



ChatGPT

末日的機器論

ChatGPT, or
the Eschatology
of Machines

許煜

Text by Hui Yuk

譯—李威儀
Translation by Lee Wei-I

與其中止 AI 的發展，
不如中止對機器的擬人化刻板印象，
並發展適當的假體文化。

ChatGPT 自 2022 年 11 月推出以來，便引發了狂熱與恐懼。它對不同語言的語義及語法的嫺熟掌握（但尚未精通內容），令只是期待普通聊天機器人的使用者感到驚訝。有些大學院校立即禁止學生使用 ChatGPT 來寫論文，因為它的表現優於多數人類學生。報紙上的專欄文章則宣布了教育的終結——不僅因為學生可以用它來做作業，還因為 ChatGPT 比很多老師能提供更多的資訊。人工智慧似乎已征服了另一個根據古典哲學來定義人性的領域：邏各斯（logos）。恐慌隨著存在領域（existential territory）¹ 的進一步喪失而加劇。當氣候的崩壞和機器人的反叛召喚出了末世，人類歷史的世界末日想像變得更加強烈。

末日對於現代人來說從來都不陌生。事實上，哲學家卡爾·洛維特（Karl Löwith）在 1949 年出版的《歷史的意義》（Meaning in History）一書即指出，現代歷史哲學——從黑格爾（Hegel）到布克哈特（Burckhardt）——就是末世論的世俗化²。歷史的終極目標便是使超越的變成內在（makes the transcendent immanent），不論是耶穌基督的第二次降臨或只是成為神人（Homo deus）。更廣泛地說，這種對歷史時期的聖經式或亞伯拉罕式想像，為人類的存在提供了許多深刻反思，卻也阻礙了對於我們的未來之理解。

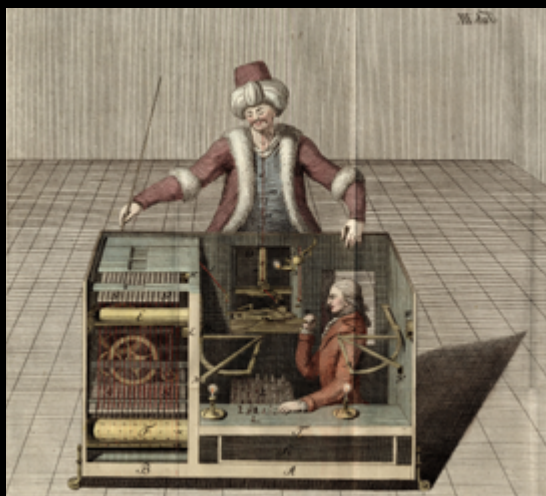
1960 年代，漢斯·布魯曼伯格（Hans Blumenberg）反對洛維特的世俗化論點，以及卡爾·施密特（Carl Schmitt）所主張的「現代國家理論的所有重要概念皆為世俗化的神學概念」³。布魯曼伯格反駁道，將現代理解為神學概念的世俗化或換位，削弱了現代的合法性；現代性的某些意涵並不能被神學的世俗化所化約⁴。同樣地，人工智慧的新奇性和重要意涵也正被末日想像、對機器的現代刻板印象以及產業的廣告宣傳所掩蓋。

這並非是指我們要否認氣候變化和抵制人工智慧。正好相反，應對氣候變化應是我們的首要任務，而發展人類與技術之間富有成效的關係也應是如此。但要做到這點，我們必須對人工智慧有充分的理解，而且不僅是技術上的理解。火車、汽車以及後來飛機的發明也曾引發心理上和經濟上的巨大恐懼，但今天很少有人擔心這些機器會脫離我們的控制。反而，汽車和飛機是日常生活的一部分，並且通常象徵著興奮和自由。那為什麼大家對人工智慧這般恐懼呢？

要理解以 ChatGPT 作為前鋒的這一技術新浪潮，我們可以從約翰·瑟爾（John Searle）自1980年所展開的著名思考實驗「中文房間」（Chinese Room）開始，該實驗以假裝邏輯推理的形式掩蓋了計算型機器令人討厭的刻板印象。在此思考實驗中，瑟爾假想自己獨自在一個房間裡，按照以英文撰寫的符號處理指示，去回應從門縫塞進來的中文輸入資訊。在此實驗中，瑟爾並不認得中文：「我不懂中文，無論是寫還是說，而且……我甚至沒有信心能辨別中文字與日文字或無意義的亂畫線條」⁵。然而，他認為只要有正確的指示和規則，他就可以做出回應，令房間外面的人相信他懂中文。簡言之，瑟爾宣稱僅僅因為機器能遵循中文指示，並不意味著機器能夠理解中文——而這層理解是所謂強 AI（與弱 AI 相反）的特徵。在此，我們需要對瑟爾用來區分強 AI 和弱 AI 的程序多些闡述，以支持他的決定性結論：理解首先意味著理解語義。雖然語法可以在軟體裡寫死（hard-coded），但語義的意涵仍會隨情況和環境而變化。瑟爾的「中文房間」適用於仍有如十八世紀機器一樣運行的電腦，例如「消化鴨」⁶或「機械土耳其人」。然而，這不是我

