意識之於人工智能通識教育課程的倫理議題探究與反思價值素養

陳 康 芬*

摘　要

本論文旨在論述置入意識探究學習設計之於人工智能通識教育課程的重要性。本論文認為人工智能的強與弱雖然不能等同於人類的心智，但具有可類比與模擬人類心智運作的客體化的觀察特質，可以將其看作是能進行學習心靈特性的開放架構，因此，可以作為關涉人工智能課程可以導入的倫理議題設計基礎—透過觀察人工智能的運作過程，瞭解作為人類「理性代理人」的他者，作為學習者的關鍵學習經驗，再以此關鍵學習經驗延伸為「人工智能的人工意識 V.S.人類心智的心靈意識」的探究學習設計。學習者透過自我意識探討等設計性的倫理議題探究，認識人類的主體與價值，笛卡兒對於自我意識的心、物探討架構，可以提供學生對人工智能的類主體的認知判斷，並進一步反思人工智能所需要的人工意識的條件，以及其仿生物相連的心智特性之於人工智能發展的意義。

關鍵字：人工智能教育、人工意識、演算法、人類自我意識、學習心靈

* 中原大學通識教育中心副教授。

A Study for the Teaching Mission and its Value about Self-Reflection of Human Consciousness and AI Consciousness to the Course "AI General Education" through Ethical Issue Quest Learning Approach

Chen, Kang-fen*

Abstract

This study aims to discuss the design of AI Ethical Issues Quest Learning based on AI Consciousness in which there is resemblance of Human Self-Consciousness. There are three statements. First, the AI Algoithm can not be equal to Human mind, but having resembling attributes as Human mind. Second, there is a framework offering a clear visible process to learn how human mind constituted through neurons to consciousness. Last, as a significant learning experience for learners, the design of ethical issue quest introduces to a comparable discourse about the essence of AI consciousness and human consciousness. Learners will construct their own knowledge about what AI is. Then, learners are able to infer the conditions of AI consciousness resembling human involved with the reflection how to define AI consciousness.

Keywords: AI Education, AI Consciousness, Algoithm, Human Self-Consciousness, Learning Mind

* Associate Professor, Center for General Education Director, Research Center for Global Hakka and Multi-culture Chung Yuan Christian University.

壹、前文

隨著人工智能演算法與相關對應科技的突破，人工智能的類人化／真人化已經不再只是科幻小說劇情，而是正在發生的世界真實。這個世界真實尚未真正發展出改變世界的本質力量，已然成為華人教育體制因應未來科技社會想像的個體知識資本，連帶引發人工智能與賦能作為改變學生學習與教師教學現場的創新方法／技術熱潮，也帶動人文知識結合人工智能在跨域研究的屬性方法，即 AI of／for／by／to Knowledge 的開發想像。

AI of／for／by／to Knowledge 的理解架構為清華大學通識教育中心主任林文源教授演講提及，但這個理解架構雖然可以客觀描述 AI 與人文科學之間所能建立的研究向度，但也擱置從傳統人文思惟或人文社會批判視角所帶來的反思研究的推進，以及跨人文與科技領域對話所需要的主體意識。通識教育作為近年高等教育體制所亟欲推動的改革對象，除了持續可從 AI of／for／by／to Knowledge 架構開啟人文與科技不同領域的整合研究與應用外，「人本」的思惟方式，面對作為現代方法的人工智慧的挑戰，可以帶來哪些值得探究的倫理議題與價值省思，並且作為推動智能教育的人文素養與知識架構設計，正是本論文所嘗試探討的向度與建議。

但是，從人文知識的核心素養「人文價值」立場來看，前述人工智能與其賦能之於教育現場的應用、或是之於跨域研究的導引，除交相指涉出人工智能作為未來社會發展主流科技與重要技術價值的「預測」，人工智能在發展過程所展現的擬人化或日漸逼近的真人化現象，開始挑戰人才能作為知識學習主體的基本認知，也啟動人工智能是否有其人工意識的重要提問，以及挑戰人類如何審視意識之於主體的意義與價值。

本論文認為人工智能是否具有其人工意識的重要提問，關涉人工智能通識教育的倫理議題的探究學習開發，也關聯人文主體在人工智能教育不能被排除的反思價值基礎。因此，本論文擬從自我意識之於人的主體性的建立根基，針對人工智能的演算法特質、與其對應的人類心智運用和意識的關係，進行人工意識的類人化／真人化的本質探討，並以此提出人工智能教育的倫理探究學習的最適人文價值設計與教學原則建議。

貳、本論

一、從人工智能的跨域研究探問到智能教育的核心人文素養

人工智能（Artificial Intelligence）的研究發展突破，某種程度已經落實人工智能機械人的想像願景，如 Hanson Robotic 公司執行長 David Hanson 主導開發的 AI 機械人蘇菲亞（Sophia）不僅是 Hanson Robotics 公司最新、最高科技機械人，也是首位取得公民資格的機械人，還曾獲邀聯合國發表演說，在聯合國開發計畫署（UNDP）被提名為世界首位創新守護者。

人工智能機械人的擬人化／真人化科技，除了能讓人工智能機械人有細膩多變的臉部表情，透過攝影機的眼睛，將影像傳達計算，還可辨識臉孔，也能隨著學習，累積對話經驗的語言應答與表情回應，智能機械人甚至帶動 AI 與人類之間的可能成真或想像衝突—如 AI 知能取代人類工作的恐慌、AI 智能機械人有自主意識後取代人類地位的焦慮等問題。因此，人類與 AI 之間的合作如何從人文知識領域學習的主體性啟動，並建構人工智能擬人化／真人化之於人文學科的探究學習，成為人工智能通識教育不可忽略的核心素養。

臺灣教育部近年來積極以相關計畫獎勵等政策導向，鼓勵拿到教育部補助計畫的大專院校，開始將人工智能教育置入相關通識課程的主題教學內容，或是由相關系所開設人工智能專業課程。在人文學科領域，科技部人文社會科學研究中心、國立清華大學人社 AI 團隊（THSSAI）、國立台灣大學等機構也積極從大專院校教師群啟動相關人工智能應用人文社會科學的跨域研究。但目前僅止於鼓勵認識與應用的推廣性質度程度，並未將人工智能教育的倫理議題視為是人文社會科學領域的研究或教學的重點討論內容。

