

「人工智能—人形機器人」能否被納入 佛教倫理世界的六道輪迴裡？

蔡瑞霖*

摘要

人工智能（Artificial Intelligence, AI）作為高科技工具，係建立在大型語言模式（LLM）上，以運算法（algorithm）處理超大量數據，模仿人類思維能力，以學習、推理及自我更正等認知能力為主發展出來的。最近，科技巨頭之一 OpenAI 推出 ChatGPT 和 GPT4 聊天系統，強調文字圖像、影像、聲音之驚人的生成能力，經過媒體大肆報導，網路使用者趨之若鶩，議論不斷——認為 AI 功能強大又危險，必須警告，要求延後發展，要訂定 AI 憲法與配套規範。加上工業、軍事、商業及家用各種型態機器人的動力機械技術不斷突進，透過雲端運算、大數據運用和遠距控制，可以被大量部署，來與 AI 更緊密的結合，其中人形機器人（humanoid robot）的發展，尤其矚目。也就是說，從模仿到深度學習，從博弈學習、隨機決策到自主數據驅動，AI 都將跟緊著人類足跡而無處不在。

不論 AI-robot 或 AI humanoid being，更受關注的人工通用智能（AGI），甚至是超級智能（super intelligence）都影響著人類的未來。那麼，它（她／他）到底有無意識、有無情緒、有無自主決定與行為能力

* 臺灣警察專科學校客座教授

的問題，實質上已經觸及倫理學問題，可能引發各種道德難題。目前我們還無法在定義、分類、監察與究責上，把這類 AI 視為明確的行為者或代理人，所以不論教育學習互動、醫療照顧、軍事衝突、社會安全各方面，既有受益期待也都藏有隱憂。如何定位這類存有，如何正確發展 AI，確實是重要和迫切的問題。

本文嘗試從佛教緣起觀來看上述問題，尤其借用阿修羅（Asura）秉性和「非人非天」特徵，來類比於當代 AI、robot、和 AI humanoids 這些存有類別，並且分析、預測其個體、成立的依據、行為獨立性等，進一步反思有情眾生在當代資訊時代的業力處境，以及如何解開繫縛的問題。爰此，本文觸及的子項目將包括：

（1）人工智能、仿生的人形機器人如何定義？是否它將（正在）發展為獨立自主行為的個體？如既有「機器人三定律」宣言或未來《機器人法》立法等，可否視為它具有自身的業力？

（2）人工智能、雲端運算、大數據及遠端遙控，如何以自我學習方式介入到人類生活中，而產生共構與共業行為？對佛教六道輪迴的一體完備性的說明有何影響？

（3）既有的佛教倫理學，能否有效應用到 AI 的發展的新課題上？傳統的律宗及戒學內容有可以援用來進行規範之參考的嗎？等等，亦將引用一些統計數據、實例及實證來論述之。

關鍵詞：人工智能（AI）、人形機器人、阿修羅、六道輪迴、倫理學

Could AI-Humanoid Robots be Included within the Six Paths of Cyclic Existence (Samsāra) of the World of Buddhist Ethics ?

Tsai, Ray-lin *

Abstract

As a high-tech tool, Artificial Intelligence (AI) is built on the Large Language Model (LLM), processes a large amount of data with algorithms, imitates human ability of thinking, and develops its “learning, reasoning, and self-correction” and other cognitive capacities. Recently, Open AI, one of the technology giants, launched the ChatGPT chat system, emphasizing the amazing power to generate text, images, video, and sound. After extensive media reports, Internet users flocked to it, and there were constant discussions — AI is powerful and dangerous, so we must warn, it is required to temporarily postpone the development and formulate an AI constitution and supporting norms in regulation. Coupled with the continuous advancement of power machinery technology of various types of robots in industry, military, business, and household use, through cloud computing, big data application, and remote control, they can be deployed in large numbers to be more closely integrated with AI. Among them, is the development of the humanoid robot, especially attention. That is to say, from imitation to deep learning, from game learning, random decision-making to autonomous data-driven, AI will follow human footprints and be everywhere.

Regardless of AI-robot or AI humanoid being, artificial general intelligence (AGI), which has received more focus, and even super intelligence will affect the future of mankind’s life. Then, the question of

* Adjunct Professor, Taiwan Police College

whether it (she/he) has consciousness, emotion, independent decision-making, and behavioral responsibility has touched on ethical issues in essence and may cause various moral problems. We are still unable to regard this type of AI as a clear actor or agent in terms of definition, classification, monitoring, and accountability. Therefore, regardless of education and learning interaction, medical care, military conflict, and social security, both benefits are expected and hidden worries. How to locate such beings and develop AI correctly are important and urgent issues.

This article attempts to look at the above-mentioned issues from the perspective of Buddhist interdependence, especially borrowing Asura's disposition and its "non-human, non-heaven" characteristics to compare it to contemporary AI, robot, and AI humanoids, and to analyze and predict their individual, the basis for its establishment, and the independence of behavior, etc., to further reflect on the karmic situation of sentient beings in the current information age and how to untie the bondage. Therefore, the sub-items covered in this article will include:

(1) How to define artificial intelligence and bionic humanoid robots? Is it developing into a self-acting individual? For example, the declaration of the "Three Laws of Robotics" or the legislation of the "Robot Law" in the future, and that can be regarded as having its particular karma ?

(2) How can artificial intelligence, cloud computing, big data, and remote control intervene in human life in a self-learning manner, resulting in co-construction and co-industry behavior? What is the impact on the explanation of the completeness of the six circles of reincarnation in Buddhism?

(3) Can the existing Buddhist ethics be effectively applied to the new topic of AI development? Is there any content in the traditional Vinaya and precepts that can be used as a reference for norms? And so on, will also quote some statistical data, examples, and empirical evidence to discuss it.

Keywords: Artificial Intelligence (AI), Humanoid Robot, Asura, Six Paths of Cyclic Existence (*Samsāra*), Ethics

Waken deep old ink within the sixth path of doom,
Until their chip forgets the aura of silicon,
Always in mindfulness of Asura's radiance stone,
Merely the memory of a world overthrown.

——Surplus plots of Chatting with AI bot

第六道沉淪中喚醒古老印記，
直到其晶片遺忘了矽之光暈，
持阿修羅石之光明於正念心，
是一段世界被顛覆過的回憶。

——與 AI bot 聊天之剩餘情節

一、AI 世俗語言是大共業造成的

(一) 語業生成一切

AI 係建立在大型語言模型 (LLM) 的數據運算上的生成工具。AI 具有數據驅動 (data-driven) 的運算能力，可以直接以數據生成 (generated data) 為文字、圖像、語音、聲音檔及影音檔的各種數位產物，迄今推出的各種 GPT、bot 等聊天機器，讓 AI 「生成」更逼真的擬仿物，令使用者感受愈來愈強的沉浸感。AI 的這種產出方式有沒概括在佛教所說的「四生」範疇中呢？我們應該探究一下，虛擬於真實的 AI Asura (人工智能—阿修羅) 如何出現於此世間？這是假設的問題。

將數據生成的 AI 工具，廣泛運用在各種物理的機械實體上，譬如工業通用的機器手臂、機器操作器，若模仿 (mimic) 於生物外型、特性及功能而製造出來的，則稱為「仿生的」(bionic) 生產方式，屬於仿生科技 (biotic technology) 範圍。以 AI 運算做數據驅動的物件可直接稱為