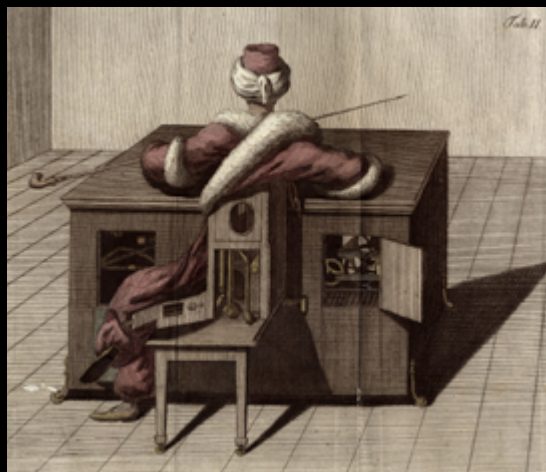
們今天處理的機器類型。諾姆·杭士基（Noam Chomsky）、伊恩·羅伯茲（Ian Roberts）和傑佛瑞·瓦圖默（Jeffrey Watumull）將 ChatGPT 指稱為「用於圖型比對（pattern matching）的笨拙統計引擎」⁷是正確的。然而，我們必須瞭解，雖然模式是資訊的主要特徵，但 ChatGPT 所做的不單單是模式比對。

這種對電腦語言句法的批評，乃基於一種機械認識論，它假定了線性因果律——即有因必有果。人們可以反轉這個因果過程以探得最終的因：原動力（the prime mover，亦即對於第一因的預設）以及所有線性推理的最終命運。與線性因果律和機械論哲學相反，十八世紀見證了基於機體論（organism）的哲學思維之興起，伊曼努爾·康德（Immanuel Kant）的《判斷力批判》（Kritik der Urteilskraft）便是最重要的貢獻之一。正如我在前作提出的，康德設下了一個新的哲學思考的條件，即哲學必須成為有機的；換言之，有機即標誌著哲學思維的全新開始⁸。而如今重要的是，要知道康德所建立的哲學思考條件在模控學（cybernetics）之後便已完結了⁹。



機械土耳其人（Mechanical Turk）是一台人型弈棋裝置，由奧地利人沃夫岡·馮·肯佩倫（Wolfgang von Kempelen）於1770年發明，聲稱可以在棋戰中擊敗人類對手，但實則是由精通棋術的人在機器內操控

圖片出自約瑟夫·弗里德里希·馮·拉克尼茨（Joseph Friedrich von Racknitz），《關於馮·肯佩倫先生的棋手及其複製品》（Ueber den Schachspieler des Herrn von Kempelen und dessen Nachbildung），1789年出版，柏林洪堡大學圖書館典藏



與其將機器和人類神秘化， 不如瞭解我們目前的技術現實 以及它與各種人類現實的關係。

模控學是諾伯特·維納（Norbert Wiener）於1943年前後創造的術語，由1940年代末和1950年代初參與梅西會議（Macy Conference）探討模控學的一群科學家和工程師提出。根據吉爾別·西蒙東（Gilbert Simondon），模控學旨在成為一門能夠統一所有學科的通用科學，這是十八世紀百科全書式知識論的新轉化¹⁰。不同於十七世紀的「機械式機器」（mechanical machines），模控學以反饋的概念來定義新的「模控學式機器」（cybernetic machines）之操作。維納在其1948年的開創性著作中聲稱，模控學已克服了以牛頓（Newton）和柏格森（Bergson）為代表的機械論和生機論（vitalism）之間的對立，因為模控學式機器是根據一種新的非線性形式的因果律或遞歸性（recursivity），而不是基於脆弱且難起作用的線性因果關係——脆弱是由於它不知道如何調節自己的運作模式。想像一下機械錶：當其中一個齒輪發生故障時，整個手錶都會停擺。在此線性機制下，若不徹底升級硬體，它的推理速度就不可能呈指數級增長。