臺灣對於人工智能在人文社會科學領域研究鼓勵與通識教育課程的教學推廣，都不曾認真探究西方發展人工智慧歷史的研究途徑，也因此忽略人工智能在試圖理解與建造智能實體過程、真正最大挑戰「以人為中心」核心價值之於西方不同學術領域所產生的相關現象與本質的探問問題—如果人工智能可以像人一樣的思考、可以有人一樣的行動、可以建立像人一樣的理性思考系統、擁有像人一樣的理性行動系統，人工智能是不是應該取得與人一樣的地位？而我們作為地球目前唯一擁有自我意識的高等智慧生物的「人」，又該如何看待人類自我意識與人工智能的意識？—人工智能的相關意識問題，挑戰了人類的意識認知，以及從意識認知而來的人的主體價值，以及從人的主體價值重新審視人工智能的客體價值判定。本論文認為這些問題與判定，涉及人工智能通識教育的倫理議題探究，並關涉人工智能通識教育絕對不能被排除在外的

人文價值核心素養。

人工智能擬人化／真人化之於人文學科的探究學習，之所以成為人工智能通識教育不可忽略的核心素養的理由，與人工智能自 1956 年被正名以來的歷史發展過程的四種研究途徑，息息相關。這四種途徑包括：類人類行為：圖靈測試方法；類人類思考：認知模塑方法；理性地思考。其「思維法則」方法則包括；理性地行動：理性代理人方法。每一種途徑都涉及以人為中心的思考運作。以人為中心的研究途徑，在某種程度上必然是一種經驗科學，牽涉到人類行為有關的觀察和架說；理性主義研究途徑則涉及數學與工程的結合（Russell, 2011, 1-2）。

這些不同研究群體彼此將互相批判與幫助對方，從不同學科的跨域研究進行合作，為人工智能領域奠下不同的想法、觀點與技術，而構成吾人之於人工智慧本身的探究基礎。這些學科與重要影響，包括：哲學—支配人類心智的理性可以等同是意識嗎？精神意識如何從物質的大腦產生出來？知識從哪裡來？知識如何導致行動？這些提問涉及「人工智能」的重要定義與界限；數學—提供人工智能躍升為一門正式科學的基礎，包括人工智能在邏輯、計算都能形成確知的各式演算法，以及如何用不確定的知識進行推理的機率理論等；經濟—從賽局理論未能提供的選擇行動的明確規定（unambiguous prescription）的延伸思考角度，啟動人工智能的作業研究範疇，以及理性代理人系統的決策理論技術；神經科學—透過大腦處理資訊的研究結論「大腦產生意識」，延伸出人工智能是否能有其人工意識的探問；心理學—將人類和動物是如何思考和行動的心理現象研究，轉擬為電腦模型的資訊處理的機制發展與認知理解；電腦工程—效率的電腦資訊軟體技術提供人工智能實體化的作業技術、程式語言、寫作現代程式（和關於它們的論文）需要的工具；控制論與模控學—人工製品的自動運作控制研究，促成人工智慧之於設計行為表最佳化的實體化期待；語言學—語言如何與思維連結的語法模型的研究視角與詮釋論述，與 A.I. 交織形成計算語言學或自然語言處理的混合領域，啟發人工智能中如何將知識轉換成電腦可用於推理形式的知識表示的研究與開發（Russell, 2011, 1-5,1-7,1-8,1-11,1-12,1-14）。

這些學科領域的知識研究成果不僅為人工智慧貢獻了許多重要的想法、觀點與技術，也形成人工智慧在現代西方發展為一門獨立科學學科的建構基礎，而在這個建構基礎上，人工智能漸次形成接近類比於「人」概念的「理性代理人」概念。因此，釐清人工智能與「理性代理人」的互屬關係與價值定位，不能只強調人工智能的功能現象，或是從人工智能本身的技術思維探討人工智能

的現在與未來應用；特別是引入人工智能的通識教育課程設計，對於人工智能之於科技人文或人文科技的挑戰與對話思考，仍有必要從「人」對於自我探索所驅動的自我價值肯定，作為探究人與人工智能之間倫理界限的重要基本提問—這也側面點出「人」與「理性代理人」之間的差異，在於人作為自我的主體存在與人工智能作為人的「理性代理人」的客體存在之間的本質差異。

也就是說，「人」與「理性代理人」兩者的自身存在，指出兩者之間最大的本質差異—「人」作為具有理性屬性的人自身，而人工智能則僅僅作為人的理性代理人的認知與類身份，毫無疑問地，「人工智能」並不能等同於「人」。但是，人工智能本身運作的類人類行為與類人類思考現象，使得人工智能並不只是單純的技術或實體，而是涉及到人之所以是人、不同於其他動物的特殊屬性與特殊條件，以及可展演過程，特別是作為一種現代方法，人與物之間的關係不再容易保有單一向度的主體對客體關係，而在本質上，則有越來越趨近的主體與類主體的關係。

這說明人工智能本身的類人化的理性機制運作，恰巧可被視為提供人的心智是如何運作的一個可拆解的過程化他者，而不是一個絕對的他者，包括認知模塑、思維法則、邏輯推理、語言處理……；這些過程化所顯示的理性代理人與其環境互動行為運作，可以清楚看到人工智能作為一種模擬人決定做甚麼、或是執行行動的系統與現代方法屬性，展現出人類本身作為高等智能生物如何運用思考解決問題、如何透過推理與規劃形成世界的知識、又如何處理不確定知識的不確定因素與其推理與決策、如何因應決策元件產生所需知識的學習方法、又如何透過視覺、觸覺、聽覺、語言處理來感知環境與回應行動……等心靈或心智智運作的交互過程。

而人工智能本身從弱人工智能到強人工智能、漸趨向於整合發展的機器人學，甚至結合生物功能應用的未來想像的潛能科技發展，都可以看到人工智能一直再再挑戰智能之於人的獨特性與人的生理性限度。這些現今技術帶來的改變與未來技術可能再突破所延伸而來的發生想像，都再再讓我們很難忽略已然存在我們心中很久、也尚未被解決的疑惑：人工智能所代表的「理性代理人」的人工意識，到底是不是、會不會、能不能、應該不應該等同於人的意識？這幾個從「是或不是」、「會或不會」、「能或不能」、「應該或不應該」的看似二元對立的選擇，其實都關乎人工智能作為現代方法後被普遍接受的合理應用的範圍界定，以及人的智能與人工智能之間的價值與認知定位。這是智能教育的核心人文素養之所以需要被重視的主要原因之一，另一個主要原因是透過人文思維的引導，才能暫時延宕科技發展的技術中心的主導思維，重新認真看待人類