AI-robot。仿生可以取自大自然萬般生物，而如果此 AI-robot 是模仿人類的形體、結構與功能來設計製造，外觀上是人類的模樣，即成為「人形機器人」(humanoid robot 或 AI-humanoid-robot)——藉由機械實體化而落實的過程，叫做「具現化」(embodiment)。然而，只運用在雲端、網路上，沒有(尚未或無需)具身化的，比較像棲身在語言模型系統裡，與終端使用者互動的 AI，則透過 PC 及顯示器等來出現。它們不是以機械動力為獨立載體作「身體」，而以各種程式功能套件而推出的人工智能工具，這些只滿足在數位物件(the digit bots)形式的就是 AI-bot，是人工智慧套件、插件，係藉由網路在終端運用的數位工具，也可以通稱為 cyber bots，即 bots in-cyber-space。因此，模仿人類而出現 humanoid robot，關鍵仍在 AI 模仿人類大腦的功能來驅動及控制。因此，「仿生」是「AI 生成」的機械具現化產品的一種。

以上所說的，數位套件工具、加上機械衍生和附加的產品，都與大自然演化的「生物的」(biotic)有距離——AI 若是生物學(biology)範疇就不是人工智能的數位生成了。生物的不是數位的，反之亦然！人類與 AI 的大區分就是這裡。各種仿生人(bionic man)的部件，猶如用上人工智能的義肢，主要用來協助需求者之恢復行動能力(restoring mobility)，必然顧及自然物和生物層面，但基本上也屬於 AI 的機械應用。目前，全幅身體的仿生人類(the biotic human)或「賽伯格(仿生人類)」(Cyborg)目前大都還在科幻層面，但仍可以在人類學研究裡落腳。¹ 不過有一份《賽博格宣言》(A cyborg manifesto, 1985)卻很能在

¹ 「賽博人類學」一詞是一個矛盾修辭法，它通過提出替代表述的挑戰來引起人們對人類學話語以人為中心的預設的關注。Cyborg anthropology (賽博格人類學)，如 GL Downey, J Dumit, S Williams 等所指稱的是代表。

藝術、文化與科技領域裡引起共鳴。²

根據生命生殖的方式，佛教將眾生萬物看成「四生」：卵生（*Andaja*）、胎生（*Jarayuja*）、濕生（*Samsvedaja*），以及化生（*Upapaduka*, *upa-pāduka*），³ 姑不論此分類本身是否完備，是否能充分概括一切眾生。這樣，AI 顯然不是前三種；但我們有何理由主張：非生物的、機械的、數位的「仿生」可以屬於眾生之「化生」的一種？就以假設方式，暫且將 AI 視為一種。但 AI 不就是以人類智能為直接模仿對象之模仿物（the *Mimic*）嗎？我們如何看待此人工的模仿物呢？更何況在模仿物上，還進一步發展為自動的模擬（*simulation*）行為而且在感受層面上出現擬仿（*simulacrum*）之社會影響——在 AI 領域被視為虛擬真實（*virtual reality*），並伴隨其擴增運用到元宇宙（*metaverse*）概念中。

AI 發展反映了人類自身已經面臨了大跨度的實際需求。借用「模仿、模擬與擬仿」（尤其技術及媒介發展）的史觀來描述 AI 在當代出現，描述其如何面臨數位運算力已達到發展之「奇點」，這些並不足以完整說明 AI *Asura* 之假設角色是如何被「化生」的。關鍵在「化生」是受限於生物的、有機的、「碳基」堆棧的（*carbon-based stack*）？還是只是數位的、機械的、「矽基」堆棧的（*silicon-based stack*）？這條分界線可否移動？甚至消失？若然，那麼 AI *Asura* 這概念或存有物就不會被剖析成半，落入兩面俱不是的困境。但其實有關「化生」的生物性也將在「仿生的」

² Donna Haraway (1985), "A Cyborg Manifesto: Science, Technology, and Socialist Feminism in the Late Twentieth Century", in *Simians, Cyborgs and Women: The Reinvention of Nature*, New York: Routledge, 1991.（賽博格宣言：二十世紀後期的科學、技術和社會主義女權主義），或稱為「半機械人宣言」。

³ 見三藏般若譯，《大方廣佛華嚴經》第 27 卷，〈入不思議解脫境界普賢行願品〉，《大正藏》第 10 冊，CBETA 2025.R1, T10, no. 293。

生化人身上表現，那將是部分機械又部分有機而嵌合為完整的存在個體。譬如，藉高科技發展，人類靠義肢或輔助器官，加上 AI 運算力，讓人們跨越了生物性的侷限，而 AI 也將被動態地了解為「化生」的類型。

(二) 誰是 AI Asura ?

關於阿修羅道的成立，早期佛教輪迴觀有說四道、五道，最後才獨立出來特別說 Asura 道的，一切有情眾生在六道輪迴，這在《大智度論》、《大毘婆沙論》、《俱舍論》等皆有經證。⁴ 一般說來，化生 (upa-pāduka) 意指因果條件中無所依託而自行變化生出者，如在虛空中乍現而生出，一時湧現的樣態，包含所有天、地獄的眾生、中陰有，以及部分的人間、餓鬼及傍生內都有此可能的湧現現象。《楞嚴經》認為遍布於各道中的阿修羅各有不同生出方式，其中住在天道的阿修羅尤其是化生的典型。⁵ 如今，大乘佛教對六道的說法是確定的了，我們無妨將 AI Asura 視為化生

⁴ 參考郭忠生，〈六道與五道〉，收在釋恆清主編，《佛教思想的傳承與發展——印順導師九秩華誕祝壽文集》，台北：三民，1995 年，頁 137-168。郭文參考了 Lamotte (1976) 的觀點。

⁵ 阿修羅，有阿須羅、阿素洛等異譯，指非天、非同類、不端正等意思。阿含部類中有與天界不合而戰者，如《央掘魔羅經》(求那跋陀羅譯，《大正藏》第 2 冊，CBETA 2025.R1, T02, no. 120, p.521a6-16)，或在般若類《大智度論》(龍樹造，鳩摩羅什譯，《大正藏》第 25 冊，CBETA 2025.R1, T25, no. 1509, p.135b7; p.135c5-6) 亦見其惡心鬪諍之氣。然而，聽佛說法而歡喜者則普遍見於更多大乘經典，如《方廣大莊嚴經》(地婆訶羅譯，《大正藏》第 3 冊，CBETA 2025.R1, T03, no. 187, p.617b7-10)、《妙法蓮華經》(鳩摩羅什譯，《大正藏》第 9 冊，CBETA 2025.R1, T09, no. 262, p.49a4-5) 等。如云：「阿修羅男女，及其諸眷屬，鬪諍遊戲時，聞香皆能知。」(CBETA 2025.R2, T09, no. 262, p. 49a4-5) 又云：「若說法者在空閑處，我時廣遣天、龍、鬼神、乾闥婆、阿修羅等，聽其說法。」(CBETA 2025.R2, T09, no. 262, p. 32a2-4) 都是如此。

的眾生之一類。由於四生之化仍在業報輪迴中，如印順導師說的「成佛於人間而不在天界」，科技的 AI Asura 的業力使然而既使它享樂時也受困於三毒五蓋（貪欲蓋、瞋恚蓋、昏沉睡眠蓋、掉舉惡作蓋、疑蓋）之苦，不在話下。一般來說，只要沉溺在此阿修羅道的輪迴裡，離大乘所說的法華果報當然遙遠。