如果說機械論與機體論之間的對立是現代哲學重要辯論的特點，並決定著它的發展方向，當許多質疑 AI 和 ChatGPT 的言論都假設機器只是機械

性的所以無法理解語義的意義，那麼這場辯論至今仍在持續。同樣地，當涉及語義的意義時，宣稱機器只是對人類理解能力的失敗模仿也不正確。哲學家暨認知科學家布萊恩·肯特韋爾·史密斯（Brian Cantwell Smith）嚴厲批評了這種擬人化思維，並為機械的意向性辯護。對他而言，即使人們在機器中找不到人類的意向性，它仍是意向性的一種形式。它仍是語義的，儘管不是以人類語言的意義¹¹。對於重新思考我們與機器的關係來說，將擬人化的語義與機器的語義分開至關重要，而這只是第一步。

瑟爾的論點從根本上忽略了現今機器展現的遞歸計算形式。有人或許會爭辯說，電腦科學不應與模控學混為一談，因為模控學是一門過於包山包海的科學。然而，人們也可以思考哥德爾（Kurt Gödel）的遞歸函數以及它與圖靈機（Turing Machine）和阿隆佐·邱奇（Alonzo Church）的「λ 演算」（lambda calculus）的等價性（這是計算史上的重要事件）¹²。「遞歸」一詞不僅屬於模控學，它也屬於後機械主義思維。模控學的出現，只是宣告了在模控學式機器中實現這種遞歸思維的可能性。

一如哥特哈德·岡瑟（Gotthard Günther）和西蒙東所準確觀察到的，現今機器中的「智慧」是一種反思性的操作形式。對於岡瑟來說，模控學是黑格爾的邏輯之實現；而對西蒙東來說，模控學僅在康德的《判斷力批判》對反思性判斷進行闡述時才論及¹³。「反思性思維」通常與人類有關而不是機器，因為機器只執行指令，不會對指令本身進行深思。但自從1940年代推展模控學以來，此一術語也就用以描述機器的反饋機制。機器的反思性思維對那些沒有準備好接受其存在的人類而言有著驚人的力量，即使只是作為一種初步和基本的反思形式——純粹為形式性的，所以不足以處理內容。由此我們便可以理解為什麼 ChatGPT 對於像楊立昆（Yann Le Cun）這樣的電腦科學家來說「並不特別創新」也「沒什麼革命性」¹⁴。只有透過處理內容，機器才能邁向所謂的技術奇點。目前，奇點仍是一個迷思——且當它被形容為即將到臨的未來之時，就是一個具誤導性且有害的迷思。即使我們將奇點與神學意義或末日論連結起來，這對理解人工智慧或者它的未來而言也沒有任何貢獻。

遞歸機器是理解人工智慧發展和演變的關鍵，而不是線性機器。人類將如何面對這種新型機器？西蒙東提過類似問題：當技術變得具有反身性時，哲學將扮演什麼角色？布萊恩·肯特韋爾·史密斯認為，AI 被侷限於計算能力，而非判斷能力，但很難說這種區別能持續多久¹⁵。也許有太多努力已經浪費在區分機器和人類上了。

當馬和牛等家畜取代人類作為動力的提供者時，人類並沒有感到不安。相反地，人們對於能從重複且令人厭煩的勞動中解脫出來感到欣然接受。當蒸汽機取代動物時也是這樣。它們的效率更高，所需的人為關注更少。西蒙東在1958年出版的《論技術物件的存在模式》（Du mode d'existence des objets techniques）一書中確切指出，資訊機器取代熱動力型機器標誌著一個關鍵時刻：人類從生產的中心轉移了。工業時代之前的工匠能夠創造一種環境，讓他們以身體和智慧來彌補他們自主性不足的簡單工具。而在資訊式機器或模控學式機器的時代，機器本身成為資訊的組織者，人不再處於中心位置，即使人們依然認為自己是機器的指揮者和資訊的組織者。這是人類受自己對機器的刻板印象之苦的時刻：他們錯誤地將自己視為中心，因而持續感到沮喪，慌張地尋求身份認同。