歷史文明中一直就存在的「人是甚麼」的重要提問，以至於認真看待人工智能的人工意識與人的意識之間所可能存在的本質差異問題。

人工意識究竟與人的意識的本質差異可能是甚麼等相關問題，之所以值得被認真看待，涉及到幾個人工智慧作為現代方法的幾個重要的問題：一、弱人工知能的模擬行動與人的智能行動之間是否有本質上的差異？二、強人工智慧從決策到行動過程可以等同於人的思考嗎？三、發展人工智慧之後的社會分工改變風險與倫理規範限制，應該如何重新界定人與人工智能相關責任歸屬。這些重要的問題都無關人工智能本身作為人類智能代理人的技術發展，而指涉出人工智能與人之間的本位思考與異同性判斷，包括：人會思考，但思考是甚麼；人工智能的智能運作是不是思考？人有意識，但意識是甚麼？人工智能的智能運作能不能等同有意識？人工智能可不可以有意識？可不可以有思考？或應不應該被認為是思考？這些探究涉及到我們如何看待「人」的整全性與「理性代理人」之間的區分與價值認定，以及「思考」作為人的獨特性的自我價值與具有類思考表現的理性代理人之間的本質差異，進而肯定以「人本」為前提的認知態度。

因此，引入「思考」與相關概念，作為人工智能與人的界限探討，可以較清楚比較人在思考過程與人工智能進行類思考過程的異同性，也可以看到人類對於自我意識存在的理性確認，以及不必依靠其他知識或外在事物、即可進行推論與證明而獲得知識的原理方法的特質屬性。從這個觀點來說，區分人類的「思考」與人工智能的「思考」，有助於幫助我們探究人類與擬人化機器架構各自對「思考」的開展，以及進一步提問—「機器可以思考嗎？」、「機器的思考跟人類的思考是否有本質差異？如果有本質差異，該如何敘述？」等系列問題。這些問題關聯我們應該如何去設定人與人工智能本身之間共存合作的相關倫理問題，以及更客觀審視人工智能的類人類意識運作與行為模仿的「理性代理人」的「他者」身分。

除此之外，在人工智能成為事實之前，人類對於自我意識存在的確認事實，一直是人類可以獲得知識的基本原理的重要認知。這個重要認知奠基於笛卡兒。笛卡兒《沉思錄》以懷疑精神與沉思推理通向理性與意志的原理自證，指出不管對事物有多少懷疑，我們不能懷疑自己的存在；這是系統推理思維中所發現的第一件事實。但是，由此，我們也發現了心靈和肉體之間的差異，或者說，思想物和非思想的物體之間的差異。笛卡兒透過懷疑與系統推理思維所獲得的自我意識的確切，反映出人在思考過程所能掌握的自我意識的兩大主體性特質：因自由意志而有的懷疑精神與系統推理思維的理性特質。從這兩大特

質反思目前人工智能作為理性代理人的身分認知，可以清楚看到人工智能即使有如人般的智能行動（弱人工智能）、與擬人的智能思考（強人工智能），除非擁有自由意志與懷疑精神，否則人工智能是不宜附之以主體性的認知定位。這個認知觀念可以協助我們區分弱人工智能與強人工智能之於類人化的程度判定，以及人工智能領域的倫理規範探討。

因此，人工智能的實踐正如人工智能作為理性代理人的定位，不管是弱人工智能或強人工智能，人工智能自身的實踐不僅僅是現象上機器本身可依循數學原理或程式系統中推理思維的結構性而來的行動結果，也展示人工智能本身對於人類理性思考與行動的「模擬」判斷，無關乎人的自然生命體與機器的非自然生命體的差異，而是關乎如何認定人工智能的「模擬」自身，能不能視同於人的思考與行動，以及類人的程度化之於社會應用的相關責任分屬問題。這些問題關係到人工智能作為智能化代理人的發展限度—如果人工智能的技術發展威脅到人類社會的生存競爭與風險危機，該領域的研究者仍有其道德上的義務限制其研究或改變研究方向。

首先，弱人工智能的「模擬行動」與強人工智能的「模擬思考」，雖然有強弱之分，但對於擬人化的程度該如何判斷？圖靈是第一位提出重要觀點的專家學者，他從機器的反應結果是否能通過人的判斷的模擬遊戲作為測試基準。圖靈的模擬遊戲測試，將機器能不能思維的根本問題，轉換為機器是否能取得和人一樣的判斷數值（圖靈，1950）。圖靈的轉向提問，暫時擱置了機器與人之間的本質界定的認知與定位問題，而開啟了人工智能可以像人一樣全面發展的想像空間。但是，當人工智能的技術與表現越來越成熟時，甚至如圖靈所預料機器可在所有純智能的領域中同人類競爭的情況，如何從人文的核心價值保障人類的智能表現，以及強化人工知能的「模擬」屬性，應該是人工智能研究領域中、不容缺席的重要認知，以及作為延伸為倫理規範基礎的探討議題。

二、人工智能領域的意識探究的脈絡化思惟與倫理導向教學設計

很多人工智能的研究者將人工智慧視為理所當然，並不會太在意人工智能的模擬智能與人的真正智能之間的差異，但是，人工智能的「類人化」是界定人工智能倫理的重要基礎，因此，區分判別人工的模擬智能與人的心智運作，進而成為界定人工倫理智能倫理的首要設準。

從笛卡兒在《沉思錄》的心物二元論觀點來看，思考中的人的心智活動具有一種可以不受限於現象空間與物質特性的過程，並且發現靈魂與身體是兩種不同屬性的存在（笛卡兒，2018）。笛卡兒的心物觀點與不用預設形上存在的