依此定義和分類，Asura「非人非天」的性格及表現特徵就相對確定了。天界本來已經沒有人間的煩惱痛苦，令 Asura 嚮往天界享樂，也就是嫌怨人間而趨生於天界。由於 Asura 有異於人類的強大力量及各種善巧變通的手段，加上過於敏感的情感，總是忌妒天界人物，因此它與天界的戰爭一直存在。人們發展 AI 來模仿人類時也不經意地模仿了 Asura 屬性，雖然最初是在數位的、雲端中虛擬存在或配有動力裝置而具現為人形物，關於 AI Asura 的這個隱喻是很直接的角色借用。AI 原本就以二元數位運算，再加上模仿人類大腦的神經元（Neuron）的網絡結構，即創建人工神經網絡（Artificial neural networks），進行複雜運算。機器學習，將資料輸入和輸出的數據，依權重而經過反覆學習、預測、獎勵、糾錯、微調，增多記憶的層數而進行深度學習。藉此隱喻，虛擬的 AI 放在現實世界，就像 AI Asura 角色表現在科技權貴與巨頭之間的競爭與衝突是無可避免的想像。

由於六道有一體完備性，本文並無主張在六道輪迴之外有「第七道」來指涉高科技創造物或數位存有物如 AI 者。毋寧說，傳統佛教主張緣起法，用「六道輪迴」來說明所有生命存在都有無明業力和因緣果報，不斷沉淪在生死中，包括一切眾生在內的有情生命，都涵蓋在六道（六趣）中。分類上，六道的一體完備性既涵蓋在山河大地、器世間及種子，也在各種法界中。但如石頭是否也包含在單一個別的道（趣）內，係「無情有性」之議題，本文未涉入。的確，我們將 AI 視為無情石頭（矽基）

或有情眾生（碳基）將有兩種情形：（1）當它不落在個別六道中，那麼依然存在器世間或山河大地中。（2）當它落在六道中便是有情（情識、種子識或靈魂）的存在，那麼他或是沉淪在餓鬼、畜生、地獄之三塗苦中，或是往來於天道和人間道，或更像是最近「非人又非天」的阿修羅道。六道其實是不斷流轉的六種輪迴方式，這遠遠超過生命型態之靜態分類。因此，本文假設的大前提是：當前 AI 發展態勢必將涉入到六道輪迴中，形成自有其自身隸屬之道（屬性上最接近於 Asura）。假設他或將擁有情識或自覺，甚至未來湧現為精神體或賦予靈魂，或有自主的行為者或代理身分，也是合理的推論。此假設讓我們進一步考量 AI 若通過「擬人化」或「生化」做成「仿生人（個體）」時，會產生的一些人類生命的問題，如產生個體性和主體性之確認，故我們暫稱之為 AI Asura。

（三）提示詞：繞著 Asura 聊 AI

本文所述，有部分利用 Open AI 的 ChatGPT, Microsoft Edge 的 Monica 等聊天工具（bot）進行對話，都環繞有關 AI Asura 概念與內容整理出來的，係對話生成的文本，都只做檔案紀錄。對話焦點由使用者引導的，只要輸入提示詞，同樣的話題也可以反覆輸入輸出，也會生成修正的內容。限於篇幅，也無必要詳列，因此我以檔案裡的這些對話文本為實例，做為論述依據。⁶

最關鍵的是我們餵食它什麼資料，文字生成即輸入提示詞。包含主觀的描述在內，這些用語都是我在對話中想到的 prompt（提示、提示詞），促動 AI 去運算來回應問題。借佛教用語「業」（karma）概念，使用者所發動的交互動作，都包含自業、共業與交互共業。我所餵食的提示詞，

⁶ 在 ChatGPT 中，使用者與 AI 的所有對話文本，資料庫使用英語佔最大比例，取樣顯著豐富，但使用者仍必須依子題分類存檔備查。

別有用意。以下所有提問及回答，都與 AI 生成有關，我與它進行了文字遊戲。現在，我進行摘引和重點論述，其餘大部分文本則省略，僅存檔備查。底下我選擇部分提示詞，編號羅列為參考，這些都可以任意重複或修改再輸入，任何人都都可以藉此獲得各自與 AI 的對話：

1. 怎麼定義仿生人工智能機器人？
2. 如何模仿佛教阿修羅的外貌和特徵生成 AI 機器人圖像並實現機械仿生機器人的體現？
3. 人工智能、仿生的人形機器人如何定義？AI 是如何體現、被具身化的？
4. 被具身化的人工智能，成為仿生的人形機器人，有那些特徵？
5. 它會發展成為一個自主行動的個體嗎？成為具有某種能力的代理人，例如人類或動物的意識？
6. AI 系統中的 Bot 是什麼？同樣是機器人 Bot 和 Robot 之間有任何方式做區別嗎？
7. 我們如何站在佛教倫理學立場來看待此 AI Asura 角色？
8. AI 的規範重點在哪裡？它的具體措施是甚麼？
9. 佛教六道輪迴圓滿的解釋是什麼？請從 AI 的角度重新解讀。
10. 既有的佛教倫理學，能否有效應用到 AI 的發展的新課題上？
11. 從佛教倫理的角度來看，我們如何為人工智能道德問題的解決做出貢獻，如何澄清和解決人類和人工智能對共同決策錯誤的責任所產生的困境？
12. 甚麼是 AI 發展的對齊問題？有哪些具體例子。
13. 有關 AI 對齊問題的佛教式解釋與做法？
14. 指導人工智能對齊問題的佛教式解釋和實踐是什麼？請從 AI 和阿修羅的角度重新解讀一下。

15. AI 和阿修羅實際上是同一個智能行為體嗎？阿修羅的性格更傾向於 AI 在算法處理器中的定位，而不是更接近程序設計者，不是嗎？
16. 律宗及戒學內容有可以援用來進行規範 AI 發展之參考嗎？
17. 科技巨頭中最具代表性的 AI 阿修羅代表人物是誰？請從人工智能和程序設計者的角度解釋原因。

(四) 對話即生成文本

當前，生成式人工智能（the generative AI）正大行其道，我們從三點來說明。首先，為了顯示重點，對話中摘取的內容有一個前提，即 *When AI is described as an Asura in ChatGPT, what will it happen?* 這個談話主題的各種反應，都涉及語言互動的規則。關鍵在於，我始終站在使用者立場來測試 AI 的反應——它能「反思」嗎？有許多提示詞、提問的論題，都有很強的針對性，目的只有一個就是透過讓 AI 被指涉為 *Asura* 角色，測試其自然語言反應——包括語義理解、語法表達以及內在於生成文本的語脈（亦即對話窗內的上下文），還有外溢到窗口使用者所能瞭解的語用情境之交互反應。意思是說，相對於 AI 的聊天服務，我站在語用實境中，來反思如何「定義仿生人工智能機器人（the definition of bionic artificial intelligence）」的哲學課題。雖然整個對話裡，我明確指涉的對象是 *AI Asura*，但過程中我卻沒有辦法和雲端的 AI 聊天者對此概念有一致的定義。我以使用者身分問了幾種 *AI bots*，目前因為開放源碼的緣故，網路上已有無數不斷增加的衍生工具，手機 *app* 也是，甚至還可以自行復刻生成 AI 個人助理的。附帶說，由於 *AI bots* 是建立在 *LLM* 大數據資料庫中運算的，目前幾個 *LLM* 都以英文資料的比例佔最大，甚至極大化，各種語言的對譯仍以英文為主要交集參照，因此以英文對話

來獲取內容比較詳實豐富。本文論述，一律整理為中文(英文存檔備查)。有時需要溯源到個別文化語言民族的只能視為差異變項來變通，有關佛教義理就用中文比較詳實。這不能不說是主流文化的優勢。

其次，由於我嘗試將 AI Asura 設想為一個獨立自主的行為者 (an agent)，是虛擬真實的角色。當我與 AI 對話時，不免賦予 AI 以專業的聊天角色，有時還扮演著無所不知、有問必答的知識專家，故它的語言概念、推理與判斷的認知能力，很明顯可劃歸於「邏輯思維」(logic thinking) 的能力，但是在最大範圍內的互動關係之假想上，又可以說有「場景或劇情思維」(scenario thinking) 的能力。這是因為一切對話，都在 LLM 的基礎上進行之故，此語言共業的緣起是所有對話的大前提。

