI 智慧體的出現，無疑地是人類生活與科技的絕大進步，但是也充滿風險與不確定性。如何控管規範，以防止失序狀態發生，不僅是公共政策實務的問題也是相關學者專家不能不去面對的倫理道德問題。

## Reference

陳良基，2017年12月，打造人工智慧創新環境機制。國土籍公共治理季刊。第5卷第4期，頁60-71。

- Amato, Fillipo et al., 2013. Artificial Neural Networks in Medical Diagnosis, *Journal of Applied Biomedicine*, 11(2).
- Anderson, Michael and Susan Leigh Anderson, 2007. Machine Ethics: Creating an Ethical Intelligence Agent. *AI Magazine*, Vol. 28, No. 4: 15-26
- Appelbaum, P. S., 2002. Clarifying the Ethics of Clinical Research: a Path toward Avoiding Therapeutic Misconception. *American Journal of Bioethics*, Vol. 2, No. 2: 22.
- Bedau, Mark A., 1996. Philosophical Aspects of Artificial Life. In F. Varela and P. Bourguine, eds., *Towards A Practice of Autonomous Systems*. Cambridge, MA: Bradford Books.
- Bostrom, Nick and Eliezer Yudlowsky, 2011. The Ethics of Artificial Intelligence. In William Ramsey and Keith Franklin Eds., *Cambridge Handbook of Artificial Intelligence*. Cambridge University Press.
- Center for Democracy and Technology, 2016 July 27. Preparing for the Future of Artificial Intelligence: In Response to Office of Science and Technology Policy.
- Chen, Nicholas, Lau Christensen, Kevin Gallagher, Rosamond Mate, Greg Rafert, 2016. Global Economic Impacts Associated with Artificial Intelligence. Analysis Group Study for Facebook, 23.
- Economist*, April 20, 2013. Look, No Hands.
- Emanuel, E. J. and d. Wendler and C. Grady, 2000. What Makes Clinical Research Ethical? *JAMA* 283: 2701-2711.
- Fadalalla, Adam and Chien-Hua Lin, 2001. An Analysis of the Applications of Neural Networks in Finance, *Interfaces*, 31(4).
- Fried, C., 1974. *Medical Experimentation: Personal Integrity and Social Policy*. New York: AmericanElsevier.
- Glickman, Seth W., John G. McHutchison, Eric d. Peterson, Charles B. Cairns, Robert A. Harrington, Robert M. Califf, and Kevin A. Schulman, 2009. Ethical and Scientific Implications of the Globalization of Clinical Research. *NEJM*, 360; 8, pp. 816-822.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville, 2016. *Deep Learning*. Cambridge, MA: MIT Press.
- Johal, Sunil and Daniel Araya, Feb. 28 2017. Work and Social Policy in the Age of Artificial Intelligence. Brookings Tech Tank.

- LaChat, Michael R., 1986. Artificial Intelligence and Ethics: An Exercise in the Moral Imagination. *The AI Magazine*, Vol. 7, No. 2: 70-79.
- Lin, Patrick, Keith Abney, and George Bekey, 2011. Robot Ethics: Mapping the Issues for a Mechanized World. *Artificial Intelligence* 175(2011): 942-949.
- Miller, Franklin G. and Howard Brody, 2003. A Critique of Clinical Equipoise: Therapeutic Misconception in the Ethics of Clinical Trials. *The Hastings Center report*, Vol. 33, No. 3(May-June, 2003), pp. 19-28.
- Munoz, Cecilia, Megan Smith, and DJ Patil. May 2016. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*. Executive Office of the President.
- Podesta et al., May 2014. *Big Data: Seizing Opportunities, Preserving Values*. Executive Office of the President.
- Rohling, Gitta, October 1 2014. *Facts and Forecasts: Boom for Learning Systems*. *Innovations Newsletter* (Siemens).
- Russell, Stuart J. and Peter Norvig. 2009. *Artificial Intelligence: A Modern Approach*, 3<sup>rd</sup> ed.. Upper Saddle River, NJ: Prentice Hall
- Smola, Alex and S. V. N. Vishwanathan, 2008. *An Introduction to Machine Learning*. Cambridge, UK: Cambridge University Press.
- Thierer, Adam, Andrea Castillo O'Sullivan, and Raymond Russell. 2017. *Artificial Intelligence and Public Policy*. Mercatus Center at George Mason University.
- Thierer, Adam. 2016. *Permissionless Innovation: The Continuing Case for Comprehensive Technological Freedom*, 2<sup>nd</sup> ed. Arlington, VA: Mercatus Center at George Mason University.
- Thompson, Derek, July/August 2015. *A World without Work*, *The Atlantic*.
- Vallor, Shannon, July 19, 2016. *On Artificial Intelligence and the Public Good*. Markkula Center for Applied Ethics at Santa Clara University.
- Venture Scanner, Nov. 30, 2016. *Artificial Intelligence Companies Founded by Year-Q4 2016*.
- West, Darrell M.. October, 2015. *What Happens if Robots Take the Jobs? The Impact of Emerging Technologies on Employment and Public Policy*. Center for Technology Innovation at Brookings. P: 1-22.
- Yampolskiy, Roman and Joshua Fox, 2012. *Safety Engineering for Artificial General Intelligence*. *Topoi*, 32:217-226