第一哲學的推導證明，會使得人工模擬智能與人的心智的異同性，會隨著論者的詮釋立場而有不同的看法。

也就是說，如果論者承認人的心智運作來自於人有靈魂屬性的關聯性，人工智能的非人與非具生物性的前提事實，可以知道人工智能之所以不能等同於人的心智、只是「模擬之物」，原因在於機器沒有靈魂，機器是人的創造性物質，而人是「神照著祂的形象造男造女」，且有「生養眾多，遍滿地面，治理這地，也要管理海裡的魚、空中的鳥，和地上各樣行動的活物」的管理祝福（舊約聖經・創世紀 1:26-28）。兩者在創造上有來自於人來自於上帝、而機器來自於人的本質脈絡差異，進而以此對照西方人權發展歷史過程中，「人」之所享有（人作為榮耀上帝之高峰創造而在神面前人人平等）的「天賦人權」的思想基礎，但人工智能並不是人，也不擁有人的主體性，只是作為人的智能模擬、具有規範人的主體性的屬性與物質他者。從這個觀點來看，人工智能並不具有成為人或是取代人的正當性的立法基礎，即使人工智能在未來演化到智慧爆炸／技術奇點（the Singularity）的階段，也能在發展倫理的規範上，盡量避免超級智慧機器所帶來的毀滅性趨勢的能力（Russell, 2011, 26-16、26-17）。

然而，這仍然有一個值得嚴肅以對的人工智能技術發展的倫理規範問題：如果人類想要延續一個以人的主體為核心價值發展的文明社會型態，應該如何從群體與個體規範限制或限縮科技技術為單一或終極價值導向的現代化文明社會發展？這個提問也涉及到人工智能究竟應該不應該、或是是否需要發展到所謂的智慧爆炸／技術奇點的階段？或是在發展至智慧爆炸／技術奇點的階段之前，能夠找到一個平衡人工智能技術發展與保護人類及其主體社會為優先考慮的控制法則或設計規範。

科幻小說作者 Isaac Asimov（1942-）是第一個討論到這個問題的人，並提出機器人三大法則：一、機器人不能傷害人類，或者允許人類遭到傷害；二、機器人必須遵循人類所給予的命令，除非這個命令違反了第一條法則；三、機器人必須保護他自己，只要這個行為不會和第二條法則有所衝突。（Russell, 2011, 26-17）這個設計法則可以視為人類目前對機器人做出倫理規範的想像設定—機器人對人類有優先保護權的絕對義務，並以此規範機器人對人類下達命令、執行命令與保護自己的基本法則。Asimov 對機器人倫理的法則設定，確實可以避免機器人一旦突破技術奇點後、對人類可能造成的傷害危機。但是，Asimov 也在故事情節發展中，提出機器人被派去開採硒礦，但偵測到危險而又轉向離開、離開後危險降低又去執行命令的圈迴狀態。這個可能性暗示機器人的法則執行不是邏輯上的絕對，而是權衡關係下的最佳能力化設計（Russell,

2011, 26-17）。關於 Asimov 所設定的機器人三大法則，可以清楚看到人類優先於機械人的絕對被保護權，以及符合人類安全優先權之下被賦予的行為自主權。這個關係反映出人與機械人之間的主從定位，以及機械人遊走於類人類自由意志的邏輯行為與其自主權限的矛盾狀況，都不能與人的自由意志與其影響行為相比擬。

再回到如何定位人工智能的心智能力開發與其身份界線的問題點上。如果論者擱置或刻意忽略笛卡兒從宗教立場而來的的創造論觀點，單純就心、物本為不同屬性、且有二元對立或彼此可斷裂特質的存在而言，則更可以顯而易見人工智能與人工意識的出現與事實，無關於「心」、而限於可見可知的「物」的範疇而言，再再挑戰人作為地球生物的「中心位置」的特殊屬性—作為人的思考與人的心智活動的物質他者的人工智能，是否可因其可見可知的現象屬性而可被視為人；另一方面，作為萬物尺度的人又該如何看到心智運作與自我意識之間的關聯性，以及人工知能模擬人類知能過程所出現的人工意識現象？包括如何在接受人工智能常態化的現代文明社會發展過程中，能夠客觀地或批判地審視人工意識與人工智能作為機器模擬人的意識與智能的他者，並認為建立以人文為中心的檢視意識與教育認知機制是重要的關鍵議題與倫理規範基礎？這些提問關乎如何從人文觀點去建立或判定對於人工智能的他者的客體價值認知。

也就是說，如果說人的思考的心智活動的事實可自證意識存在，人工智能的「模擬」的事實，亦能自證「人工意識」的存在；兩者的意識因其本質不同，而可類歸為不同系統的各自運作過程，且人工意識可因「模擬」而成為人的意識運作的投射他者，具有可觀察性；進而，根據人工智能的弱與強，分別提出「弱人工智能：機器能夠智能行動」、「強人工智能：機器真的能夠思考」的層次性針對問題，並順勢形成兩種可彼此參照的意識探討架構。

但是，笛卡兒的心物分離觀點，會產生一個根本問題—如果心智與身體分屬不同，作為一個整全個體的人是如何以心智控制身體？笛卡兒推論兩者可以透過松果腺（pineal gland）進行互動，並且將這個問題簡化為心智如何控制松果腺，但是，仍然無法解決或懸而未能解心智是甚麼、心智又是如何產生的根本問題。繼續心物分離的難題，相對於笛卡兒的心智二元主張，另有認為心智與身體應為一體的心智一元論的主張，通常被稱為唯物主義。這類避開心物二元分屬的難題，從現象事實的統一性，提出心理狀態即為身體狀態的看法，而將探討焦點放在心智與身體的共在性—如以腦神經的傳導與脈衝現象同時產生心理狀態的解釋框架。但是，這樣的解釋框架也容易導致延宕笛卡兒心物分

離對整全個體的人是如何以心智控制身體的難題，直接訴諸心智一元論心理狀態是根自「腦神經」的生物／生理性運作的認知，進而轉向心智一元論的預設基礎。

心智一元的唯物論傾向將人的複雜性簡化為「物」屬性的現象解釋與詮釋邏輯，如人的心智運作來自身體，而身體的運作可以決定人的心智的意圖。以這個觀點會強化人工智能的「模擬」與人的心智運作之間的無差異性、或是將兩者之間的差異性暫時擱置，專注探討「功能」與「功能導向的意圖狀態」。當人工智能的「模擬」與人的心智運作的「功能」可以對等、或甚至超越時候，兩者之間也就自然消弭了意圖探究的必要性，可能翻轉人之為人（不同於其他地球物種）的尊貴本位認知，喪失繼續探究或承認人有自我意識的獨特性的興趣，甚至否定人有不同於「物」的「心」屬性存在，將人的可能學問與價值發展都限制在「物的真實」層面，而將人關於心智與心靈的意識問題，一概簡化為生物性的實體物（如大腦）輸出過程。