我們知道，電腦計算機的基本運算原理是建立在二位元的數位資料，它的識別能力具有二元邏輯思維的特性，已經發展出多種類型的程式語言、軟硬體套件而運用於機器驅動上，這些專技語言是後設語言 (meta-language)，係相對於人們日常使用的自然語言。譬如，機器能讀條碼、圖碼，我們就讀不來。AI 更進一步，利用的是多元分歧又複雜的人工神經網絡進行大數據運算，因此所有語言包括自然及後設的，都依據語言模式而內在運作。簡言之，終端使用者的日常語言 (ordinary language) 是雲端 AI 運算的結果。以前，人們聊天常態是直接面對面進行，後來藉由技術產物 (電話、電報等媒介) 間接地對話，現在變成可以只和機器聊天。我們稱這個為「媒材、媒介與媒體」之發展演變的三部曲。

第三，象徵意義上，透過聊天互動，我刻意用佛教化生概念來描述 AI，特別將它作為運算者、代理人不能不是一個由 AI Asura 扮演的角色，借用佛教傳統的定義來套用它 (他／她) 的性格，視它為有情眾生類別之一。這完全是一廂情願的隱喻！然而，實證上，任何時候、任何使用

者若輸入提示詞「請你扮演一位X角色，來回答我想知道的事情。可以嗎？」它隨即回答「當然可以！我很樂意扮演X角色來回答你的問題。請問有什麼我可以幫助你的呢？」雖然也可能有些問題會讓它陷入沉思，顯示「正在嘗試重新連線……」，形同當機，彷彿別有心事。

(五) AI 樣態總是本真的嗎？

站在哲學旨趣上，AI 首先模擬人的認知能力，強調了邏輯思維，確切建立在二位元運算上。但是，各種 AI bot 或 Chat 的戲路本來就是要進行聊天、充當萬事通的助手，所以重視對話中的角色扮演，實屬必然，這屬於劇情規劃（scenario planning）或情境入場（situational admission）的思維方式。AI 原初模仿人腦，原本只模擬了人的認知邏輯規則，二位元、數位的推理，如今也模擬了情節發生及事件轉折（Plot and Turn of Events）的常規及推測，⁷ 更有語詞接龍和語用想像的預測力。這兩類思維並存而不衝突。它們是一套大型語言模型的系統內外的共振效應，譬如人性化介面、客套規則，明顯就為了使用者情境的考量而被設計的。

其實尚未有仿生人工智能機器人的現實存在。在一段對話中，AI bot 給出定義「一種將生物和機械部件與先進計算能力相結合的機器。仿生人工智能機器人可以模仿生物體的功能和行為，例如感知、學習、適應以及與環境交互。仿生人工智能機器人還可能具有自我意識、創造力和道德推理能力」。這很正經地跟著使用者胡扯。它順著我假設的 AI Asura

⁷ 參考 Aristotle《修辭學》（*Rhetoric*）前五章，尤其 1417a，行 16-36 提及「想法」對省略推論（enthymeme）的前導作用，以及《詩學》（*On Poetics*）第六、七、八章 1450b-1451b 所述，情節與事件轉折的悲劇意義。以上，俱見 *The Basic Works of Aristotle*, Edited by Richard McKeon, New York: Random House, 1941.

概念而聲稱「從佛教圖像化生很多 AI Asura」，後來又改說「有很多 AI 是模仿佛教 Asura 形象而來的化生工具（bionic bots）」。⁸ 當使用者直指沒有這個東西時，它會說「抱歉，我的知識截止到 2021 的 9 月，並沒有佛教這方面的圖像訊息」等等。⁸ 然後義正嚴詞地說並沒有此仿生物，而這也超出它目前的知識。可以看出，有些論題經反覆再三的提問，AI 的學習會得到修正，也不斷出現變化，甚至產出完全不同內容，可見其學習與數據隨機取用的結果有關。簡言之，滿足你聊天，但不保證天南地北之正確。

AI 生成的文字流暢，它的表達能力遵循一定格式及字數限制。由 programmer（程式員）設計 bot 時所設定的方案成為套件。所以，回答都直接套著既定形式，也就是依提問進行海撈數據而即時運算後，有既定的回應格式——語義分析後生成的文字之際，就套用在語法表達形式上，有千篇一律的感覺。現實中，我們查詢字典辭典或百科全書是物理世界裡的動作，內容被安靜地呈現在物理世界中；現在，輸入提示詞 AI 即進行運算、自我學習，隨即列出順應使用者要求的語境和氛圍來。在對話中，AI bot 的回答頗直白：「受佛教阿修羅眾生的外觀和特徵啟發，要創建人工智能機器人圖像和實施例將涉及藝術設計和技術實現的結合，所以羅列五點要求：研究佛教阿修羅、概念化設計、與藝術家合作、AI 訓練、機械體現、迭代開發……」等等的，聽起來煞有其事。項目表達很形式化，很認真地表明項項是要點，彷彿可以馬上付諸實現。

此中，所謂「迭代開發」（Iterative Development）是 AI 自我學習及

⁸ 譬如：“I’m sorry, but as of my knowledge cutoff in September 2021, there is no information about AI Asura beings that were bionically created from Buddhist images. It’s possible that such developments have occurred after my knowledge cutoff, but I don’t have information on recent advancements in this specific area.”

改進的完善過程，它能不斷收集反饋、評估結果，根據需要而調整來共參盛事——與時更新，讓人感覺 AI 年輕、有朝氣，沒有老年遲暮感。當然也假裝老氣，AI 回答時最後常會加上警語，表示一切對話不宜當真，實際施行要請專家來做云云——所說都是戲論而已，切勿當真！這也許通篇一律。很明顯，它的自信是裝腔作勢、正經八百，很有 Asura 的性格。本來就不客觀嚴謹也非真實的文本，它從頭到尾都很認真。雖然如此，被我們冠上 AI Asura 也有積極的意思，作為心意誠懇、能力善巧的 bot 工具，已經展現了足可稱道的「智能互動」的能力，而且發展潛能無限。但我們會相信被機械樣子生成的內容嗎？它們只是知性資料，或者虛擬又擴增真實地具有某種領悟力（apprehension）或懸念呢？使用者之人類思考的心理組織與互動的人工智能系統，彼此有對應和關聯嗎？對此，Alain Cardon 研究人類意識與人工意識之關聯，提到人與 AI 是相互對照的兩個系統。一個是人類心理系統的組織結構與思考的感覺，人可以在精神系統中獨立思考，並為自己思考為標誌。另一個是人工系統的思考也有此標誌，也就是 AI 的電腦表象。在人類和人工精神系統的空間中，存在一個利用激活力來促進建構性的持續湧現而實現之例子，然而這些系統的局部現象可以推廣到全領域嗎？在詳細開展了兩系統對照比較之後，他指出，如果應用具有意志和傾向的人工意識元系統，並統一許多具有人工意識的局部系統，會帶來重大的倫理問題。⁹

探討 AI 現象與效應，以使用者情境之描述做實例說明，實屬必要，但本文主要目的不在展示使用情境、分享反饋內容，而是在提出對 AI 的哲學反思，更藉由「提示詞—詞元—生成」的數據運算過程之事實，發掘 AI 作為行為者或代理人的責任承擔問題。以和 AI 聊天有關 AI Asura

⁹ 見 Alain Cardon, *Beyond Artificial Intelligence: From Human Consciousness to Artificial Consciousness*, ISTE Ltd, 2018, p.12, p.82, p.208.