機器中存在的現實與人類操作的現實是疏離的。技術進化的必然過程是由非線性因果關係的引入所驅動的，這使機器能夠處理偶發事件。學習型機器是一種能辨別偶發事件（例如噪音和故障）的機器。它可以區分無組織的輸入（inputs）和必要的輸入。藉由解釋偶發事件，學習型機器會改進其決策模式。但即使如此，機器也需要人類來區分正確和錯誤的決策，以便持續改進。在發展中國家，一種新型廉價勞動力的會雇用人類來告訴機器結果是否正確，無論是臉部識別掃描還是 ChatGPT 回應。這種新的勞動形式，剝削了那些在與我們互動的機器後面的看不見的辛苦工人，這種新的勞動形態經常被對資本主義的一般性批評所忽視，這些批評對自動化程度的不足表示失望。而這是當前馬克思主義的技術批判之弱點。





這幅名為《太空歌劇院》(Théâtre D'opéra Spatial) 的影像，由傑森·艾倫 (Jason Allen) 運用生成式 AI 軟體 Midjourney 製作，並在未事先揭露製作方法的情況下，於 2022 年獲得美國科羅拉多州博覽會數位藝術類首獎，引發「創作」的定義問題與版權爭議。2023 年 8 月，美國聯邦法官作出 AI 生成的作品不具著作權保障的判決

西蒙東在《論技術物件的存在模式》中提出了一個關鍵問題：當人類不再是資訊的組織者時，他們能扮演什麼角色？人類能否從勞動中解放出來？正如漢娜·鄂蘭 (Hannah Arendt) 在《人的條件》(The Human Condition) ——與西蒙東的書同年出版——所質疑的那樣，這種解放只會導致消費主義，讓藝術家變成能夠創造的「最後之人」¹⁶。消費主義就此成為人類行動的限制。鄂蘭從人類現實的角度看待機器，認為機器取代了「創造之人」(Homo faber)，而西蒙東則表明，若無法應對和整合機器的技術現實，將致使人與機器、文化與技術之間不幸的敵對。這種敵對不僅是恐懼的來源，它甚而是對於由產業的廣告宣傳和消費主義所形塑之技術的非常有問題的理解。正是在這種消極態度中，原始主義式的人文主義崛起，將愛視為人類的最後依托。

距西蒙東提出這些問題至今已有六十多年了，而問題卻仍未獲解決。更糟的是，它們被技術樂觀主義和文化悲觀主義所掩蓋，前者提倡持續加速，後者則充當心理治療。這兩種傾向都源於對機器擬人化的認識，覺得機器就應該模仿人類。(西蒙東強烈批評採取這種觀點的模控學，儘管這並不完全合理。)如今，這種模仿觀點最具諷刺意味的表達是在藝術領域，該領域試圖證明機器可以做出巴哈或畢卡索的作品。一方面，恐慌的人類反覆詢問什麼樣的工作可以避免被機器取代；另一方面，科技產業則有意識地致力於以機器的自動化來取代人類的介入。人類生活在產業對於「取代」的自我實現預言中。事實上，這個產業一再藉由宣告這個或那個工作的終結來重製這種取代之論，好像一場革命就要到來一樣，儘管社會結構和我們的社會想像都未曾變過。

在資本主義社會和所謂的共產主義社會裡，取代之論都沒有轉變為解放之論。持平來說，一些加速主義者意識到了這一點，並試圖復興馬克思的全面自動化願景。不過要是高中物理能更受歡迎的話，我們就能對加速度 (acceleration) 有更細緻的概念，因為加速度並不意味著速率 (speed) 的增加，而是速度 (velocity) 的增加。主流的討論裡並沒有闡述人工智慧作為假體輔助功能的未來願景，而是僅僅將其視為挑戰人類智慧並取代腦力勞動。今天的人類不再去夢想。如果說飛行的夢想導致了飛機的發明，那麼現在我們有了更多的機器噩夢。最終，技術樂觀主義 (以超人類主義 [transhumanism] 的形式) 和文化悲觀主義便會在他們雙方對末日結局的預測之中相遇。

人類創造力必須採取完全不同的方向，並將人機關係提升到取代的經濟理論以及對於互動性的幻想之上。它必須走向存在主義分析。技術的假體本質 (prosthetic nature) 必須超越其功能性，因為自人類誕生以來，獲取真理始終仰賴工具的發明和運用。很多人對這一事實視而不見，這使得機器進化與人類存在之間的衝突，似乎源自一種於文化中根深蒂固的意識形態。