心智一元論的唯物觀點對於人工智能的發展影響，雖然可以從論理的正當性，提出開發人工智能技術的興趣與必要性，將人工智能的研究價值集中於功能開發，而懸吊意識之於主體、之於人的獨特性的人本價值認知。因為，將以大腦的功能性決定或對等於意識的出現，就等於承認人的意識的運作只是神經元輸出、輸入的傳導現象的副作用物，人的心靈、心智運作也只是更巨大或更複雜傳導現象的必然產出結果，進而無關自由意志，或懸而不論人有意志自由的事實。

自由意志之於人類的重要心智與心靈現象，並不只關乎人在各種條件與環境之下的意志選擇，也關乎人生而為人必須要去維護的尊嚴與存在高貴性。人對自我高貴的自重與對他人也同樣高貴的認知對待，是人生而皆享平等的人權原理，以及人必須先以道德自重原則（如康德為道德而道德的義務與令式）為人之本的探究限度，才能審度道德實踐中的意志自由價值與人文精神。以此理路對比人工智能的「模擬」，會發現人工智能的機器透過程式運算的「意識」副作用，只是依照程式命令執行的類意識，其中所衍生的人工心智與人工心靈，並不能等同於人的心智與心靈。

然而，心智一元論對功能導向的專注與探討，仍有可能衍生出一種從結果推論的可能性：人工智能的意識與其衍生的人工心智與人工心靈，雖然缺乏自由意志的先天條件，但人工意識衍生其心智與心靈的類人思考的模擬本身的「思考」，是甚麼樣性質的思考？這個提問雖然在前提上，承認人工智能的意識是對思考的模擬作用與產出結果，即使不能等同於人的心智與心靈，又應該

如何解釋來自模擬思考的思考？這可以讓機器本身擁有或等同於一種精神活動嗎？

如果回到笛卡兒心物二元觀點的理路，繼續探究下去，人工智能的模擬思考與人的思考活動，在認知上的判斷，都是可以分別獨立於機器與身體而存在的事實，但是，兩者的思考運作又有甚麼差異性？以致我們可以判斷人的思考是精神活動，而人工智能的模擬思考不是？如果模擬思考的思考與真正的思考之間，確實存有本質上的差異，人工智能的物之思，也是一種具有真實性的思考現象，而自證「機器能夠思考」的事實。

不過，對於「機器真的能思考」的問題，仍有極大的異議空間—機器的人工意識雖然因不具先天自由意志而缺乏自我獨立與自主性，但有沒有可能會發生心智一元論專注於功能導向而可能發生的假設現象：當人工智能的技術使人工智能的智能運作與行動，都能精細到完全真人化，甚至做到人工智能的模擬可以同化人類心智的程度，機器不具人先天存在的自由意志的前提條件，已然不是那麼重要—並進而承認人工智能的模擬思考本身即使是一種類精神活動，但因為具有思考屬性，就算不能等同於真正思考，也可能被視作人類思考的異化現象而理所當然。一旦發生了，擁有人型的強人工智能與人之間的界限該在哪裡？人工智能人該不該擁有「人」權？擁有甚麼樣的「人」權？人工智能的製造者／使用者與人工智能的智能之間的應對關係、及其生產消費使用關係，人工智能的智能行為與心智產出又該如何分配責任分屬與其合理性？

這些預設性想像問題，提醒我們選擇繼續無限制發展人工智能之後的風險控管，以及從計算的道德規範（Ethic of computing）延伸出人工智能社會可能出現的危機現象，像是：人們可能由於自動化而失業、人們可能擁有過多（或過少）的閒暇時間、人們可能會失去作為人的獨一無二的感覺、人工智能可能導致非預期的結果、人工智能系統可能會導致責任歸屬的喪失、人工智能的成功可能意味著人類種族的終結等種種難題（Russell, 2011, 26-13）。

這些種種難題顯示繼續發展人工智能與相關技術的潛在風險，以及為何在人工智能領域中必須預設人文思維與倫理價值，作為回應人工智能繼續發展下去的核心素養能力的理由。同時，將人文思維與倫理價值納入人工智能知識與技術開發不可或缺的反思素養能力，主要的原因在於，人工智能的極化挑戰了人類心智在地球生物中的獨一無二性，並以科技的技術實踐了機器模擬人類心智運作的想像。

這意謂未來人工智能的極化發展，勢必影響人類心智分工的社會型態，就像十八世紀工業革命的機器技術，既創造了前所未有的文明生活，但也徹底改

造傳統社會的勞力技術分工型態。人文思維與倫理價值導入人工智能知識與技術領域的反思探究，可以提醒使用者與技術開發者，未來以人工智能為開發技術導向的社會型態發展，「人」與「人工智能」的本質有何不同？人工智能的人工意識與人類自我意識、學習心靈的現象之於「人」的意義又各自代表甚麼意義與價值？人工智能可以取代人的甚麼？不能取代人的又是甚麼？這些問題都關涉到課程為何採取人工智能意識與人類自我意識的架構對照設計，作為探究人工智能倫理的教學設計與學生引導基礎。

但是，這些以人類文明歷史發展與人文價值向度為核心思惟的探究架構，都未被目前人工智能的通識課程所重視，顯示人工智能的通識課程設計本身的跨域探究，並未普遍或理所當然導入人文通識領域教學者本身所能提供的人文思考或人文批判觀點，以至於人文主義向度對人工智能發展所可能啟動的基本提問的缺遺，如人工智能模擬人類心智運為何可以/不可以等同人類心智的認知觀點是甚麼？這些認知觀點如何設定人工智能在現在社會或未來生產應用的各種權限關係與倫理界限？人工智能的「理性代言人」作為人類心智運作的可顯現他者，是否挑戰、如何挑戰人類心智的創造性與特殊性？人工智能的人工意識與自主學習機制之於人類自我意識與心靈學習的異同性，如何挑戰或改變過去人文主義的傳統？我們接受挑戰的同時，如何重建人文主義之於人本核心的普世價值？以及重新喚起我們對「人是甚麼？」、「人的價值是甚麼？」「人的自我價值如何透過探問而形成普遍知識？」笛卡爾透過自我懷疑而推論出對於自我意識的確定性，之於人工智能是否能因而產生人工意識，提供了一個值得教學者設計開發的議題思辨與探究進路。