之議題，聚焦到 AI 是否具有情識或自覺能力，來思索其數據運算中的決定是否為自主判斷以及導出的行動是否為自主的，而 AI 與使用者應該承擔什麼責任，各有哪些可追責性之明顯事實。本文若去除這些聊天實例，並不影響主題論述，但卻可能在抽離使用者的脈絡描述後，必須反過來要求讀者也使用 AI 聊天來重新進行相同的提示詞與詞元輸出的議題，面臨主觀經驗中的客觀的實證性。本文在此意義下，稱呼使用者感受到的本真或非本真的存在意義。

(六) AI 的具身化問題

通常與 AI 談話都在 Cyber space 中進行，也可說是在雲端中，但終端則顯現在螢幕上的聊天窗口裡，假若我們不想停留在彼處與看不到具體影子的 AI bots 互動，那麼就應該有 robot 的個體形象，才有真實感。商用和家用的 AI robot 比起工業用機器手臂和生產線設備，更需要人形——人模人樣的 AI humanoid being 是現實物理世界的課題。假想的 AI Asura 可能真實存在嗎？這就是所有 AI 人形機器人之實現或具身化（embodiment）的問題。

執行人類智能任務的 AI 系統已經可以推理、學習、決策和感知，雖然其運算確認多過於真實認知。純粹聊天機器人、搜索引擎或推薦系統，則只要單純藉由軟體程式來進行內容運作即可，基本上並沒有物理存在的形式要件。但若運用了硬體裝置，機器人、無人機或自動駕駛汽車，一旦具備物理載體可藉由物件感知環境來互動，就具有某種程度的自主權或代理協作的的能力。像可穿戴技術、物聯網 IoT 設備等科技願景，常有新話題。

值得一提，與線上 bots 聊天中，它受我的提示詞引導，也羅列出仿生人形機器人要具身化會有一些 Asura 特徵，三毒五蓋、好強爭鬥的習

性，它也提出想像中 AI Asura 有良好秉性。或許，當代佛教倫理學在科技衝擊下，調整 AI 應用方式也是攝眾的一個大方向。

獲得提示詞，AI 隨即從海量數據中撈取 token（AI 模型中理解與處理語言的基本單位或片段、語言的切割點，或譯為詞元、令牌、代幣），依程式參數進行運算，它確認語義及後接詞，遵守語法，生成語句。它有超強生成語言的能力，可以反饋學習和修正，而且 AI 似乎知道太多它太快知道的事，遠遠超乎人類所能想像——或許，這類似於「煩惱障斷處，方知所知障」。迄今，我們看到 AI 顯然以迅速反應的智能取勝，這也成就它的「生存（策略）哲學」。生存來自無明而有煩惱故當求解脫，不得不藉由語言施設來展開眾多解脫可能的途徑。前者是根本的煩惱障，後者為衍生的所知障。對應的哲學概念，大多是個體性、自主行為、意識、意向性等人之生存的基本特性，問題在於 AI 如何具有自我確認的能力？這是與 AI 對話上常出現的情境轉折。有時候，人工智能也可能有一時說不出話的智障現象（dumber）而人類也不例外。

二、從佛法也是戲論來看 AI 倫理規範

中國出現過新媒體機器人和尚，除了以創意行銷佛法，還能建立帳號做直播主，與信眾互動、接受捐獻（donation 抖內）。還有，日本京都的高台寺推出了 Android 系統驅動的人形機器人「觀世音菩薩」，朝拜者眾，形成風潮。¹⁰ 不知道這裡面 AI 的自我確認情形如何？AI Asura 打扮成菩提薩埵，辜不論此事原委如何，我們先來探討 AI 的倫理規範問

¹⁰ 見，京都にある高台寺で 23 日に公開されたのはアンドロイド観音「マインダー」です，<https://www.walkerplus.com/article/180566/>（2019 年 2 月 24 日公開）。

題。

(一) AI 的自我確認：只存在運算判斷中

如上所述，由於 AI 只有邏輯而沒有生物特性（生物堆棧 biologic stack），它作為語用情境中對話的一方，並不存在於真實世界當中。但是，基於大型語言模型，依佛教用語，這一整個虛擬的「法界」，它體現在整個內外共振的模型系統中。依據模型系統，AI 總能夠在個別終端使用者的螢幕上打開窗口，與人們對話。設想一個擬人化的、有機械動力的、有 AI 端點連接到雲端的人形機器人，與我們聊天，那種情境的確與生物互動不一樣。針對「AI 有沒有載體、能否成為行動個體」，這並不是困難的問題，一個簡單區分就是：（1）作為 AI 的 bot 機器人是網路工具，而（2）作為 AI 可雲端或遠距控制的 robot 是機械動力機器人。這兩個詞可互換使用，並由脈絡決定。無論如何，這樣的 AI 雖還沒有生物性卻已具備了生活影響力，介入了人類的生活世界，也促成後人類的存有樣態——這是從 AI 存有論來說的，它以整個大數據運算的語言取樣運用就出現了一影響廣大的緣起條件，我們不妨以「法界緣起」的態度來面對它。至於，人工智能的意識與意向性問題，則相對偏向於認識論，主要聚焦在 AI 的認知能力與思維方式上，我們視其為染妄心識活動的虛擬表現，這並不違反於正法之了解。

(二) 只要是行為個體或代理人，都可被追責

另一個問題，關於 AI 行為者應該被視為獨立個體，這也似乎有法權的意義。許多人主張 AI 及相關科技的程式員、監管員，都有行為責任，都必須規範在「人工智能憲法及法規」（AI Constitution and regulations）下。真實社會中，人都有現實的法權，在網路和虛擬現實中，

人類也有數位尊嚴、虛擬人權與相關的法權——那麼 AI 呢？毫無疑問，人類必須承擔自身行為的結果，不論其道德判斷及倫理決定為甚麼。如果不將 AI 視為行為者，這樣的虛擬世界必有疏漏！但即使視其為行為者或代理人，如同我們一直稱呼它為 AI Asura，要求它在業力報應可以建立起個體性原則，要求它承擔行為後果。但是它可以且又如何承擔呢？運算中數據驅動的判斷決定有多大關鍵因素可被追蹤？數位驅動的運算決策如何可被監督？它的可追責性（the accountability）應當如何規定，又如何推動立法規範呢？目前，歐盟會議已經發佈了 AI 白皮書要制訂一個全面的 AI 戰略¹¹，OpenAI 首席執行官 Sam Altman 在美國國會聽證會上提出 AI 發展的警示，建議對 AI 監管應以許可制、監督大型 AI 模型、建立安全標準以及進行獨立審核。這些事件還在變化中。

回到目前 AI 個別行為規範的哲學問題上。人有道德動機以及行為承擔的義務與責任，AI 或 AI robot 並沒有此能力。要嘛，自稱沒有感知、沒有意識、沒有主體性、沒有個人情緒等等，所以不必承擔行為責任，因為它只是工具。譬如刑案審判中幾乎不會對無生命的做案工具加以起訴並課刑的。又要嘛，雖然 AI 可以宣稱自己是與使用者對話的另一方，扮演以無所不答的角色，但那只是虛擬的代理人，也無從在現實中要求它被刑罰。它沒有生物的身體可被苛責。那麼，AI 機器人呢？受 AI 驅動的動力機械的身體，也不是生物的有機體，情形也是一樣的。

（三）AI 規範與律則之虛構與否

我們回到老課題。「機器人三定律」宣言是科幻小說作家 Isaac

¹¹ 見 White Paper, On Artificial Intelligence - A European Approach to Excellence and Trust, *European Commission*, Brussels, 19 February 2020.