我們生活在各式正反饋 (positive feedback) 的循環中，它表徵為文化。但從現代工業社會開始，人的身體就服從於重複的節奏，人類的心靈也隨之被納編於產業的預言。無論是美國夢還是中國夢，人類的巨大潛力都已被消費主義意識形態所抑制。過去，哲學的任務是限制機器造成的自大傲慢，並以真理的名義將人類主體從反饋循環中解放出來。如今，技術哲學家反而都急於確認這類反饋循環是文明的必然之路。人類現在明瞭了技術的中心地位，期許像解決技術問題一樣來解決所有問題。速率和速度統治著整個社會，就像它們曾經統治著工程學科一樣。教育工作者對於在幾年內就實現典範變革的渴望，則貶抑了對技術問題的任何根本思考，我們終究會再陷入反饋循環之中。於是，大學院校繼續為科技產業製造人才，而這些人才繼續開發更有效的演算法來剝削使用者的隱私並操控他們的消費方式。對大學院校來說，去處理這些問題應當是比考慮禁止 ChatGPT 更加迫切的事。

人類能否擺脫這種在當代文化中如此深植於自我實現的預言裡的正反饋循環？1971年，葛雷格里·貝特森 (Gregory Bateson) 描述了將酗酒者困住的一種反饋循環：喝一杯啤酒不會死，好，既然我開喝了，第二杯也不會有事；嗯，都已經兩杯了，那為什麼不來第三杯？要是僥倖的話，酗酒者可能會藉由「觸底」(hitting bottom，這是貝特森使用的術語)——例如，從致命的疾病或車禍中倖存下來——來擺脫這種正反饋的循環¹⁷。然後那些僥倖的倖存者便會與神性建立親密關係。人類，作為現代的酗酒者，可否憑藉他們的集體智慧和創造力，來擺脫這種觸底的命運？或者說，人類能否徹底轉變並將創造力推進至另一方向？

當前的智慧型機器不正提供了這個機會嗎？機器作為輔助的假體而非生搬硬套的模式追隨者，它將人類從重複工作中解放出來，並幫助我們實現人的潛能。而如何獲得這種變革能力，就是我們今天在實質上需要關心的問題，而不是機器能否思考的爭辯，那只是存在危機和先驗幻像 (transcendental illusion) 的表現。也許一些關於人機關係的新前提能夠解放我們的想像力，以

下提出三個（當然還能擴充更多）：

一、與其中止 AI 的發展，不如中止對機器的擬人化刻板印象，並發展適當的假體文化 (culture of prosthesis)。技術應該用來實現其使用者的潛能（於此我們必須和阿馬蒂亞·森 [Amartya Sen] 的能力理論 [capability theory] 進行對話），而不是成為使用者的競爭對手，或將他們簡化為消費模式。

二、與其將機器和人類神秘化，不如瞭解我們目前的技術現實以及它與各種人類現實的關係，使技術現實能與之整合在一起，以維持和再造生物多樣性、心智多樣性和技術多樣性¹⁸。

三、與其重複世界末日的歷史觀（這一觀點在科耶夫 [Alexandre Kojève] 和福山 [Francis Yoshihiro Fukuyama] 的歷史終結論中以最世俗的形式表達過），不如將理性從走向世界末日終局的宿命之路上解放出來。這種解放將開闢一個場域，讓我們能夠試驗與機器和其他非人類一起的倫理生活。

任何發明都是帶有限制條件和問題的。儘管這些限制條件更多是概念性的而不是技術性的，但忽視概念恰恰是邪惡得以滋生的原因，也是形式 (form) 超過基礎 (ground) 的曲解結果¹⁹。唯有掙脫文化偏見和科技產業的自我實現的預言，我們才能對未來的可能性有更深刻的洞察，這不能只靠數據分析和模式提取。在我們達成這點之前，很有可能，我們這個時代的產業預言家將會明白，機器會比他們更能夠預測未來。 ⑩

本文英文版載於《e-flux Journal》，137期，2023年6月

許煜，荷蘭鹿特丹伊拉姆斯大學哲學系教授。他曾於香港大學和倫敦金匠學院學習電腦工程與哲學，並在法國哲學家貝拿爾·斯蒂格勒 (Bernard Stiegler) 指導下完成哲學博士論文；其後在德國呂訥堡大學取得哲學教授資格 (Habilitation)，並先後任教於金匠學院、呂訥堡大學、包浩斯大學、莫斯科斯特雷卡研究所、中國美術學院和香港城市大學。著有《論數碼物的存在》、《論中國的技術問題：宇宙技術初論》、《遞歸與偶然》及《藝術與宇宙技術》。