三、人工智能的人工意識探究之於人類自我意識與心靈學習的開放式觀察框架與其相關探討議題

不管是從笛卡兒的心物二元論，或是延宕心智如何控制身體的難題，轉向心物一元論，都可以發現人工智能的科技發展，挑戰心物二元論或是心物一元論一致以「人」為本位的心智討論範圍，進而我們可以透過笛卡爾為探究思惟的起點，進而提問一個可能性：如果人類的心靈真的有多重的可實現性？笛卡爾的心物二元論與心物一元論的理解與詮釋框架，則可以為人工智能的人工意識的可能性與如何理解其異同於人類意識，提供了一個既可以分解討論、也可以綜合討論的探究上的思惟架構。而人類本身的心靈的多重可實現性的前提，也可以讓我們避開以人為本位、或以人為唯一有心智能力的高等智慧生物的認知本位觀點，而將焦點放在大腦的運作／執行方式本身，即大腦本身的運作／

執行方式與程式驅動有異曲同工之妙、甚至有同屬於「心」屬性的「本質」可能性，則心靈的多重可實現性的問題將可以得到解決或被證實。

這個理解的轉向，會解消人類意識所對應的心靈狀態，不再執著於只限於有靈魂的人類、或取決於神經元激發與運作大的腦狀態，而將重點放在「功能運作／執行本身」。這個觀點傾向支持以功能狀態解釋意識的存在與心靈的運作，但是，卻不宜直接從此觀點混淆或直接推論人工智能的人工意識可以等同於人類的意識。因為從身、心、靈三個整面向而來的人觀，或是由身與心整合的人對自我的認知，所強調的「人」都不是權衡於功能的「工具」，而是會趨向於存在與意義的價值探問脈絡，不同於「功能運作／執行本身」的複製經驗，僅是一種輸入與輸出的程式現象，並不能說明這種現象作為人的一種生存表現向度，可以等同於機械的複製功能。但是，這樣的對照性理解，可以較清楚看到人作為擁有主體意識的生物個體，以及人工意識作為可能擁有客體的人工意識的物質個體的差異。

塞爾根據圖靈測試的構想，假設了一個「中文屋」的思想實驗，說明人類心智的複雜程度是遠遠超過功能論所言及的複製狀況。在這個思想實驗中，他假設該系統在一個屋子裡，有一個小縫隙與外部相通，內有一個只懂英語的人和一本英文的手冊，以及各式各樣的紙；這個人將縫隙出現一連串的中文語言符號的紙片，依循英文手冊的說明，找尋所對應的符號，再轉錄到一張紙上回傳出去。塞爾宣稱，屋子裡的人不懂得中文，但可以根據指令，以流利的中文回覆，然而，「理解」對這個人來說，並未發生，系統回應只是正確的輸入輸出行為，並不能成為一個心靈的充分條件（塞爾，1980, 75-78）。

塞爾的中文屋的思想實驗點出一個具有功能性的系統與輸入輸出的程式，並不能等同人類心智狀態，如果程式必須執行像人類神經元中由低階物理過程所引起的高階突現特徵，不能只有執行的功能性，而是必須要能擁有與人一樣相同因果能力的架構，才能對等而論。塞爾的中文屋思想實驗，有助於從人類智能產生於生物性實體、以及人工智能產生於非生物性質實體的事實，更專注於「理性」功能於多重實現性中的不同架構的系統運作事實。

也就是說，人工智能像人一樣思考的系統與使用計算模型（演算法）的「理性」功能，即使有可能使得人工智能擁有人工意識、像人一樣有可以深度學習與進行決策等系統思維，但是，人的智能不管來自有機生物的大腦實體或是心靈的精神內容，就是與人工智能來自電腦程式的語法實體不同。賽爾的「中文屋」的思想實驗，雖然只是一個思想上的推論，但是，對於整個實驗的說明與過程推論，卻可以提醒我們對於「理解」之於心智運作的關鍵位置，以及有機

生物來自大腦實體或是心靈精神所產生的「理解」，與人工智能來自電腦程式的語法實體而產生不需有理解而能達偵一致的等同現象，我們仍有懸而未解的問題存在？「理解」之於人的心智運作的肯定價值，以及「不需經過人類理解過程」而仍可以達臻相同結果的應用效能？我們的選擇與判定是甚麼？

賽爾的「中文屋」思想實驗的時代，人工智能技術仍停留在初始發展的階段，但隨著圖靈的構想實踐越來越能擁有整全與普遍深化的發展結果時，人工智能—特別是模擬人類知能運作的強人工智能，作為人類的「理性代理人」的他者存在，以及這個他者進入人類文明社會與系統體制的廣泛性與結構性影響。這已經不只是賽爾在「中文屋」思想實驗中的心靈條件判定問題而已，而是人類文明社會歷史中、相對更複雜的技術革命改變社會結構的文明進步結果。

我們不妨循著塞爾的「中文屋」思想實驗進路，將之繼續推進為一個對於人類未來社會發展的隱喻想像實驗，將「中文屋」作為多種人工智能應用於人類社會各類型工作職業的隱喻，「中文屋」裡原本是人所從事的工作，應該/不應該被人工智能所取代？亦或是從中庸之道提出合作模式解決前者是否被取代的尖銳衝突？不管選擇為何，任何的可能選擇都應該被納入充分的討論中，不宜理所當然地「視人工智能為以未來事實」而全盤接受。這也是人工智能通識教育必須突破目前過於偏重科技思惟或技術導向的教學內容，而需要有更多傳統人文學者願意進入人工智能的通識教學領域，引入不同人文知識立場與倫理價值選擇的多元思惟與思辯議題，形成教學設計核心的重要原因。從人文知識的形成與價值，探究人類知能中的理性與工具理性的現象與本質，也許是一個值得借鏡探究人的智能與人工智能異同性的理解框架。