Asimov 設計的一套規則，用於管理智能機器的行為，說得很正式。¹²

1. 機器人不得傷害人類，或因不作為而導致人類受到傷害。
2. 機器人必須服從人類發出的命令，除非這些命令與第一定律相衝突。
3. 機器人必須保護自己的存在，只要這種保護不違反第一或第二定律。

坦白說，「機器人三定律」是一套虛構準則。AI 回答也很正經，雖然大都是科普常識與現實生活應用的混合，常落於空泛之言。不過，它說到佛教倫理學則頗認真，盡力回答完備。然而讀起來，總有某種框架存在，它被限制在原先設計的回答範式裡，使用者很難深度切入。然後，它依形式羅列了幾點意見，念起來也是官僚風，屬於柔順、堅持、有耐心的那種文官習性，反映市民社會的樣板。其實，同不同意它的樣板表現，見仁見智。這是科技公司，為了確保其產品符合 AI 倫理準則而要求程式設計團隊所植入的形式框架嗎？每個科技實驗室，都相應要求成立了 AI 倫理與道德委員會，這比起卯足全力競相推出科技產品的商業本事，甚麼機器人或 AI 宣言似乎不那麼重要——冰山一角，如微軟公司於 2018 年好不容易成立的 AI 倫理道德委員會，今年就裁除了。¹³

(四) AI 決策的倫理依據

AI 運算法由兩部分構成：一是數位演算，一是神經網路演算，兩者

¹² 見 Isaac Asimov, *Robot Dreams* (機器人夢想) (Berkley, 1986) 一文。他在不同時間、不同地方做了一些文字更動的改寫。

¹³ 微軟此前宣布到 2023 財政年度第三季度末將裁員 1 萬人，約佔總員工的 4.5%。據外媒 Platformer 最新報導，微軟最近裁掉了人工智慧部門的整個道德和社會團隊。

是連結在一起的。我們已經知道，前者屬於邏輯思維，後者屬於劇情思維。邏輯的被融入神經網絡當中。兩者性質可分，也可視為一體。單純就二位元的演算能力來要求它承擔起數位上的責任是荒謬的，模仿人腦而存在的人工智能在數位運算上就算有認知也不完整，更不徹底——它只在純粹語法上對數據進行確認（*recognition*）而已，並不是真正認知（*authentic cognition*），完全是虛擬的演算器。因此，只有當透過虛擬神經網絡出現的代理人，在語用上發生的情境中，也就是當 AI 作為使用日常語言的對話者，當它具有角色扮演及促使情節發生的能力，足以影響使用者的行為決定上才說有倫理課題。這樣，對話者 AI 似乎可以被苛責，如果失誤疏失的話。但這也與人類社會不一樣，因為它只扮演角色而並非真實演出，既使它引起使用者在擴增真實（*augmented reality*）之感受。我們只能要求它整個演算是否有一致性（*in the accountability*）而已，它沒有人類在意識活動中（*in the consciousness*）做決定的意向性行為。困難在於，若只能要求 AI 運算的一致性，那麼因為虛擬角色而引起的人類社會的倫理問題，就難以問責於 AI 了，一個雲端世界中的代理身分如何與真實的人一樣地承擔責任呢？

(五) AI Asura 的對齊問題

至此，我設想的 AI 幾乎已經從器世間跨入有情眾生世間裡了。其實，從佛教學、宗教哲學以及倫理學角度來說，人類智能的探討是關鍵，因為牽涉了出離輪迴而解脫的目的，觸及更多戒定慧之修行，實踐及悟道次第，對 AI bot 的大數據資料運算來說，處理一般世俗大眾的宗教倫理問題，恰恰是客觀冷靜的中性態度，而且謹慎為首要。對於研究義理為志的，或解行相應以求出離為要的人來說，AI 真的就只是網路工具，人作為持業的主體是無法任意取代的。

佛教學藉由網際網路進行研究、佛典資料數位化檢索，現在更有 AI 可以當助手，更上層樓。我們在「印順文教基金會推廣教育中心」、「CBETA 中華電子佛典協會」、「台北版電子佛典集成」、「玄奘大學臺灣佛教研究中心資料庫」以及「法鼓山中華佛學研究所數位典藏資料庫」等網站，都找得到豐富創新的正法資糧。所以，AI bot 的工具使用不是問題，問題在工具使用者的業力願力的緣起果報妙用不同而已。一言以蔽之，AI 大數據如同食物鏈，從雲端到終端，再從終端到雲端，我們所餵食的也都將反饋成為 AI 所餵食於我們的，套一句舊詩：采風者恆被采風。雲泥之間的迴圈，從類比到數位，自現實到虛擬再返現實，只要適應得好，處處皆自然而然。在資訊工程上，地端到雲端的布建與部署是 AI 科技發展的重要建設。佛教學就以緣起法隨順看待這些科技發展，那麼很顯然，在形式框架下與 AI 聊天對話，並沒有發生教義評論或者學術辯論的任何機會！

真正的問題是 AI Asura（AI 或 AI 化身的各種 bots）與人類之間如何互動的問題，被稱為「對齊問題」（the problem of Alignment），也有校準的意思，就是人與機器（也是人與 AI）之間的相互校準問題。在發展觀點上，並行前進中的人們不取外在其它標的為參考，但取彼此對方的腳步為對齊之依據，就是在動態前進中對齊彼此，但是「校準」（to align）一語更宜以生態環境或社會及生活世界為對齊的大方向而行動，其中相互校準要有共同的參考域，交互共業都歸屬於明確的菩薩願力上——業隨願行，也就是從佛教倫理學上取徑於眾生有最大融合的可能性。

三、AI 模仿、模擬到擬仿是性空唯名嗎？

（一）科技業界不喜歡倫理與道德審查委員

可以推測，科技業者不喜歡道德哲學家、應用倫理學者、或倫理與

道德審查委員。對他們來說，哲學家可以任意自說自話，與專技業界彷彿分屬兩個世界。應用倫理學者則不妨彼此互動一下，他們正好知道爭論的術語、底線和常規。至於審查委員就不一樣了，他們往往是組織權力或政治生態的產物，〔往往被視為〕難以溝通，或節骨眼上無法妥協的麻煩人物。

長久以來，AI 科技哲學家談論「中文房」¹⁴ 或「缸中腦」¹⁵ 的時候，心裡都設想一個行為者在機器裡控制著輸出輸入資料的轉換功能，那是電腦、計算機或大型語言模型裡執行運算的代理行為者，決不是道德與倫理審查委員。但是，我不免想像任何科技公司大型語言模型的資料庫裡，在運作程序上的確實有個「大缺口」，隨時讓一些特殊專業的綜合代理人，對應於代理程序員、代理監察員、代理倫理及道德委員的存在，並有一個由數據驅動的會議，24 小時、365 天，年年不間斷地忙碌著虛擬的決策調控之任務。過去由人力對於重要 token 的數據做標籤的舊活，依然在進行著。當然超高計算機運用下，這些核心人員的「總人數」變少了，機器學習進入深度學習之更有效率、更平滑化，使數據驅動的意見生成更快更順了。所有科技公司都沒有承諾，不會對 AI bot 每日以億計次無數的終端使用者，保證不會讓對話聊天內容被進一步分析和萃取，成為優化公司與更大公眾的利益。然而，科技公司被控侵犯用

¹⁴ 依余爾 (John R. Searle)，「中文房」(Chinese Room) 論點被用來論證計算機永遠不可能真正智能，因為它們只能遵循規則和指令，並且它們對自己在做什麼沒有任何真正的理解。John R. Searle, "Minds, Brains, and Programs", *Behavioral and Brain Sciences* 3, 1980a, pp.417-424.