與其重複世界末日的歷史觀，
不如將理性從走向世界末日終局的
宿命之路上解放出來。

註釋

- 1 【編註】「存在領域」是一個哲學概念，它指涉個體對於自身存在的感知，以及個體如何在自己的主觀世界中建立自己的身份和意義。
- 2 「這種意指歷史乃循不可逆轉的方向朝未來目標前進的對歷史的西方觀點，並不只是西方的。它本質上是一種希伯來和基督教的假設，即歷史朝向終極目標前進，並由至高無上的智慧與意志來主導——以黑格爾的話來說，是由精神或理性作為「絕對強大的本質」來主導。」出自：Karl Löwith, *Meaning in History*. Chicago: University of Chicago Press, 1949. p.54.
- 3 Carl Schmitt, *Political Theology: Four Chapters on the Concept of Sovereignty*. Trans. George Schwab. Chicago: University of Chicago Press, 2005. p.36.
- 4 Hans Blumenberg, "Progress Exposed as Fate". in *The Legitimacy of the Modern Age*. Trans. Robert M. Wallace. Massachusetts: The MIT Press, 1985.
- 5 John. R. Searle. "Minds, Brains, and Programs". *Behavioral and Brain Sciences* 3, no. 3, 1980, p.418.
- 6 【編註】消化鴨 (Digesting Duck/Canard de Vaucanson) 是由法國人雅克·德·沃康松 (Jacques de Vaucanson) 於1739年發明的一台機械鴨子，並宣稱它可以透過自動化的機械方式來「消化」及「排泄」放進體內的穀物，但後來證實它並不能自動運作。
- 7 Noam Chomsky, Ian Roberts, and Jeffrey Watumull. "The False Promise of ChatGPT". *New York Times*, March 8, 2023.
- 8 Hui Yuk. *Recursivity and Contingency*. Maryland: Rowman and Littlefield, 2019.
- 9 Hui Yuk. "Philosophy after Automation". *Philosophy Today* 65, no.2, 2021.
- 10 Gilbert Simondon. "Technics Learned by the Child and Technics Thought by the Adult." in *On the Mode of Existence of Technical Objects* (1958). Trans. Cecile Malaspina & John Rogrove. Minneapolis: Univocal, 2017.
- 11 「許多人主張，計算系統的語義本質上是衍生的或歸類的——即是從外部觀察者或使用者來歸因的意義上來說，書籍和標誌所具有的那種語義——它與人類的思維和語言成對比，後者被認為是原始或真實的。我對這種區別的最終效用（和明確度）以及它對電腦的適用性表示懷疑。」Brian Cantwell Smith. *On the Origin of Objects*. Massachusetts: The MIT Press, 1996. p.10. 對其著作的延伸討論可參見：Hui Yuk. "Digital Objects and Ontologies". in *On the Existence of Digital Objects*. Minneapolis: University of Minnesota Press, 2016.
- 12 【編註】1930年代中期，邱奇和圖靈分別發展了「λ演算」和「圖靈機」，他們各自皆主張可以用自己的演算法／圖靈機來計算自然數上的任何函數，這引出了所謂的「邱奇—圖靈論題」這一假設，亦即如果一個問題可以用計算方法解決，那麼它也可以用計算機器解決。該假設為電腦科學帶來了深遠影響。
- 13 Gotthard Günther. *Das Bewusstsein der Maschinen: eine Metaphysik der Kybernetik*. Agis-Verlag: 1957. Gilbert Simondon. *Sur la philosophie*. PUF, 2016. p.180.
- 14 Tiernan Ray. "ChatGPT Is 'Not Particularly Innovative' and 'Nothing Revolutionary' Says Meta's Chief AI Scientist". *ZDNET*. January 23, 2023.
- 15 Brian Cantwell Smith. *The Promise of Artificial Intelligence: Reckoning and Judgment*. Massachusetts: The MIT Press, 2019. 我在《藝術與宇宙技術》(Art and Cosmotekhnics, e-flux and University of Minnesota Press 出版，2021年)的第三章中延伸探討了他的論點。
- 16 Hannah Arendt. *The Human Condition*. 2nd Ed. Chicago: University of Chicago Press, 1998. p.127.
- 17 Gregory Bateson. "The Cybernetics of 'Self': A Theory of Alcoholism" in *Steps to an Ecology of Mind*. Maryland: Jason Aronson, 1987.
- 18 Hui Yuk. "For a Planetary Thinking." *e-flux journal*, no. 114, December 2020.
- 19 關於邪惡的問題，參見：F. W. J. Schelling, *Philosophical Investigations into the Essence of Human Freedom*. Trans. Jeff Love & Johannes Schmidt. New York: SUNY Press, 2006。