理性與工具理性同樣作為人類表現主體的運作方式，所涉及到的關鍵不是表現的現象，事實上，理性與工具理性的運作，是屬於心智運作隱而未顯的過程；通常顯明而可直接觀察的結果行為現象，很難馬上可以推測、判定是基於理性或工具理性。理性與工具理性的差異性，不防可以藉助於康德在實踐道德批判脈絡中，將理性區分為純粹理性與一般理性的概念—純粹理性與一般理性都可以導致道德行為，但有兩者卻有價值上的天壤地別；純粹理性的實踐來自為道德而道德的動機與義務行為，而一般理性則否，涉及人為計算功利的非純正性。純粹理性的運作關乎道德人格自我認知的完整與獨立性，而一般理性則是將理性工具化，視道德行為為道德實踐，缺乏對行為動機的自我審視與道德意識的自我認識能力。

康德對於人類道德理性的批判，提供我們對於看似一樣的現象行為、但因動機的運作原理不同而產生不同價值的認知判斷的合理性思考，應用在區分人類本尊應用理性與人工智能作為人類「理性代言人」的「理性他者」之間價值判斷的類比推論上。人文主義之於理性與科技主義之於工具理性的價值認知與選擇，有助於更中立化審視理性之於人的智能與理性他者之於人工智能的開放觀察與其之後的價值選擇問題，特別是人工智能與其產生的「理性」現象與其產生的高效能與社會應用，人工智能可資參與或改變的社會結構，究竟是朝向哪個方向發展？理性的主體與理性他者的客體，以及兩者之間在應用過程所產生的複雜的經濟、文化、社會生產與互動，其倫理規範的一致性與合理性，都涉及到人工智慧作為現代方法與接受應用之後對人類自我認知的挑戰。

其中，作為人類知能與其結構理性功能的他者，是否足以取代人類知能與理性功能，或是應該／不應該取代人類知能與理性價值，都涉及到人工知能科技研究發展與倫理界限設定發展的根本爭議。而人工知能倫理的學習設計，之所以不能被排除在人工知能的認知教育課程，就是因為人類社會自工業革命，開始啟動以現代化與相應的進步作為人類文明核心的歷史進程，並未解決人類的生存資源分配與相應創造出更好的社會發展型態，反而衍生出更多領域相對不公平與發展失衡的現象；而相關社會而到了人工知能時代，更是極有可能將人類的自動化技術與電腦資訊的應用，達臻至人類文明前所未有的質變可能，甚至正如既是數學教授、也是科幻小說家的凡納・文區（Vernor Vinge）於 1993 年所預測的毀滅性發展趨勢:「在 30 年，我們將擁有創造超人智慧的技術方法。其後不久，人類時代將會結束。」（Russell, 2011, 26-16）。

因此，人工智能也像其他新技術一樣，不能只專注在發展，仍須要有同時進行是否應該發展的風險評估與道德規範，才不會引發未來人工知能取代人類自主社會的恐慌，或是理所當然接受人工知能具有等同於或凌駕人類精神文明的價值認知觀點。而人工智能通識教育課程的倫理議題設計與人文價值為核心的討論框架，有助於探討人工知能與（人類）真正知能之間的邊界，以及從人工智能作為模擬人類知能的他者反思「人」的主體性意義與價值。

人的主體性的普遍認知關乎「人」的意義與價值判斷，也區分了人在地球生物群中的特殊定位，其中，人的自我意識、思考、道德等非關生物性身體本能的特殊存在現象，成為人與萬物之分的尺度界限。如何從人的主體性意義與價值為起點，審視定位人工智能的「模擬機械」（virtual machine）與其產出仿人類行動的「弱人工智能」、仿人類意識的「強人工智能」，包括這些人工知能模擬人類行動與心智運作過程所出現的類人類勞動與類人類意識的價值定

位，以及其進入人類經濟、人際社會之後的運作效能，且能依此研究認知，進而能合理制定出可回推於生產者與使用者彼此應該遵守的經濟與社會倫理規範，甚至研訂相關法規及對應法律制度。

這些相關人工智能本身與其相關經濟與社會倫理規範的度量衡，都關乎人工智能之於人的模擬程度與相似性，以及提供吾人審慎思考人工智能本身的商品化與商品經濟化之後、對於原本現代社會專業分工與技術的取代評估。同時，專業分工作為人類維持社會文明運作的一個驅動力，不管是「弱人工智能」或「強人工智能」，甚至未來人工智能技術突破奇異點後所出現的人工智能人，這些與人的主體的本質差異到底是甚麼？如何重新認真探索人的主體價值？而能透過人的主體探索與肯定，檢視人工知能所產生的類主體現象與價值評估。

從這個思考進路來看，人工智能的人工意識與其類人類自我意識及學習心靈的客觀開放特質，可以成為探問人工智能之於人類智能的異質性起點。Igor Aleksander 提出的 Magnus（Multi Automata General Neural Units Structure）計畫，以及 Magnus 作為學習能力屬性的神經性軟體，可以將之看成是進行學習心靈特性的開放架構的想法，並在 1994 年撰寫《A.I.人工智慧—不可思議的心靈》（Impossible Minds: My Neurons, My Consciousness）的科普專書，嘗試從一個虛構的「Molecula」故事啟迪想像、並對應真實的 Magnus 計畫執行「神經元如何發展意識」的主軸探討，與讀者分享與 Magnus 一起工作後，「開始覺得能夠了解我的神經元如何導致意識的產生」的特殊經驗（Aleksander, 2001, iv-v）。

Igor Aleksander 指出了人類作為一個個獨立的有機體、從發現自我與他人、並產生意識，有其限制在「被包裹與私密的東西。我們都生活在自己的繭裡面」的困難點，與 A.I.的 Magnus 作為非有機體的他者，他能「真切看到 Magnus 於任意時點思考的總和」的圖像化過程，完全不同。他描述了這樣的經驗，並得出 Magnus 之所以可以產生「類人工意識」的合理推測：

> 一言以蔽之，我能進入 Magnus 的繭中，並開始推論成為 Magnus 會是怎樣。同時也因為我建構了 Magnus，我知道它能如何運作，也能推測它將會怎麼做，我還能提供 Magnus 心理世界為何是像它那樣解釋。所以無論 Magnus 的意識可能屬於哪種層次，我都能基於對它的設計與經驗而提出解釋。（Aleksander, 2001, 11-12）