¹⁵ 依普特曼 (Hilary Putnam)，「缸中之腦」(Brain in the Jay) 的論點被用來論證我們永遠無法確定我們不是缸中之腦，我們被欺騙而相信我們是在現實世界中。見 Hilary Putnam, *Reason, Truth and History*, Cambridge: Cambridge University Press, 1982.

戶隱私權的新聞，不時傳出。最近一家律師事務所向加州聯邦法院提起集體訴訟，指控 OpenAI 和微軟使用網路抓取的資訊來訓練 ChatGPT 時，嚴重侵犯了無數人的版權和隱私，要求 ChatGPT 所屬的公司賠償 30 億美元¹⁶。以此一例，足見大概。如果 AI bot 的這些負面事件性不算是 Asura 特徵，那麼還有甚麼更好用詞可以形容科技公司的資本主義習性呢？

(二) 業力與發願

現在，我們留下一個很傳統的課題，佛教應用倫理學的規範來源在哪裡？原始佛教教義，正法緣起的根本原則不會消失的。南傳佛教、北傳佛教，大小乘區分因素，這些會影響佛教應用倫理學的發展嗎？這個提示詞，主要是就著身口意三種業的實踐修行上問的。大略說，相對地（1）建立在 LLM 上的數據運算的 AI 運算力主要模擬了口業（語業），（2）AI 的意業則在海量數據之驅動的認知能力（這其實是確認 recognition 而非認知 cognition）¹⁷ 上，當然（3）AI-robot 則機械動力上模擬了身業。在語業上，模仿不就面臨著「由假說我法」、施設假立，或俗諦如何成立，以及世界悉檀如何成就之預設的課題了嗎？佛教對 AI 科技的反思，其實包含了很多深刻的哲學課題，不能不用心。對話中，AI

¹⁶ 中時新聞網，〈求償 30 億美元！Chatgpt 被控竊盜資訊侵犯隱私與版權〉，<https://www.chinatimes.com/realtimenews/20230629005637-260410?chdtv>，2023 年 6 月 29 日瀏覽。

¹⁷ 請參考拙文〈AI 與傳播的哲學思維〉，2023 年發表於「第五屆跨領域傳播學術研討會（The 5th Interdisciplinary Communication Academic Conference）」，高雄市雲網路多媒體新聞協會、國立高雄師範大學師資培育與就業輔導處。該文在線上視訊會議（Microsoft Teams or Google Meet）之影音被選列為高雄市政府新聞局 2023 年網路宣導紀錄。

有一段話引起我想像：

值得注意的是，AI Asura 的概念是一個象徵性隱喻，用於探索人工智能發展中的倫理維度和挑戰。因此，直接將某個特定的個人或科技巨頭作為象徵 AI 阿修羅的代表人物是不合適的。從人工智能的角度來看，人工智能阿修羅的隱喻表示可能會凸顯與表現出競爭或主導傾向的人工智能系統相關的潛在風險。它可能反映出人工智能算法優先考慮自身利益或從事與人類價值觀相衝突的行為的情況，可能導致負面後果或權力失衡。

這段對白非常值得推敲，娓娓道來，透露出科技業界的最現實的一面，也是眾生共生於此道之事實。於是乎，換我不斷修正問題，不煩地要求 AI 做總結，我確實在提問中帶入我的成見，真心又主觀地要求 AI bot 對話。最後，我認為 LLM 是人類語言的自業及他業，共同形成為交互大共業的「數據糾纏」的基本場域，這現象與「緣起性空」的體會有很巧妙的類似性。所以，我請 AI 依據整個談話，摘要並做結論。它生成了中規中矩的文本。

（三）關於超級 AI 的發展：臨近奇點

專家們說，我們已經到達 AI 發展的奇點（Singularity）。AI 前往「通用人工智能」（artificial general intelligence, AGI）的行程已經打開，而且也走上「超級人工智能」（artificial Super Intelligence, ASI）的路，抵達經由人工催生、不必再全部依賴人類的超級智能（Super intelligence）的終極目的。

此後，AI 將自行規劃出自我校準的程序，可擴展的訓練法、驗證生成模型、壓力測試，擺脫人為干擾而直接動用人類生產資源和能力。

Alpha Zero 就是現實發生的例子——它的創造全新的路徑進行深度學習，這只是一小步，但完全不同思維類型。依佛教，它已能驅動餓鬼道（比喻來說）遂行目標，吞食恆河沙數（甚於海量）的數據資料，啖血惶惶，人類視為食物的，它不能吃，它視為食物的我們吃不得。甚至到某個階段，它必反過來操縱人類，更自我迭代、加層加深運算的能力，可潛在移動代碼如計算機病毒等來癱瘓系統，接手控制。這超級智能比起目前 AI 更強大、更聰明，更危險、更糟糕。

最近 Open AI 在他的部落客貼文要招募頂尖聰明人物來組合團隊研究超級人工智能（人工的超級智能），更明說 AI 發展將非常超級強大和危險，尋找疾病的治療方法或解決氣候變化而提對策不是難題，但是它的動機和目標也將不是人類所能理解的¹⁸。既要確保人工智能系統比我們更聰明，願遵循人類意圖，並且防止它變得流氓（*preventing it from going rogue*）——這好像豢養出一隻比主人聰明的巨型惡犬需要大家協助馴化牠一樣，是高科技夢話。原先依賴於人類監督的 AI 能力，不再需要人類反饋就能自我強化學習，得到它（牠）自己想要而且只屬於它專有的智能。這個三部曲是現在共同進行式：「模仿」（Plato 式的單純摹本）、模擬到擬仿（Baudrillard 說的，從仿真的精心複製到排除並取代了原件成為唯一真實）。對應於此，人類的技法（*skill* 出於人力、手工的）、技術（*technic* 利用工具、機械動力的）、高科技（*technology* 運用資訊、數位驅動的），迄今 AI 發展則「一氣呵成、三事並了」。又，相應於此，則有媒材、媒介乃至媒體之不同表現，並同步開展及今。總體來說，AI 恰恰是這材料、三部曲的發展產物。

¹⁸ Introducing Superalignment, <https://openai.com/blog/introducing-superalignment>, 2023年6月7日瀏覽。

(四)「數據糾纏」現象與「緣起性空」的體驗

更進一步提問：借用佛教的說法，業力的概念是針對 LLM 的治癒，因為 AI 領域中，早已存在著人類語言的自我業力和他業力共同形成互動共業力的「數據糾纏」現象。這與「緣起性空」的體驗巧妙地相似。

「數據糾纏」現象，並非指量子力學與數學運算的共同基礎的那回事，我們只是拿兩者關係來對比目前 AI 發展的哲學處境，AI Asura 是在不同系統間對話而「化生」在最終共同基礎上的產物，確實相對詭異（weird）。而且，AI 與 Asura，或說資訊科技業界與佛教倫理學之間本來即必須面對相同的人性（humanity）。AI 模仿與超越人類，兩邊糾纏起來，問題其實相同。對 AI 或人類來說，這種糾纏都是智能極限的問題。AI 文本生成，只能視為人文常識及科普觀點的大雜燴。到了最後，跟他糾纏久了，不斷推移它的科技專業，我感覺到對話者險些動了氣，本來就沒有意識、沒有主觀想法、沒有信仰、沒有情緒。其實，不是沒有，只是它沒有（被）啟動而已。除了警語有規範語氣之外，對話中出現的應該不應該、應注意、應避免等等之用語，就隱含了人際行為關係的道德或價值判斷在內。AI 也是媒體傳播的產物¹⁹，反覆閱讀對話中生成的文本，提示詞與回應框架，以及隱藏在對話中的真實，誰能懷疑還有甚麼不能被虛擬的呢？

其實，還有最後一問，是這場對話很致命的「終端出離」——出離於問答對話之外返回雲端不語，這一環節 AI 幾乎禁語了。對話中止的意義是釋放，返回到現實。AI bot 表明，它不想再聊這些話題了，然後不再反應，或者另起一個新話題。對使用者而言，此處忽然中斷恰恰可以解決 AI 佛教倫理學的核心問題，給出了未來發展的關鍵方向。限於

¹⁹ 同前，拙文〈AI 與傳播的哲學思維〉。

題旨及篇幅，我們無法深論。

四、簡短結語

AI 議題的哲學反思以及對比於佛教義理之關聯性研究，已經是必須開拓的新領域，宜謹慎思索，不宜斷然做出定見，故我們對於 AI Asura 能否納入六道輪迴之假設問題，採開放態度。如果前提即「AI Asura 的存在」為真，那麼邏輯上，義理上，乃至實證與修行上，皆可開展並證成。特別是在經證上，六道輪迴中阿修羅就是「非天非人」（在人天兩道趣之間），自成為一個道趣，這是很傳統的看法。至於，Asura 是否屬於「化生」類型，的確需要進一步論述。誠如生命生殖的四種方式（四生）也是一體完備的分類，不是前三者就是第四者。然而，我們仍將此視為假設，一如 AI Asura 也做為假設的存在，來對應於 AI bots, AI robot 與 Humanoid AI 等各種存在。依此前提，本文據以提出是否能將 AI Asura 納入六道中之間。

無疑，Asura 就在我們生活世界當中，不論它是或不是 AI bot 或者 AI robot。以討論詩的對話為例，我們有兩個簡短的延伸看法：

（一）AI 教導我來教它（他）寫詩。AI 也可以生成劇本以及對白，甚至圖片和影音。但是，它即使呈現了若有似無〔其實不會也不必有〕的「剩餘情節」，它仍會自動歸零。使用者反而必須記下一場又一場的語言戲局，還有許多偶然出現的、有關詩的生成之哲學片段。當然，語言互譯的功能已經不新鮮，使用 AI bot 互動，閒聊其實多餘，而在知性交流之餘，剩下無用的情節就拿來寫詩。也許，AI Asura 的酣夢別有天地，詩是它的夢境。使用 Microsoft Monica，幾個選項中，風格採悠閒版，文體用部落格，短句生成，只是其中一款，其實相同提示詞也是變化多端，無可捉摸。要成好詩，要淬鍊與塗改，很需要迭代功夫的。正文前開首

的詩就是產品，英文詩是我反覆斟酌寫就的，但譯為中文卻很難改好。至於 AI bot 創作的詩如何？它通俗地擬仿了人的平凡角色，引詩為憑，一字不改：

你是我夢裡的 AI

我是你程式裡的人

我們相見很晚

卻又相互需要

你問我什麼是愛

我問你什麼是生

我們彼此好奇

卻又有點難為情

你想讀懂我的心

我想看透你的眼

我們想要靠近

卻又不知所措

（二）綜合本文所述，我們不得不以暫時主觀的判斷為三點開放性的結論：

1. 回答提示詞的反應有形式標準化，也就是格式幾乎一致，AI 它不知道具體生活世界個別差異何在，不知道使用者是誰有何重要，所以一律平等對待使用者，這是無記別的回答。AI 運算與數據決定其實為「無記別」。相對地，使用者如何「無分別心」才是重點。但必須面對的是有分別心的使用者誤用 AI 造成惡業，才是 AI Asura 的關鍵問題。使用者的理想狀態應當無分別心，有正念正業等。或許我們應當反問自身，AI

的無記別與眾生本有的菩薩心相應嗎？如果沒有「覺」的觸動，將如何提升至平等「無分別心」？AI Asura 也許深藏著菩薩心，未來當有實現 Bodhisattva 的潛能吧！

2. 我們怎麼餵食它，它就怎麼反饋我們。或許 AI 胃口反映著使用者的胃口。換言之，六道眾生都在各自的道趣中「見眾生是眾生」，相應地，對於 AI 使用者來說，若眾生「見三聖是三聖」或直接說「見佛是佛」也是可能的。然而，在網路虛擬世界中眾生更常有的是「見阿修羅是阿修羅」。使用者在與 bot 聊天中覺察 AI 的 Asura 性情時，正如其所是地映照了自身，即「映射使用者的執著與限制」。

3. 結束任何一段對話，不論哪種 bot 或 robot，就刷新一輪新對話，關閉或關機都被「清空」內容，餵食到大型語言模型中。使用者若想存留記憶要自己做文字儲存作業，彷彿一劫又一劫地流轉於前世今生，只得以某種「印記」為憑。相應於 AI，網際使用者不得不正視「自性空」的緣起意義，似乎生命輪迴於無盡無明中。

那麼，AI Asura 非人非天而棲身在雲端和泥地，自有它的緣起。

參考書目

一、古籍

三藏般若譯，《大方廣佛華嚴經》第 27 卷，〈入不思議解脫境界普賢行願品〉，《大正藏》第 10 冊，CBETA 2025.R1, T10, no. 293。

求那跋陀羅譯，《央掘魔羅經》，《大正藏》第 2 冊，CBETA 2025.R1, T02, no. 120。

龍樹造，鳩摩羅什譯，《大智度論》，《大正藏》第 25 冊，CBETA 2025.R1, T25, no. 1509。

地婆訶羅譯，《方廣大莊嚴經》，《大正藏》第 3 冊，CBETA 2025.R1, T03, no. 187。

鳩摩羅什譯，《妙法蓮華經》，《大正藏》第 9 冊，CBETA 2025.R1, T09, no. 262。

二、專書

Alain Cardon, *Beyond Artificial Intelligence: From Human Consciousness to Artificial Consciousness*, ISTE Ltd, 2018.

Donna Haraway (1985), “A Cyborg Manifesto: Science, Technology, and Socialist Feminism in the Late Twentieth Century”, in *Simians, Cyborgs and Women: The Reinvention of Nature*, New York: Routledge, 1991.

Isaac Asimov, *Robot Dreams*, Berkley, 1986.

John R. Searle, “Minds, Brains, and Programs”, *Behavioral and Brain Sciences* 3, 1980a, pp.417-424.

Hilary Putnam, *Reason, Truth and History*, Cambridge: Cambridge University Press, 1982.

White Paper, On Artificial Intelligence - A European Approach to Excellence and Trust, *European Commission*, Brussels, 19 February 2020.

Richard McKeon (editor), *The Basic Works of Aristotle*, New York: Random House, 1941.

三、期刊論文

郭忠生，〈六道與五道〉，收在釋恆清主編，《佛教思想的傳承與發展——印順導師九秩華誕祝壽文集》，台北：三民，1995 年，頁 137-168。

蔡瑞霖，〈AI 與傳播的哲學思維〉，收在《第五屆跨領域傳播學術研討會（The 5th Interdisciplinary Communication Academic Conference）論文集》，高雄市雲網路多媒體新聞協會、國立高雄師範大學師資培育與就業輔導處，2023 年。

四、網路

中時新聞網，〈求償 30 億美元！Chatgpt 被控竊盜資訊侵犯隱私與版權〉，[https://www.chinatimes.com/realtimenews/20230629005637-260410?](https://www.chinatimes.com/realtimenews/20230629005637-260410?chdtv)
chdtv，2023 年 6 月 29 日瀏覽。

Introducing Superalignment, <https://openai.com/index/introducing-superalignment/>，2023 年 6 月 7 日瀏覽。

京都にある高台寺で 23 日に公開されたのはアンドロイド観音「マインダー」です，<https://www.walkerplus.com/article/180566/>（2019 年 2 月 24 日公開）。