Igor Aleksander 的經驗描述與其相關研究論述，提供了透過 A.I.作為人類觀察自我意識的類自我的他者的可能性，並且以此為可觀察的他者，進而使得原本難以證明或陳述的關於人的主體條件建構的重要概念，如自我、意志、質感（qualia，心靈經驗的品質），可以透過 Magnus 的類人類神經元運作與類人類學習方式，揭露出 Magnus 在進行模擬類人類意識化的過程，以及設計者根據模擬設計所做出的現象心理化詮釋，進而合理推測 Magnus 有其所屬的類人工意識與類人工心靈。

Igor Aleksander 透過推測 Magnus 的類人工意識與類人工心靈的合理性，試圖展開對心理學，對語言學、、對哲學領域等重要與人有關的領域對話。Igor Aleksander 的立論與探究進路，指出目前人類觀察自己主體形成、與觀察人工智能本身的類人類主體的客體存在的最大界限，在於人類對於自身意識理解的限制，以及目前 A.I.技術尚未能實踐.的自由意志化。而對於 Igor Aleksander 以 Magnus 類人類神經元網絡的機體的學習模型，作為人工智能有其類人工意識與類人工心靈的推理預設，以此形成「人的意識」與「A.I.」類人工意識的開放討論框架，為意識之於人工智能的通識課程與人工智能的倫理探究，提供了一個很好的議題化設計與跨領域對話的基礎，探究人與人工智能之於其產生意識與心靈的條件，與其相似現象描述、但有不同本質的各自意義探討。Igor Aleksander 的經驗描述與研究論述，從人與科技互動立場所提出的類主體的他者的存在事實，也指出人工智能的人工意識探究之於人類自我意識與心靈學習的開放式觀察框架與設計教學的想像空間。

回到人工智能的人工意識探究之於人類自我意識與心靈學習的開放式觀察框架的探討議題，人工智能透過模擬人類理性思考、實踐人類理性行動的「理性代理人」的現代方法，以至探討人工智能的「理性代理人」的客體身份與定位認知，都涉及到人工智慧的發展歷史與逐漸成熟的應用技術。人工智能與過去任何人類文明的發明物不同的地方，在於人工智能對於人類的「理性」特質的實體他者化與強化應用效能。「人工智能人」也成為「人工智能」領域中相當具有吸引力的極大化整合夢想。這個夢想如果被人工智能技術實踐出來，人工智能的人工意識與人工心靈能不能從類人類的客體化認知，取得人工知能本身的主體性位置？人工智能的類人類理性與其實體的純粹化，是否可優位於人類的理性主體？兩者之間的倫理邊際又該如何界定？

改編布萊恩・奧爾迪斯（Brian Aldiss）小說的著名電影《人工智慧》（A.I.）（Spielberg, 2001），講述一個被設計相信自己就是人類的智慧型機器人，其所遭遇無法理解自己最終被主人—母親拋棄的命運（Russell, 2011, 26-18）。這個

故事挑戰人工智能人作為資本主義經濟商品定位，與其被設計成具有自我意識的主體屬性的「人」，兩者之間所產生的矛盾問題。故事走到最後，揭露這位智慧型機器人小男孩，之所以遭到母親拋棄命運的原因，在於他並不是自己依照設計所宣稱的「獨一無二」，而是機器人工廠透過技術「大量複製」的其一商品。

這個故事之所以觸目警世，在於當人類的科技技術可以突破過去人類知識文明累積對人所下的定義，並將之商品化與交易化時，「人」作為一個主體的意識與心靈獨特性，將如何被看待？故事中的智慧型機器人的純真小男孩形象，更加凸顯人工智能超越人類「理性代理人」的界限，越位到有關情感、意志的人際互動與關係中，人的意識與心靈，類人擬真到幾乎能等同與真人程度的人工智能意識與心靈，又該如何界定彼此的身份與倫理規範？人類對機器產生的情感，只是人類自我情感投射的個人問題，無關乎倫理；但當機器可以超越機器本身限制，像人一樣擁有自由意志時，人應該如何對待一個「有自由意志的機器體」？而這個擁有自由意志的機器體，是否可以因自由意志而自主決定自己的命運？甚至有權主張取而代之？

這些問題雖然目前都因技術限制而未能構成問題，但這些問題都指向人工知能機械對「人」的定位與範疇的「代理程度」的底限規定，也讓我們不得不思考，人類是否有其必要持續挑戰人工智能領域在未來發展「人工知能人」的欲望與技術？這是人文主義者對於「人」的主體本位，或針對「人」的內涵本位探討時，必須審慎向人工智能朝向技術奇點發展過程提出的重要倫理問題思考與提醒。

叁、結論

人工智能的人工意識、人工心靈與其未來可能演化的人工自由意志，作為人思索主體屬性與價值意義的參照框架，確實可以幫助吾人重新檢視人對人之意識、理性與心智形成的特殊性。然而，人工智能不只作為 21 世紀的主流科技發展，也作為一種現代方法的的全面實踐，大學通識教育如何回應人工智能、並提出有益於大學通識觀點之下的人工智能的知識判斷，人文觀點的反思設計是絕對不能排除在外，並且要積極引入相關對人工智能的「理性代理人」現象與本質的探索，形成重要的倫理議題，進行思考探索。

其中，從人類意識到人工智能的人工意識問題的探究學習，就是一個很值得導入人工智能倫理課程的設計路徑。學習者可以透過「意識」的提問，觀察

人工智能模擬人類心智而出現的人工意識與其相關現象，以及發展願景中，可能對人的主體性認知所造成的衝擊或困境。但是，另一方面來說，這也提供吾人如何通過通識教育中導入人工智能倫理探討的重要切入點。

參考文獻

Stuart Russell (2011)。《人工智慧現代方法》。臺灣：全華圖書。

Igor Aleksander (2001)。《A.I.人工智慧－不可思議的心靈》。臺灣：楊智文化

笛卡兒（2018）。《沉思錄》。臺灣：五南出版社。

J.R. 塞爾（1980）。〈心靈、大腦與程序〉。《人工智能哲學》。上海：上海譯文出版社。

Amblin Enterainment.(Steven Spielberg). (2001)A.I. Artificial Intelligence (also known as A.I.) [DVD]。